**BAYESIAN IDENTIFICATION OF DISCONTINUOUS FIELDS WITH AN ENSEMBLE-BASED VARIABLE SEPARATION MULTISCALE METHOD**

**NA OU**, GUANG LIN, AND LIJIAN JIANG

**Abstract.** This work presents a multiscale model reduction approach to discontinuous fields identification problems in the framework of Bayesian inference. An ensemble-based variable separation (VS) method is proposed to approximate multiscale basis functions used to build a coarse model. The variable-separation expression is constructed for stochastic multiscale basis functions based on the random field, which is treated Gaussian process as prior information. This end, multiple local inhomogeneous Dirichlet boundary condition problems are required to be solved, and the ensemble-based method is used to obtain variable separation forms for the corresponding local functions. The local functions share the same interpolate rule for different physical basis functions in each coarse block. This approach significantly improves the efficiency of computation. We obtain the variable separation expression of multiscale basis functions, which can be used to the models with different boundary conditions and source terms, once the expression constructed. The proposed method is applied to discontinuous field identification problems where the hybrid of total variation and Gaussian (TG) densities are imposed as the penalty. We give a convergence analysis of the approximate posterior to the reference one with respect to the Kullback-Leibler (KL) divergence under the hybrid prior. The proposed method is applied to identify discontinuous structures in permeability fields. Two patterns of discontinuous structures are considered in numerical examples: separated blocks and nested blocks.

**Key words.** stochastic multiscale basis functions, ensemble-based VS method, TG prior, discontinuous field identification

**AMS subject classifications.** 65N99, 65N30, 35R60

1. **Introduction.** The characterization of the permeability is critical for the modeling subsurface flow and transportation. Properly quantifying the uncertainties induced by the permeability field is very important for making reliable probabilistic based predictions and future decisions. The involved unknown spatially distributed parameters are need to be estimated in some quantified sense, such that the observations can be best explained.

Because the dimension of measurement data is usually much smaller than the dimension of the unknown spatially distributed parameters, the limited number of measurements can only depict the permeability field with high uncertainty. The weak sensitivity of measurements to the unknowns and the noise in the observations, often results in ill-posedness of the inverse problems. The prior regularizing information is naturally utilized as a penalty to remedy this issue. Gaussian measures are widely used as the prior distributions, which also work as the Tikhonov regularization term used from the view of optimization problems. A rigorous Bayesian framework for the inverse problems in function spaces is developed in [27] for smooth unknown random fields.

However, in practical situations, the spatially varying fields can be better represented as consisting of a few relatively uniform geologic facies or zones with abrupt
changes at their boundaries. One approach to facies detection is the level-set method \cite{24, 5, 11}, in which the shapes of facies are not required to be predetermined, to identify zonal structures. The geometry of the discontinuity of the permeability is represented implicitly by level set functions. The authors of \cite{6, 8} combine the level-set method with the total variational (TV) norm \cite{26} to control both the jumps in the permeability and the length of the level sets. The TV norm is also used as the prior density in \cite{19} to describe the sharp changes of the spatially varying field in the framework of Bayesian inference. The well-defined posterior measure in the function space is developed in paper \cite{30}, where TV-Gaussian (TG) prior is presented to address the infinite-dimensional Bayesian inverse problems, for discontinuous random fields. We choose the TG prior to identify the discontinuous permeability field in this paper.

Markov chain Monte Carlo (MCMC) methods \cite{21, 3, 4} are the standard techniques for sampling the posterior distribution in Bayesian inferences. An infinite number of samples are required to pursue the convergence and reliability of the chain, which implies we need to simulate large-scale PDE-based forward models hundreds of thousands or even millions times, for characterizing the posterior distribution. The discretization of the spatial domain results in high-dimensional parameters inevitably, while the efficiency of many standard MCMC chains degrades as the dimension of the parameters increases. The computational defect of standard MCMC method has promoted the development of efficient MCMC techniques, such as incorporating the data information to improve the sampling efficiency: the two-stage and cheaper approximation strategies \cite{12, 7}, the stochastic Newton MCMC \cite{22} and the dimension-independent likelihood-informed MCMC \cite{10}. The preconditioned Crank-Nicolson MCMC (pCN-MCMC) method described in \cite{9}, of which the acceptance probability is invariant with respect to the dimension of the parameter space, is applied to draw samples from the Bayesian posterior in the present work.

As noted before, exploration of the posterior distribution requires repeated evaluations of the forward operator, and a large number of samples are needed to ensure reliable estimates of the inference. One attempts at accelerating Bayesian inference in computationally intensive inverse problems have relied on reduction of the stochastic forward model \cite{2, 16, 18, 29}. Multiscale Finite Element Method (MsFEM) \cite{17} is one of the model reduction methods and gives a coarse-scale equation through variational formulation by multiscale basis functions, which contain fine scale information. The generalized Multiscale Finite Element Method (GMsFEM) \cite{14, 13} inherits the important features from MsFEM, once the multiscale basis functions constructed, it can be used repeatedly for models with different source terms and boundary conditions, while the coarse space in GMsFEM is more flexible. A novel variable-separation (NVS) method proposed in \cite{20} provides a variable-separated form approximation of the functions dependent with the stochastic inputs, recursive formulations can be derived for stochastic functions. The author of \cite{25} apply the NVS method to the time-dependent fractional equations and the involved Bayesian inverse problems, resulting in good approximation of the posterior density. Proper orthogonal decomposition (POD) reduced-order model is incorporated into the ensemble-based method in \cite{15} to significantly reduce the costs of determining multiple solutions.

In the paper, we develop an ensemble-based variable separation multiscale (VS-Ms) method in accelerating the forward model simulation. First, we build local problems based on GMsFEM, with multiple determined inhomogeneous Dirichlet boundary conditions, to obtain the reduced snapshot space. We use the VS method to solve
the local problems in each coarse neighborhood such that the local functions can be expressed in variable separation form. However, it would be time consuming if we approximate each local function by the VS procedure. On the other hand, as the number of multiscale basis functions retained in the coarse block increase, the time for calling the variable separation form may increase to a large magnitude, even the separated terms is fixed at a small number.

To efficiently construct multiscale basis functions in random space, the ensemble-based method is used in this paper. We run VS process to solve the local problems with the ensemble of the inhomogeneous Dirichlet boundary conditions, and then reverse the physical basis for each local function basing on the mechanism and residual equation of the VS process. Then we have stochastic multiscale basis functions in one coarse block sharing the same interpolate rule, i.e., for any random parameter, the separated stochastic functions from the VS form need only to be called one time, and the multiscale basis functions can then be approximated by multiplying the calculated stochastic part with the corresponding physical parts. The derived multiscale basis functions can capture most characteristics of the ones built by directly solving local problems.

The proposed method shares the merit from the GMsFEM and can be applied to the model with different boundary conditions and source terms, once the stochastic multiscale basis functions expression is constructed. We apply it to discontinuous field identification problems under the framework of Bayesian inference. The variable separated form of the multiscale basis functions are constructed with respect to Gaussian random field, of which the covariance matrix depends on the relative distance of the spatial points. In practice, the ensemble-based VS-MsFEM approach is used only 4 times in constructing variable separated form of multiscale basis functions. More details on the proposed method will be displayed in Section 3. The number of local basis functions selected in each coarse neighborhood dominates the effects on the accuracy. The proposed method is applied in recovering discontinuous fields with different structures, e.g., two separated blocks and nested blocks. We give a brief convergence analysis of the approximate posterior to the reference one with respect to the Kullback-Leibler divergence, and the numerical results verify the convergence of the proposed method.

The outline of this paper is organized as follows. In Section 2, we briefly introduce the Bayesian inference with hybrid prior, the pCN-MCMC algorithm and the general convergence result of the approximated posterior, which is induced by the reduced model, under the TG prior. Section 3 describes the proposed model reduction method, where the local basis functions stem from the snapshots of GMsFEM, and the VS method is applied to inhomogeneous Dirichlet boundary conditions and the ensemble-based VS method is used to obtain the variable expression for the problems with multiple boundary conditions. Some numerical examples are presented in Section 4, we carry out a few numerical simulations to verify the efficiency and accuracy of our proposed method. The proposed approach is also applied to posterior densities approximation, where discontinuous fields with different structures are identified. Some conclusions and comments are made finally.

