

## EXPLORING THE LOCALLY LOW DIMENSIONAL STRUCTURE IN SOLVING RANDOM ELLIPTIC PDES\*

THOMAS Y. HOU<sup>†</sup>, QIN LI<sup>‡</sup>, AND PENGCHUAN ZHANG<sup>†</sup>

**Abstract.** We propose a stochastic multiscale finite element method (StoMsFEM) to solve random elliptic partial differential equations with a high stochastic dimension. The key idea is to simultaneously upscale the stochastic solutions in the physical space for all random samples and explore the low stochastic dimensions of the stochastic solution within each local patch. We propose two effective methods for achieving this simultaneous local upscaling. The first method is a high-order interpolation method in the stochastic space that explores the