2. Problem setting. We consider a system described by partial differential equations

\[ \mathcal{F}(u; \kappa(x)) = f, \quad x \in D \subset \mathbb{R}^2, \]  

(2.1)
where the unknown discontinuous coefficient $\kappa(x) \in L^{\infty,+}(D)$,

$$L^{\infty,+}(D) := \{L^{\infty}(D)|\kappa(x) > 0\}.$$  

The operator $F$ represents the forward model, which describes the relation between the coefficient $\kappa(x)$, external source/sink term $f$ and the output state $u$. The coefficient identification problem for (2.1) seeks to determine the coefficient $\kappa(x)$ in such a way that the observed output matches the indirect measurement $d$ in a prescribed sense, the data $d$ is related to $\kappa(x)$ via

$$d = G(\kappa) + \zeta,$$

where $G : L^{\infty,+}(D) \rightarrow \mathbb{R}^n$ and $\zeta$ is an $n_d$-dimensional independent and identically distributed (i.i.d.) Gaussian random vector with mean zero and standard deviation $\sigma$. The likelihood is given by

$$\pi(d|\kappa) = (2\pi\sigma^2)^{-\frac{n_d}{2}} \exp\left(-\frac{\|d - G(\kappa)\|_2^2}{2\sigma^2}\right).$$

Due to the sparsity of the given measurement, some regularization terms are required. In order to guarantee the well-posedness of the inverse problem, the weighted $L_2$ norm penalty is imposed. We set

$$\log \kappa(x) \sim \mathcal{GP}(0, C(x_1, x_2)),$$

where $C(\cdot, \cdot)$ is the kernel operator, $\mathcal{GP}(\cdot, \cdot)$ represents the Gaussian process. We concentrate on the squared exponential kernel in this paper, i.e., it has the form

$$C(x_1, x_2) = \sigma_x^2 \exp\left(-\frac{(x_1^1 - x_2^1)^2}{2l^2_{x_1}} - \frac{(x_1^2 - x_2^2)^2}{2l^2_{x_2}}\right),$$

where $\sigma_x^2$ is the variance of the random field, $l_{x_1}$ and $l_{x_2}$ are the length scales along the direction $x_1$ and $x_2$, respectively. Other penalty term is also required to pursue the discontinuous property of coefficient. We follow [30] and the edge pursuing penalty term $\mathcal{R}(\kappa)$ is set as total variation norm, i.e.,

$$\mathcal{R}(\kappa) = \lambda \|\kappa(x)\|_{TV} = \lambda \int \|\nabla \kappa(x)\|_2 dx,$$

where $\lambda$ is a prescribed positive constant. The infinite problem has been proved well-defined in paper [30], we discretize the spatially varying field matching with the grid, on which the forward model is solved, i.e., $\kappa(x)$ can be approximated by

$$\kappa(x) := \kappa(x; \xi) = \sum_{i}^{N_{x_1}N_{x_2}} \mathbf{I}_{\{x \in D_i\}} e^{\xi_i},$$

where $N_{x_1}$ and $N_{x_2}$ are the number of partition nodes in the $x_1$ and $x_2$ direction, respectively. $\bigcup_i D_i$ are the non-overlapping partition of the domain $D$, $\mathbf{I}_{\{x \in D_i\}}$ is the indicator function, and

$$\xi \sim \mathcal{N}(0, \Sigma),$$
where $\Sigma$ is the covariance of the joint Gaussian distribution. Then we make inference to the unknown $\xi \in \mathbb{R}^{n_\xi}$, where $n_\xi = N_{x_1}N_{x_2}$ is the dimension of $\xi$. Then the posterior expression is

$$\pi(\xi|d) \propto \exp \left( - \frac{1}{2\sigma^2} \left( \frac{d - G(\xi)}{\sigma^2} \right)^2 - \frac{1}{2} \|\xi\|^2_{\Sigma} - R(\xi) \right),$$

where

$$\|\xi\|^2_{\Sigma} = \|\Sigma^{-\frac{1}{2}}\xi\|^2_{2},$$

and $R(\xi)$ is the discretized approximation with respect to $\xi$, more details about the discretization of the TV norm can be referred to [28, 19]. Posterior samplers are obtained by pCN-MCMC scheme, as following [30].

We use the ensemble-based VS multiscale (VS-MS) method, which will be described in detail in Section 3, to accelerate the sampling. Let $G_N$ be the approximated forward observation operator. Then the corresponding approximated posterior is

$$\pi_N(\xi|d) \propto \exp \left( - \frac{1}{2\sigma^2} \left( \frac{d - G_N(\xi)}{\sigma^2} \right)^2 - \frac{1}{2} \|\xi\|^2_{\Sigma} - R(\xi) \right).$$

For the convenience of notation, we will use $\pi^d(\xi)$ and $\pi^d_N(\xi)$ to denote the posterior density $\pi(\xi|d)$ and $\pi_N(\xi|d)$, respectively.

We use Kullback-Leibler (KL) divergence [23] to measure the difference between the approximated posterior and the reference one. For probability density functions $\pi^d_N(\xi)$ and $\pi^d(\xi)$, KL divergence is defined by

$$D_{KL}(\pi^d_N||\pi^d) = \int \pi^d_N(\xi) \log \frac{\pi^d_N(\xi)}{\pi^d(\xi)} d\xi.$$

$D_{KL}$ measures the difference between the two probability distributions and is nonnegative. For a given discretization, we have a brief convergence analysis without regard to the discrete error of the unknowns and the total variation penalty.

**Assumption 2.1.** For every $\epsilon > 0$, there is $M = M(\epsilon)$ such that, for all $\xi \in \mathbb{R}^{n_\xi}$,

$$\|G(\xi)\|_2 \leq \exp(\epsilon\|\xi\|_\Sigma^2 + M),$$

**Assumption 2.2.** The function $R : \mathbb{R}^{n_\xi} \to \mathbb{R}$ satisfying: For every $r > 0$, there is $K = K(r)$ such that, for all $\xi \in \mathbb{R}^{n_\xi}$ with $\|\xi\|_\Sigma < r$,

$$R(\xi) \leq K(r).$$

**Theorem 2.1.** Assume that the functions $G$ and $G_N$ satisfy Assumption 2.1, $R$ satisfies Assumption 2.2, and the observational error in (2.2) has an i.i.d. Gaussian distribution. Then the approximation posterior $\pi^d_N$ and the true posterior density $\pi^d$ are close with respect to the Kullback-Leibler divergence, i.e., there is a constant $C$ independent of $N$, such that

$$D_{KL}(\pi^d_N||\pi^d) \leq C\|G(\xi) - G_N(\xi)\|_{L^{2_0}}^2,$$

where $\pi_0$ is the Gaussian prior $\mathcal{N}(0, \Sigma)$.

The proof of the Theorem is provided in the Appendix A.
3. Ensemble-based VS-MS method. The dimension of $\xi$ depends on the discretization of the physical domain $D$, which could be very high. If constructing the surrogate model directly, it will suffer from severe curse of dimensionality. For a change, we construct surrogate model over local domains. We construct local problems stemming from GMsFEM, and then approximate the local functions on each coarse neighborhood by the ensemble-based VS scheme. The resulted variable separation expressions for stochastic multiscale basis functions share the common interpolate rule, which save the computational cost further.

We use the following model equation subject a boundary condition to show the construction of ensemble-based VS-MS basis functions,

$$-\text{div}(\kappa(x;\xi)\nabla u) = f(x) \quad \text{in} \quad D.$$ 

This equation can model the single-phase subsurface flow. We use generalized finite element method [1] to solve this equation on a coarse grid, where the basis functions are constructed by the proposed ensemble-based VS method. We refer this method to an ensemble-based VS-MS method. The generalized finite element method is standard (see [1]). In the paper, we will focus on the construction of multiscale basis functions on each coarse block.

3.1. Local problems based on GMsFEM. For GMsFEM, we need to precompute a set of multiscale basis functions. For a random sample taken from the prior (2.4), we solve the following local problem on each coarse block $\omega_i$ to construct the snapshot space,

$$\begin{cases} -\text{div}(\kappa(x;\xi)\nabla \varphi^i_l) = 0 & \text{in } \omega_i, \\ \varphi^i_l = \delta^h_l(x) & \text{on } \partial \omega_i, \end{cases} \quad (3.1)$$

where functions $\delta^h_l(x)$ are defined with respect to the fine grid boundary point on $\partial \omega_i$. The space of snapshots $V_{\omega_i}^{\text{snap}}$ can be constructed as

$$V_{\omega_i}^{\text{snap}} = \text{span}\{\varphi^i_1, \ldots, \varphi^i_{N_H}, 1 \leq l \leq M_{\text{snap}}^i\},$$

where $N_H$ is the number of coarse nodes and $M_{\text{snap}}^i$ is the number of boundary nodes on $\partial \omega_i$. The snapshot functions can be stacked into a matrix as

$$R_{\text{snap}} = [\varphi^1, \cdots, \varphi^{M_{\text{snap}}^i}].$$

In order to reduce the dimension of the snapshot space, we perform a spectral decomposition of the space of snapshots and solve the local problems

$$\begin{cases} -\text{div}(\kappa(x;\xi)\nabla \psi^i_j) = \lambda_j \kappa(x;\xi)\psi^i_j & \text{in } \omega_i, \\ \kappa(x;\xi)\nabla \psi^i_j \cdot \vec{n} = 0 & \text{on } \partial \omega_i, \end{cases} \quad (3.2)$$

where $\psi^i_j \in V_{\omega_i}^{\text{snap}}$. We choose the smallest $M_i$ eigenvalues of

$$\hat{A}\psi^i_j = \lambda_j \hat{S}\psi^i_j,$$

and take the corresponding eigenvectors in the space of snapshots by setting $\psi^i_j = \sum_l \psi^i_{j,l}\varphi^i_l$, for $j = 1, \cdots, M_i$, to form the reduced snapshot space, where $\psi^i_{j,l}$ are the coordinates of the vector $\psi^i_j$ and

$$\hat{A} = [a_{mn}] = \int_{\omega_i} \kappa(x;\xi)\nabla \varphi^i_n \nabla \varphi^i_m = R_{\text{snap}}^T \hat{A} R_{\text{snap}},$$
Bayesian discontinuous fields identification

\[ \tilde{S} = [s_{mn}] = \int_{\omega_i} \kappa(x; \xi) \varphi_n \varphi_m = R_{\text{snap}}^T S R_{\text{snap}}. \]

where \( A \) and \( S \) denote fine-scale matrices corresponding to the stiffness and mass matrices with the coefficient \( \kappa(x; \xi) \). The relationship between a coarse neighborhood and its coarse elements is illustrated in Figure 3.1.

The goal of Eqs. (3.1) and (3.2) is to construct a few effective boundary conditions in each coarse block, so that the functions \( \psi^i_j(x, \xi), j = 1, \cdots, M_i \) expanding the reduced snapshot space can be used to approximate the solution space. The construction of the reduced snapshot space is equivalent to solving the equation directly

\[
\begin{aligned}
&-\text{div}(\kappa(x; \xi) \nabla \phi^i_j) = 0, \quad \text{in } \omega_i, \\
&\phi^i_j = \psi^i_j(x, \xi), \quad \text{on } \partial \omega_i
\end{aligned}
\]

in the fine grid finite element space, where \( \psi^i_j(x, \xi) \) is treated as the effective boundary condition. However, it is inadvisable to build the reduced snapshot space by solving the equation above, since \( \psi^i_j(x, \xi) \) depends on the stochastic variable \( \xi \), and we have no explicit expression of it. As a compromise, we solve local problems

\[
\begin{aligned}
&-\text{div}(\kappa(x; \xi) \nabla \phi^i_j) = 0, \quad \text{in } \omega_i, \\
&\phi^i_j = \bar{\phi}^i_j(x), \quad \text{on } \partial \omega_i
\end{aligned}
\]

(3.3)

to construct the reduced snapshot space, where \( \bar{\phi}^i_j(x) = \psi^i_j(x, \bar{\xi}) \) is built with \( \kappa(x; \bar{\xi}) = \kappa(x; \xi) \). Here \( \bar{\xi} \) is an arbitrary realization taken from the prior, e.g., the mean of the random field. In this way, the boundary condition \( \bar{\phi}^i_j(x) \) is deterministic and pre-computed for any \( \xi \), then the local problem (3.3) can be solved directly.

It is reasonable to treat \( \bar{\phi}^i_j(x) \) as the effective boundary condition, since it is constructed based on the model information. Even when the model information is rough, the boundary conditions can sometimes be constructed effectively, which further leads to good approximation of the solution. Finally, we multiply the partition of unity functions (e.g., linear hat functions over coarse elements) by \( \phi^i_j(x, \xi) \) obtained through solving (3.3) to construct the local basis space.

3.2. VS method. We use the VS method proposed in [20] to construct surrogate model for solutions to Eq. (3.3). For the convenience of notification, we leave out the

![Fig. 3.1. Illustration of the 4 types of coarse neighborhood and the coarse element.](image-url)
scripts in $\phi^i_j(x, \xi)$, $\bar{\phi}^i_j(x)$ and $\omega_i$ in this section, to demonstrate the application of VS method to inhomogeneous Dirichlet boundary condition problems. Let $H^1_0$ be the usual Sobolev space. A weak formulation of Eq. (3.3) can be written as: find $\phi^0 \in V$ such that

$$a(\phi^0(\xi), v; \xi) = -a(\bar{\phi}, v; \xi), \quad \forall v \in V,$$

where $V := H^1_0(\omega), a(\cdot, \cdot)$ is a bilinear form on $V$ from the diffusion equation, and $\phi = \phi^0 + \bar{\phi}$ is the solution. The bilinear form $a(\cdot, \cdot, \xi)$ is assumed to be affine with respect to $\xi$, i.e.,

$$a(w, v; \xi) = \sum_{p=1}^{m_n} \kappa_p(\xi) a_p(w, v), \quad \forall v, w \in V, \forall \xi \sim \mathcal{N}(0, \Sigma), \quad (3.4)$$

where $\kappa_p(\xi)$ is stochastic functions with respect to $\xi$, $a_p : V \times V \to \mathbb{R}$ is a symmetric bilinear form. If $a(\cdot, \cdot, \xi)$ is not affine, the VS method [20] can still be used. Let $V_h \subset V$ be given finite dimensional approximation space. We seek the solution of Eq.(3.3) approximated by

$$\phi(x; \xi) \approx \phi^0_N(x; \xi) + \bar{\phi}(x),$$

and $\phi^0_N(x; \xi)$ has the separated form

$$\phi^0_N(x; \xi) = \sum_{q=1}^{N} \Phi_q(x) \eta_q(\xi), \quad (3.5)$$

where $\eta_q(\xi)$ are stochastic functions and $\Phi_q(x) \in V_h$ are deterministic functions, the separated term $N$ may be much less than the degree of freedom of the original problem.

Let the residual for the VS procedure be

$$e(\xi) := \phi^0(\xi) - \phi^0_{k-1}(\xi).$$

Then, for $\forall v \in V_h$,

$$a(e(\xi), v; \xi) = -a(\bar{\phi}, v; \xi) - a(\phi^0_{k-1}(\xi), v; \xi).$$

Let $r(v; \xi) \in V_h^*$ be the residual

$$r(v; \xi) = \begin{cases} -a(\bar{\phi}, v; \xi), & k = 1, \\ -a(\bar{\phi}, v; \xi) - a(\phi^0_{k-1}(\xi), v; \xi), & k \geq 2. \end{cases} \quad (3.6)$$

We have

$$a(e(\xi), v; \xi) = r(v; \xi), \quad \forall v \in V_h. \quad (3.7)$$

Let $\Xi$ be a small sample set taken from $\mathcal{N}(0, \Sigma)$. At step $k$, we choose

$$\xi^k := \begin{cases} \text{randomly chosen from } \Xi, & k = 1, \\ \arg \max_{\xi \in \Xi} \| \hat{e}(\xi) \|^2_V, & k \geq 2. \end{cases}$$
where the error indicator $\|\hat{e}(\xi)\|_{V}^{2}$ has the form

$$
\|\hat{e}(\xi)\|_{V}^{2} = m_{a} \sum_{p=1}^{m_{a}} \sum_{p'=1}^{m_{a}} \kappa_{p}(\xi)\kappa_{p'}(\xi)(\mathcal{L}_{p}, \mathcal{L}_{p'})_{V} + \sum_{q=1}^{k-1} \eta_{q}(\xi) \sum_{p=1}^{m_{a}} \kappa_{p}(\xi) \times \left\{ 2 \sum_{p'=1}^{m_{a}} \kappa_{p'}(\xi)(\mathcal{L}_{p}, \mathcal{L}_{p'})_{V} + \sum_{q'=1}^{k-1} \eta_{q'}(\xi) \sum_{p'=1}^{m_{a}} \kappa_{p'}(\xi)(\mathcal{L}_{q'}, \mathcal{L}_{q'})_{V} \right\}. 
$$

(3.8)

and $\mathcal{L}_{q}$ is the Riesz representation of $a_{p}(\Phi_{q}, v)$, i.e., $(\mathcal{L}_{q}, v)_{V} = -a_{p}(\Phi_{q}, v)$ for any $v \in \mathcal{V}$ and $1 \leq p \leq m_{a}$, $1 \leq q \leq k-1$. Similarly, $\mathcal{L}_{q'}$ is the Riesz representation of $a_{p}(\tilde{\phi}, v)$, for any $v \in \mathcal{V}$ and $1 \leq p \leq m_{a}$. The derivation and calculation of $\|\hat{e}(\xi)\|_{V}^{2}$ is similar to the one in paper [20].

Let $e(\xi) = e_{h}(x)e_{\xi}(\xi)$, and $e_{h}(x)$ be the solution of Eq. (3.7) with $\xi = \xi^{k}$. If we take $\Phi_{k}(x) = e_{h}(x)$ in Eq. (3.5), then

$$
e_{\xi}(\xi) \sum_{p=1}^{m_{a}} \kappa_{p}(\xi) a_{p}(e_{h}, v) = r(v; \xi).
$$

(3.9)

If we take $v = e_{h}$ in Eq. (3.5), then it follows that

$$
e_{\xi}(\xi) = - \sum_{p=1}^{m_{a}} \kappa_{p}(\xi) a_{p}(\tilde{\phi}, e_{h}) - \sum_{q=1}^{k-1} \eta_{q}(\xi) \sum_{p=1}^{m_{a}} a_{p}(\Phi_{q}, e_{h}) \sum_{p=1}^{m_{a}} \kappa_{p}(\xi) a_{p}(e_{h}, v).
$$

(3.10)

We take $\eta_{k}(\xi) = e_{\xi}(\xi)$ in Eq. (3.5). As can be seen from Eq. (3.10), $\eta_{k}(\xi)$ depends on $\{\eta_{q}(\xi)\}_{q=1}^{k-1}$ computed previously while they are stochastic functions with respect to $\xi$. Let $b^{k} := [a_{1}(\tilde{\phi}, \phi_{1}^{0}), \ldots, a_{m_{a}}(\tilde{\phi}, \phi_{k}^{0})]$, $\kappa_{\xi}(\xi) := [\kappa_{1}(\xi), \ldots, \kappa_{m_{a}}(\xi)]^{T}$, and $\eta^{k}(\xi) := [\eta_{1}(\xi), \ldots, \eta_{k-1}(\xi)]$,

$$
A^{k} := \begin{bmatrix}
a_{1}(\phi_{1}^{0}, \phi_{1}^{0}) & \cdots & a_{m_{a}}(\phi_{1}^{0}, \phi_{k}^{0}) \\
\vdots & & \vdots \\
a_{1}(\phi_{k}^{0}, \phi_{k}^{0}) & \cdots & a_{m_{a}}(\phi_{k}^{0}, \phi_{k}^{0})
\end{bmatrix}.
$$

Then the matrix form of $\eta_{k}(\xi)$ can be written as

$$
\eta_{k}(\xi) = \frac{b^{k} \kappa_{\xi}(\xi) - \eta^{k}(\xi)A^{k}_{1} \kappa_{\xi}(\xi)}{b^{k} \kappa_{\xi}(\xi)},
$$

where $A^{k}_{1}$ is the first $k-1$ rows of matrix $A^{k}$, $a^{k}$ is the last row of matrix $A^{k}$. Because all the inner product in physical space can be computed and saved at each step, the simulation for $\xi$ is efficient due to the affine form.

3.3. Ensemble-based VS method. As can be seen, multiple inhomogeneous Dirichlet boundary condition problems are required to be determined on each coarse neighborhood $\omega_{i}$. It will demand significantly computational cost if we approximate each solution $\phi_{j}^{i}(x, \xi)$ for $j = 1, \ldots, M_{i}$ applying the VS method. Inspired by the idea of ensemble-based method used in [15], we construct variable separated forms of $\phi_{j}^{i}(x, \xi)$ for different $j$ on each coarse neighborhood, such that they can share the same interpolate rule $\{\eta_{q}^{i}(\xi)\}_{q=1}^{N}$. 
By summing up the Eq. (3.10) for different $j$, we can obtain the local problem

\[
\begin{aligned}
&-\text{div}(\kappa(x; \xi) \nabla (\bar{\phi}^i)) = 0, \quad \text{in } \omega_i, \\
&\langle \bar{\phi}^i \rangle = \langle \tilde{\phi}^i(x) \rangle, \quad \text{on } \partial \omega_i,
\end{aligned}
\] (3.11)

where $\langle \tilde{\phi}^i(x) \rangle = \sum_{j=1}^{M_i} \bar{\phi}_j^i$, and $\langle \phi^i(x) \rangle = \sum_{j=1}^{M_i} \phi_j^i, M_i$ is the number of local basis functions selected in the coarse neighborhood $\omega_i$. $\langle \phi^i(x) \rangle$ is treated as the solution to the new Eq. (3.11), of which the inhomogeneous Dirichlet boundary condition is $\langle \bar{\phi}^i(x) \rangle$. Thus we have the solution

$$
\langle \phi^i(x) \rangle = \langle \phi^i(x) \rangle_N^0 + \langle \tilde{\phi}^i(x) \rangle,
$$

where $\langle \phi^i(x) \rangle_N^0$ has the variable separated approximation

$$
\langle \phi^i(x) \rangle_N^0 = \sum_{q=1}^{N} \Phi_q^{\omega_i}(x) \eta_q^{\omega_i}(\xi).
$$

This can be sought by running the VS procedure described in Section 3.2. Here the superscript 0 refers to the zero boundary condition, and the subscript $N$ represents the number of truncations terms.

At the step $k$ ($k \geq 2$) of VS iteration, we solve the equation analogous to Eq. (3.7) with $\xi = \xi^k$, i.e.,

$$
a(\Phi_k^{\omega_i}(x), v; \xi^k) = -a(\langle \bar{\phi}^i(x) \rangle, v; \xi^k) - \sum_{q=1}^{k-1} \eta_q^{\omega_i}(\xi^k) a(\Phi_q^{\omega_i}, v), \quad \forall v \in \mathcal{V}_h,
$$

to obtain the $k$-th physical basis $\Phi_k^{\omega_i}(x)$. For $q = 1, \ldots, k$, let

$$
\Phi_q^{\omega_i}(x) = \sum_{j=1}^{M_i} \Phi_{j,q}^{\omega_i}(x).
$$

Then the equation above can be rewritten as

$$
\sum_{j=1}^{M_i} a(\Phi_{j,k}^{\omega_i}(x), v; \xi^k) = -\sum_{j=1}^{M_i} a(\bar{\phi}_j^i(x), v; \xi^k) - \sum_{q=1}^{k-1} \eta_q^{\omega_i}(\xi^k) \sum_{j=1}^{M_i} a(\Phi_{j,q}^{\omega_i}(x), v), \quad \forall v \in \mathcal{V}_h.
$$

We decompose Eq. (3.12) to be the system

$$
\begin{aligned}
a(\Phi_{1,k}^{\omega_i}(x), v; \xi^k) &= -a(\bar{\phi}_1^i(x), v; \xi^k) - \sum_{q=1}^{k-1} \eta_q^{\omega_i}(\xi^k) a(\Phi_{1,q}^{\omega_i}(x), v), \\
&\vdots \\
a(\Phi_{M_i,k}^{\omega_i}(x), v; \xi^k) &= -a(\bar{\phi}_{M_i}^i(x), v; \xi^k) - \sum_{q=1}^{k-1} \eta_q^{\omega_i}(\xi^k) a(\Phi_{M_i,q}^{\omega_i}(x), v).
\end{aligned}
$$

We note that Eq. (3.11) is one of the decompositions. It seems like the residual equation of problem (3.3) for different $j$ in the $k$-th step of VS method, but the samples $\{\xi^q\}_{q=1}^k$ and interpolate rule $\{\eta_q^{\omega_i}(\xi)\}_{q=1}^{k-1}$ has been determined based on problem
Let $\Phi_{j,k}^{\omega_i}(x)$ be the $k$-th physical basis of the corresponding variable separated form of the $j$-th local problem. We then obtain the separated form for each $\phi_j^i(x;\xi)$

$$\phi_j^i(x;\xi) = \sum_{q=1}^{N} \Phi_{j,q}^{\omega_i}(x)\eta_q^i(\xi), \quad 1 \leq j \leq M_i, \quad x \in \omega_i,$$

which is considered as approximations of $\phi_j^i(x;\xi)$ satisfying Eq. (3.14) for $x \in \omega_i$.

**Remark 3.1.** The ensemble-based method we used here leads to a rough approximation of solutions to (3.3). We use the VS method to approximate the local problems with average boundary condition $\langle \hat{\phi}(x) \rangle$, and then derive the separated forms (mainly the physical parts) for each $\phi_j^i(x;\xi)$ through decomposing the residual equation.

**Remark 3.2.** The approximation may not be as good as running VS for each $\phi_j^i(x;\xi)$, but the compromised scheme makes them share the same interpolate rule. The derived expression (3.14) is an approximation of $\phi_j^i(x;\xi)$ in some sense. As $\Phi_{j,k}^{\omega_i}(x)$ are fixed, the use of a common interpolate rule significantly improves the computation efficiency.

**Remark 3.3.** We can obtain the physical basis functions for each $j$ along with the interpolate rule $\{\eta_q^i(\xi)\}_{q=1}^{N}$. Or as an alternative, we run VS process for problem (3.11) to obtain the selected samples $\{\xi_q\}_{q=1}^{N}$ and interpolate rule $\{\eta_q^i(\xi)\}_{q=1}^{N}$. Then calculate the physical basis by the similar technique used in paper [25], for any $j = 1, \cdots, M_i$.

We can finally multiply the physical basis functions with the partition functions $\{\chi_i(x)\}_{i=1}^{N_M}$, and then the stochastic multiscale basis functions are presented as

$$\phi_j^i(x;\xi) = \sum_{q=1}^{N} \chi_i(x)\Phi_{j,q}^{\omega_i}(x)\eta_q^i(\xi), \quad 1 \leq j \leq M_i, \quad x \in \omega_i.$$

In particular, we set $\phi_j^i(x;\xi) = \chi_i(x)\hat{\phi}_j^{i}(x)$ on $\partial\omega_i$. The physical part $\chi_i(x)\Phi_{j,k}^{\omega_i}(x)$ can be calculated and stored for any random sample $\xi$. Thus the stochastic multiscale basis functions can be efficiently obtained. Algorithm II describes the procedure of ensemble-based VS-Ms method.

The random samples of $\xi$ are taken from the prior (2.4), where the covariance matrix is determined by the relative distance of the spatial points. Due to the property of the Gaussian process, the parts of $\xi$ taken from the same local domain structures obey the same Gaussian distribution. In this paper, $\hat{\xi}$ is chosen as the mean of the random variable $\xi$, the local boundary conditions $\hat{\phi}_j^i(x)$ for $j = 1, \cdots, M_i$ are the same for the same types of local structures. Hence, we only need to use the proposed method to obtain stochastic multiscale basis functions for 4 types of local domain structures (as shown in red box of Figure 5.1). We apply the corresponding 4 types of expressions to the whole domain $D$.

4. **Numerical examples.** In this section, we consider the discontinuous field identification problems for a single-phase Darcy flow problem. The flow is described in terms of fluid pressure $u(x,t)$ which is governed by the following equation

$$c \frac{\partial u}{\partial t} - \text{div} \left( \kappa(x) \nabla u \right) = f(x,t), \quad x \in D, t \in (0,T],$$

where the studied domain is $D = [0,1] \times [0,1]$, and the initial condition is set as $u(x,0) = 0$. $\kappa(x)$ is the unknown permeability field. $c$ is the specific storage, which is
Algorithm 1 Ensemble-based VS-Ms method

1. Take an arbitrary realization \( \log \kappa(x; \xi) \) from the prior (2.4), and generate the effective boundary conditions \( \phi^j_i(x) \) for \( j = 1, \cdots, M \), by GMsFEM in coarse neighborhoods;

2. Run Ensemble-based VS procedure for local problems (3.11), to obtain the interpolate rules \( \{ \eta^q_{\omega_i}(\xi) \}_{q=1}^N \);

3. In each coarse block, reverse the physic basis for \( \phi^j_i(x; \xi) \) and obtain the corresponding separated form

\[
\phi^j_i(x; \xi) = \begin{cases} 
\sum_{q=1}^N \Phi_{j,q}^\omega_i(x) \eta^q_{\omega_i}(\xi), & x \in \omega_i, \\
\phi^j_i(x), & x \in \partial \omega_i.
\end{cases}
\]

4. Multiply the physical basis with the union partition functions, and obtain the stochastic multiscale basis functions as

\[
\phi^j_i(x; \xi) = \begin{cases} 
\sum_{q=1}^N \chi_i(x) \Phi_{j,q}^\omega_i(x) \eta^q_{\omega_i}(\xi), & x \in \omega_i, \\
\chi_i(x) \phi^j_i(x), & x \in \partial \omega_i.
\end{cases}
\]

5. For any random realization taken from the prior (2.4), calculate the multiscale basis functions efficiently and simulate the model.

taken to be constant 1 for simplicity. The source term \( f(x, t) \) is defined by a weighted Gaussian plume with standard deviation 0.1, centered at \((0.55, 0.4)\), with weight 1, which keeps constant with respect to the time \( t \). The log permeability is parameterized as a Gaussian random field, the parameters in the Gaussian kernel \( C(x_1, x_2) \) are set as

\[
\sigma_x^2 = 0.1, \quad l_{x_1} = 0.07, \quad l_{x_2} = 0.07.
\]

Eq. (2.5) is treated as the affine form of the field \( \log \kappa(x; \xi) \), to characterize the unknown random field accurately. Then we have the required affine of the bilinear form

\[
a(w, v; \xi) = \sum_{p=1}^{m_a} \xi^p a_p(w, v), \quad \forall v, w \in \mathcal{V}, \quad \forall \xi \sim \mathcal{N}(0, \Sigma),
\]

where

\[
a_p(w, v) = \int_D \mathbf{1}_{\{x \in D_i\}} \nabla w \nabla v dx.
\]

Note that there are only a fixed number of nonzeros in the corresponding stiffness matrix for each \( a_p(w, v) \), we use the proposed method on 4 types of local coarse neighborhood, the dimension \( m_a \) would not be too large to slow down the computation. For the simplicity of notation, let \( m(x; \xi) = \log \kappa(x; \xi) \). The backward Euler scheme is used for the temporal discretization. Some model simulations are presented in the first subsection to demonstrate the efficiency of the scheme (3.3) and our proposed method in solving the forward model. We also apply the proposed model reduction method to the inverse problems to show the performance of the method in identifying discontinuous fields.
4.1. Reduced model approximation. We study the problem (4.1) with no-flow boundary conditions. The end time is $T = 0.2$, and the time step is set as $\Delta t = 0.001$. The solution obtained by finite element method is considered the reference one. We solve it on a uniform $80 \times 80$ fine grid, and we set the coarse grid $8 \times 8$ for multiscale finite element method.

The local boundary conditions in problem (3.3) are generated by setting $m(x; \bar{\xi})$ to be the mean value of the random field. The ensemble-based VS-Ms method is used to obtain the variable separated form of multiscale basis functions, where 4 local stochastic problems are solved on 4 coarse neighborhood as shown in Figure (3.1). The parameter dimension $m_\alpha$ on the coarse neighborhood I, II, III, IV is 100, 200, 200, and 400, respectively.

We show the effects of the ensemble-based VS-Ms method in solving the problem (4.1) numerically. Let $u_h$ be the reference solution using the finite element method on fine grid, $u_H$ the solution obtained by GMsFEM. $u_{\bar{m}}$ represents the solution obtained by the multiscale method, where the reduced snapshot functions are obtained by solving local problems (3.3), $u_{\langle \bar{m} \rangle}$ is the approximation resulted from the ensemble-based VS-Ms method. Let $\varepsilon^n_s = u^n_h - u^n_s$, $s = H, \bar{m}, \langle \bar{m} \rangle$. Then the $L^\infty$-error and $L^2$-error at $n\Delta t$ are defined as

$$\|\varepsilon^n_s\|_\infty = \max_{\xi \in \Omega} \left\{ \max_{x \in D} |u^n_h(x; \xi) - u^n_s(x; \xi)| \right\},$$

$$\|\varepsilon^n_s\| = \sqrt{\int_{\Omega} \int_D (u^n_h(x; \xi) - u^n_s(x; \xi))^2 dx d\pi_0}, \ s = H, \bar{m}, \langle \bar{m} \rangle,$$

where $\pi_0$ is the Gaussian prior $\mathcal{N}(0, \Sigma)$. We compare the maximum $L^\infty$-error and the maximum $L^2$-error for the different strategies, and define

$$\varepsilon^\infty_s = \max_{0 \leq n \leq n_T} \|\varepsilon^n_s\|_\infty,$$

$$\varepsilon^2_s = \max_{0 \leq n \leq n_T} \|\varepsilon^n_s\|,$$

to show the corresponding performance, where $T = n_T \Delta t$.

For $s = H, \bar{m}, \langle \bar{m} \rangle$, $\tau_s$ represents the time cost in constructing the multiscale basis functions for an arbitrary realization of $\xi$. The numerical results are shown in Table 4.1. 81 local problems are required to be solved for all the methods displayed above. As for GMsFEM, the effective boundary conditions are obtained by solving eigenvalue problems (3.3), which depends on the stochastic random field. Solving local problems (3.3) directly or using the ensemble-based VS-Ms method cost much less time, since the local boundary conditions are fixed and can be stored overhead. The separated form provided by the ensemble-based VS-Ms method significantly accelerates the speed of calculating multiscale basis functions.

As can be seen from the table, both the two types of errors decrease as the number of multiscale basis functions increase, for the studied strategies. Under the prior condition, the scheme (3.3) we adopt in generating the reduced snapshot space works well, it reaches almost the same magnitude accuracy as GMsFEM. The accuracy of the PDE is preserved when using the ensemble-based VS-Ms method. Due to the constructed explicit expressions, the basis functions can be efficiently computed. Hence it costs the least time in constructing local basis functions.
When the number of multiscale basis functions is fixed by 3, we solve the problem \((4.1)\) by varying \(N = 5, 15, 25, 35\), to study the effect of the number of separated terms. It can be shown from Table 4.2 that, the trend of the two types of errors is flat against the number of separated terms, which is resulted from the high efficiency of the VS method with small separated terms. The number of multiscale basis functions dominates the effects on the accuracy for our proposed method. Hence, we will concentrate on the influence of the number of multiscale basis functions on the posterior density approximation in the next section.

**Remark 4.1.** Though the problem can be solved well with even small separated terms, the moving of samplers in the Markov chain leads them away from the prior, we set \(N\) not be too small to make the proposed method stable and still work in Markov chain sampling.

### 4.2. Posterior densities approximation

In this section, we use the proposed method to approximate the likelihood, along with the imposed TG hybrid prior to identify discontinuous fields. We construct the variable separated expression for the multiscale basis functions by the ensemble-based VS-Ms method, and then use it to improve the efficiency of MCMC. Measurement data are generated by using finite element method in a fine grid with time step \(\Delta t = 0.001\). For any given values of \(\xi\), we solve the equation using the ensemble-based VS-Ms model reduction method with time step \(\Delta t = 0.002\). The measurement noise is set to be \(\sigma = 0.01\).

#### 4.2.1. Case 1: separated blocks

First we apply the model reduction method to the example, where there are two separated blocks on the target permeability field.
Namely, suppose we have the target piecewise constant permeability

$$
\kappa(x) = \begin{cases} 
  e^2, & \frac{(x_1 - 0.35)^2}{0.15^2} + \frac{(x_2 - 0.3)^2}{0.15^2} \leq 1, \\
  e^2, & \frac{(x_1 - 0.65)^2}{0.15^2} + \frac{(x_2 - 0.6)^2}{0.15^2} \leq 1, \\
  1, & \text{otherwise.}
\end{cases}
$$

i.e., there are two faces to be identified. The boundary condition is set as

$$
u(x, t) = 1.7 - 1.4x_1, \quad x \in \partial D,$$

Fig. 4.1. The empirical quantile-quantile, histogram and empirical cdf of (a) $\xi_{2753}$, (b) $\xi_{6363}$, (c) $\xi_{6367}$ and (d) $\xi_{2731}$. 
for the problem, the forward model is studied during the time interval \([0, 0.15]\). We observe the data starting from \(t = 0.02\), at every 0.01, up to \(t = 0.11\). For each time layer, 49 pressure measurements are uniformly collected in \([0.05, 0.95] \times [0.05, 0.95]\).

The governing equation in (4.1) is solved on a uniform \(80 \times 80\) fine grid, resulting in 6400 unknowns to be estimated. The coarse grid is set to be \(8 \times 8\), the effective local boundary conditions are generated by setting \(m(x; \tilde{\xi}) = 0\), we select 3 basis functions on each coarse neighborhood, 20 separated terms are kept in the variable separation expressions. The posterior samplers are drawn with \(\lambda = 300\), \(\beta = 0.015\), resulting in the acceptance rate 38.55\%. We draw \(10^5\) samplers to ensure the reliability of the inference, and remain the last 50\% for statistics computation.

We display the empirical quantile-quantile, histogram and empirical cumulative distribution function (cdf) of posterior samplers to show the non-Gaussian of the posterior density. The plots of \(\xi_{2753}\), \(\xi_{6363}\), \(\xi_{6367}\) and \(\xi_{2731}\) are shown in Figure 4.1. The quantiles illustrate that the distributions differ from Gaussian significantly. As can be seen from the histogram, the distributions of \(\xi_{2753}\) and \(\xi_{2731}\) are right-skewed and left-skewed, respectively. For \(\xi_{2731}\), the cdf curve reaches 0.5 fast, but the interval gets wider when it reaches the rest 0.5. \(\xi_{6363}\) and \(\xi_{6367}\) seem to be multi-modal, which obviously are non-Gaussian.

The true profile, posterior mean and posterior standard deviation of the log permeability field are shown in Figure 4.2. The two blocks are separately identified, no adjoin between them exists as expected. Both the values of the two blocks are underestimated, the block near the lower left corner is estimated better. It may be resulted by the higher values condition imposed at the left boundary when solving the forward problem.

We plot the surface from the view of plane \(y = 0.3\), \(y = 0.6\), \(x = 0.35\), and \(x = 0.65\) in Figure 4.3, the mean and 95\% credible intervals are presented as well. The shapes of the curves imply that the background and jumps are recognized well, and the value of the log permeability is very closed to the true one for the block near the lower left corner. Though the jumps estimated are not as sharp as the reference ones, there are no oscillations near them. It can also be reflected from Figure 4.4 where the spatial gradient of the true and posterior mean of the log permeability field are presented. Compared with the reference face boundaries, the estimated ones are wider because of the less sharp estimated jumps. The face boundary near the upper right corner is darker since the severe underestimated values at this block. However, the locations of the face boundaries are separately identified generally.

The 95\% credible intervals and prediction intervals for the pressure at \(u(x =
Bayesian discontinuous fields identification

0.35, y, t = 0.05), u(x, y = 0.65, t = 0.05) and u(x = 0.35, y = 0.5, t) are constructed. As illustrated in Figure 4.3, the reference values are mostly located in the credible intervals, measurements are almost contained in the predictive intervals. Due to the deterministic Dirichlet boundary condition on \( \partial D \), both the credible and prediction intervals become tight as \( y \) and \( x \) gets closer to the end points for \( u(x = 0.35, y, t = 0.05) \) and \( u(x, y = 0.65, t = 0.05) \). For \( u(x = 0.35, y = 0.5, t) \), as the uncertainty from the input propagates, the uncertainty associated with the model fit and predictions grows, which naturally leads to loose intervals as time moves on.

Figure 4.4 shows convergence of the forward model and posterior approximation with respect to the number of multiscale basis functions selected in the coarse neighborhood. We plot the \( L_2 \) error in the forward model, \( \| G - G_N \|_{L^2_{\nu_0}}^2 \), and the

**Fig. 4.3.** The true profile, posterior mean and 95% credible estimates of the log permeability at the surface (a) \( y = 0.3 \) (left) and \( x = 0.35 \) (right), (b) \( y = 0.6 \) (left) and \( x = 0.65 \) (right).

**Fig. 4.4.** Spatial gradient of (a) true and (b) posterior mean of the log permeability field.
Fig. 4.5. Data, point estimates, and 95% credible and prediction intervals produced by the Bayesian analysis for \( u(x = 0.35, y, t = 0.05) \), \( u(x, y = 0.65, t = 0.05) \) and \( u(x = 0.35, y = 0.5, t) \).

Fig. 4.6. Approximation of the forward model and the posterior density with respect to the number of multiscale basis functions. Dashed line represents the Kullback-Leibler divergence \( D_{KL}(\pi_d^N || \pi^d) \) (denoted by \( D_{KL} \)); solid line shows the \( L^2_{\pi_0} \) error \( \| G - G_N^N \|^2_{L^2_{\pi_0}} \) (denoted by \( E_{L^2} \)).

Kullback-Leibler divergence of the exact posterior from the approximate posterior, \( D_{KL}(\pi_d^N || \pi^d) \). Both the errors \( D_{KL} \) and \( E_{L^2} \) decrease exponentially as the number of multiscale basis functions increase. The convergence rate of the Kullback-Leibler divergence is basically the same as that of the \( E_{L^2} \) error in the forward model. This confirms Theorem 2.1.

**4.2.2. Case 2: nested blocks.** As the last example, we consider recovering a discontinuous permeability with nested blocks structure. The problem is studied with boundary condition

\[
u(x, t) = 2 - 2x_1, \quad x \in \partial D,
\]

and the end time \( T = 0.1 \). The reference log permeability field and measurement locations are shown as Figure 4.7 where the two blocks are nested with different permeability value. We observe data in the time interval \([0.02, 0.1]\) with the time step 0.01.
In this subsection, we solve the problem on a uniform $100 \times 100$ fine grid, which results in $10^4$ unknowns. For the sake of using the explicit variable separation expression of the multiscale basis functions constructed before, we set the coarse grid as $10 \times 10$. Then the structure of local covariance is the same, but the relative distance becomes smaller. Correspondingly, we put the influence of the change into the length scale in the covariance kernel, and modify the length scale to be $0.8 \times l_{x_1}$ and $0.8 \times l_{x_2}$ in this example. This will result in the same Gaussian prior as the two examples presented above. It is convenient to use the constructed variable separated expression but add some basis index and rearrange the corresponding basis functions.

The Markov chain is run by the proposed method, where the number of basis functions is 3. We draw $1.4 \times 10^5$ posterior samplers with $\lambda = 250$, $\beta = 0.012$, and keep the last $5 \times 10^4$ ones for statistical inference, the resulted acceptance rate is 38.80%. Figure 4.8 displays the maximum a posterior (MAP) estimate, posterior mean and posterior standard deviation of the log permeability field. The uncertainty is small overall, the MAP reflects the concave part of the heart shaped block better than the posterior mean, the lower half of the heart shaped block are well embodied by both the two estimates. Besides the background, the circle block is also well identified, whose value is between the background and the heart shaped block.

We present the spatial gradient of the true and posterior mean of the log permeability field in Figure 4.9. The face boundary for the nested block is brighter than the circle’s. This implies the spatial gradient value is larger, which matches the spatial gradient of the true log permeability. The histogram of the reference, 0.25 quantile, mean and 0.95 quantile estimate of the log permeability are demonstrated in Figure 4.10. The reference has a trinomial distribution in the histogram, the high frequency value concentrates around 0 and 1 for the estimates, the stack around 4 or 3.5 is less
prominent due to the underestimation of the permeability value.

In Figure 4.11, we plot the 95% credible and predictive intervals for the model response at \( u(x = 0.35, y = 0.05, t = 0.05) \), \( u(x, y = 0.65, t = 0.05) \) and \( u(x = 0.35, y = 0.5, t) \). Most part of the reference values lie in the credible intervals, apart from one or two points, measurements are located in the predictive intervals. Similar to the response studied in Section 4.2.1 for \( u(x = 0.35, y = 0.05, t = 0.05) \) and \( u(x, y = 0.65, t = 0.05) \), the uncertainty of the model response mainly concentrates in the domain, and gets smaller as the response distributes closer to \( \partial D \). The accumulated propagation uncertainty results in loose intervals for \( u(x = 0.35, y = 0.5, t) \).

The Kullback-Leibler divergence and the forward model \( L_2 \) norm error are plotted against the number of multiscale basis functions in Figure 4.12. We observe the same trend for all curves, the errors decay as the number of multiscale basis functions increase. Moreover, the curve of the figure confirms that the posterior density of Kullback-Leibler divergence \( D_{KL} \) decreases with the same speed as \( \| G - G_N \|_{L_2}^2 \), and the error \( D_{KL} \) also can be bounded in terms of the error \( E_{L_2} \) in the forward model.

5. Conclusion. This paper presented the ensemble-based VS-Ms model reduction method with application in Bayesian inverse problems for discontinuous field
identification. The stochastic multiscale basis functions are approximated by the ensemble-based VS method, with respect to the reference Gaussian measure. The adopted ensemble scheme and residual decomposition strategy lead to a common interpolate rule for local functions on the same coarse neighborhood, which improves the efficiency of basis calculation. The way of generating local problems and the relative distance characterized covariance structure simplify the basis construction problem to 4 local problems, coarse neighborhood with the same partition of grid share the same variable separated form. With the same local partitions, only some basis index and rearrange the corresponding basis functions are needed for refinement of the fine grid partition. The influence of changing the relative distance is considered as length scale modification. The presented approach leads to an accurate approximation of the full forward model while save much computation cost, due to the efficient multiscale basis functions construction and the degree of freedom reduction. Rigorous analysis is carried out for the approximation of model reduced method in the Bayesian inverse problem with the TG prior. The numerical examples confirm that the approximated
posterior matches the reference one better and better as the enrichment of multiscale basis functions. We demonstrate our approach in the single-phase inversion problems, where the discontinuous permeability with different structures are identified. The constructed explicit expression of the basis functions can be applied to problems with different right-hand conditions, which is inherited from the GMsFEM, while the computational time is less than the GMsFEM.

Appendix A. Proof of Theorem 2.1

**Lemma A.1.** (Cotter, Dashti, and Stuart [27]) Let \( G \) satisfy Assumption 2.1 and define
\[
Z(\xi; d) = \frac{\|d - G(\xi)\|^2}{2\sigma^2}.
\]
Then \( Z \) satisfy the following: for every \( r > 0 \), there is \( K_G = K_G(r) \) such that for all \( \xi \in \mathbb{R}^n \) and \( d \in \mathbb{R}^n \) with \( \max(\|\xi\|_\Sigma, \|d\|_2) < r \),
\[
Z(\xi; d) \leq K_G(r).
\]

**Lemma A.2.** Let \( G: \mathbb{R}^n \rightarrow \mathbb{R}^n \) be a function that satisfy Assumption 2.1. Then there exists a \( C_1 > 0 \) such that for any sequence \( G_N \) of functions that converge to \( G \) for \( N \rightarrow \infty \),
\[
\|Z - Z_N\|_{L^2(\Sigma)}^2 \leq C_1\|G(\xi) - G_N(\xi)\|_{L^2(\Sigma)}^2,
\]
where
\[
Z_N(\xi; d) = \frac{\|d - G_N(\xi)\|^2}{2\sigma^2}.
\]
The Lemma A.2 can be proved following the method of [29]. Our proof of Theorem 2.1 will also closely follow the method of [29].

**Proof of Theorem 2.1** We first define the normalizing constants \( \gamma \) and \( \gamma_N \) as
\[
\gamma = \int L(\xi)\pi_0\pi_1 d\xi,
\]
\[
\gamma_N = \int L_N(\xi)\pi_0\pi_1 d\xi,
\]
where
\[
L(\xi) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp(-Z(\xi; d)),
\]
and
\[
L_N(\xi) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp(-Z_N(\xi; d)),
\]
\( \pi_1(\xi) \) is the total variational prior
\[
\pi_1(\xi) \propto \exp(-R(\xi)).
\]
It is obvious from Lemma A.1 and Assumption A.2 that for every $r > 0$, when $\|d\|_2 < r$
\[
\gamma \geq (2\pi\sigma^2)^{-\frac{d}{2}} \exp(-K_G(r) - K(r)),
\]
We have an analogous lower bound for $\gamma_N$.

Similar to the derivation in [29], we have
\[
D_{KL}(\pi^d_N||\pi^d) \leq D_{KL}(\pi^d_N||\pi^d) + D_{KL}(\pi^d||\pi^d_N)
\]
\[
\leq \frac{|\gamma - \gamma_N|}{\gamma_N} \int L|Z - Z_N|\pi_0\pi_1 d\xi + \frac{1}{\gamma_N} \int |L_N - L||Z - Z_N|\pi_0\pi_1 d\xi
\]
\[
:= I_1 + I_2,
\]
where
\[
I_1 = \frac{|\gamma - \gamma_N|}{\gamma_N} \int L|Z - Z_N|\pi_0\pi_1 d\xi,
\]
and
\[
I_2 = \frac{1}{\gamma_N} \int |L_N - L||Z - Z_N|\pi_0\pi_1 d\xi.
\]
For the first term, due to the definition of $Z(\xi; d)$ and $R(\xi)$, they are non-negative, we have
\[
\int L|Z - Z_N|\pi_0\pi_1 d\xi \leq (2\pi\sigma^2)^{-\frac{d}{2}} \|Z - Z_N\|_{L^2_0}.
\]
Using the inequality $|e^{-x} - e^{-y}| \leq |x - y|$ for $x > 0, y > 0$, and the non-negative $R(\xi)$, we have
\[
|\gamma - \gamma_N| \leq (2\pi\sigma^2)^{-\frac{d}{2}} \int |Z - Z_N|\pi_0 d\xi,
\]
\[
\leq (2\pi\sigma^2)^{-\frac{d}{2}} \|Z - Z_N\|_{L^2_0},
\]
where the Holder’s inequality has been used to estimate the inequality. Combine the above results, we have
\[
I_1 \leq \exp(2K_G(r) + 2K(r))\|Z - Z_N\|_{L^2_0}^2,
\]
\[
= K_1(r)\|Z - Z_N\|_{L^2_0}^2.
\]
We note that
\[
|L - L_N| \leq (2\pi\sigma^2)^{-\frac{d}{2}} |Z - Z_N|,
\]
therefore,
\[
I_2 \leq \exp(K_G(r) + K(r))\|Z - Z_N\|_{L^2_0}^2,
\]
\[
= K_2(r)\|Z - Z_N\|_{L^2_0}^2.
\]
Note that the bounds on $I_1$ and $I_2$ from above are independent of $N$, we have
\[
D_{KL}(\pi^d_N||\pi^d) \leq (K_1(r) + K_2(r))\|Z - Z_N\|_{L^2_0}^2,
\]
\[
\leq (K_1(r) + K_2(r))C_1\|G(\xi) - G_N(\xi)\|_{L^2_0}^2,
\]
\[
= C\|G(\xi) - G_N(\xi)\|_{L^2_0}^2.
\]
REFERENCES

[1] I. Babuška, U. Banerjee, and J. E. Osborn, Survey of meshless and generalized finite element methods: a unified approach, Acta Numerica, (2003), pp. 1–125.

[2] L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, Y. Marzouk, L. Tenorio, B. V. B. Waanders, and K. Willcox, Surrogate and Reduced-Order Modeling: A Comparison of Approaches for Large-Scale Statistical Inverse Problems, John Wiley & Sons, Ltd, 2010.

[3] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, Handbook of markov chain monte carlo, CRC press, 2011.

[4] T. Bu-Thanh, O. Ghattas, J. Martin, and G. Stadler, A computational framework for infinite-dimensional bayesian inverse problems part i: The linearized case, with application to global seismic inversion, SIAM Journal on Scientific Computing, 35 (2013), pp. A2494–A2523.

[5] M. Burger, A level set method for inverse problems, Inverse problems, 17 (2001), p. 1327.

[6] T. F. Chan and X.-C. Tai, Level set and total variation regularization for elliptic inverse problems with discontinuous coefficients, Journal of Computational Physics, 193 (2004), pp. 40–66.

[7] J. A. Christen and C. Fox, Markov chain monte carlo using an approximation, Journal of Computational and Graphical statistics, 14 (2005), pp. 795–810.

[8] E. T. Chung, T. F. Chan, and X.-C. Tai, Electrical impedance tomography using level set representation and total variational regularization, Journal of Computational Physics, 205 (2005), pp. 357–372.

[9] S. L. Cotter, G. O. Roberts, A. M. Stuart, D. White, et al., Mcmc methods for functions: modifying old algorithms to make them faster, Statistical Science, 28 (2013), pp. 424–446.

[10] T. Cui, K. J. Law, and Y. M. Marzouk, Dimension-independent likelihood-informed mcmc, Journal of Computational Physics, 304 (2016), pp. 109–137.

[11] M. M. Dunlop, M. A. Iglesias, and A. M. Stuart, Hierarchical bayesian level set inversion, Statistics and Computing, 27 (2017), pp. 1555–1584.

[12] Y. Efendiev, A. Datta-Gupta, V. Ginting, X. Ma, and B. Mallick, An efficient two-stage markov chain monte carlo method for dynamic data integration, Water Resources Research, 41 (2005), p. 12423.

[13] Y. Efendiev, J. Galvis, and T. Y. Hou, Generalized multiscale finite element methods (gms-fem), Journal of Computational Physics, 251 (2013), pp. 116–135.

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

[15] M. Gunzburger, N. Jiang, and M. Schneier, An ensemble-proper orthogonal decomposition method for the nonstationary navier–stokes equations, SIAM Journal on Numerical Analysis, 55 (2017), pp. 286–304.

[16] M. Hinze and S. Volkwein, Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: Error estimates and suboptimal control, in Dimension reduction of large-scale systems, Springer, 2005, pp. 261–306.

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

[18] L. Jiang and N. Ou, Bayesian inference using intermediate distribution based on coarse multiscale model for time fractional diffusion equations, Multiscale Modeling & Simulation, 16 (2018), pp. 327–355.

[19] J. Lee and P. Kitanidis, Bayesian inversion with total variation prior for discrete geologic structure identification, Water Resources Research, 49 (2013), pp. 7658–7669.

[20] Q. Li and L. Jiang, A novel variable-separation method based on sparse and low rank representation for stochastic partial differential equations, SIAM Journal on Scientific Computing, 39 (2017), pp. A2879–A2910.

[21] J. S. Liu, Monte Carlo strategies in scientific computing, Springer Science & Business Media, 2008.

[22] J. Martin, L. C. Wilcox, C. Burstedde, and O. Ghattas, A stochastic newton mcmc method for large-scale statistical inverse problems with application to seismic inversion, SIAM Journal on Scientific Computing, 34 (2012), pp. A1460–A1487.

[23] Y. M. Marzouk and H. N. Naim, Dimensionality reduction and polynomial chaos acceleration of bayesian inference in inverse problems, Journal of Computational Physics, 228 (2009), pp. 1862–1902.

[24] S. Osher and J. A. Sethian, Fronts propagating with curvature-dependent speed: algorithms
Bayesian discontinuous fields identification

based on hamilton-jacobi formulations, Journal of computational physics, 79 (1988), pp. 12–49.

[25] N. OU, L. JIANG, AND G. LIN, Variable-separation method based bi-fidelity model for bayesian inverse problems, submitted to Journal of Computational Physics.

[26] L. I. RUDIN, S. OSHER, AND E. FATEMI, Nonlinear total variation based noise removal algorithms, Physica D: nonlinear phenomena, 60 (1992), pp. 259–268.

[27] A. M. STUART, Inverse problems: a bayesian perspective, Acta Numerica, 19 (2010), pp. 451–559.

[28] C. R. VOGEL, Computational methods for inverse problems, vol. 23, Siam, 2002.

[29] L. YAN AND L. GUO, Stochastic collocation algorithms using l1-minimization for bayesian solution of inverse problems, SIAM Journal on Scientific Computing, 37 (2015), pp. A1410–A1435.

[30] Z. YAO, Z. HU, AND J. LI, A tv-gaussian prior for infinite-dimensional bayesian inverse problems and its numerical implementations, Inverse Problems, 32 (2016), p. 075006.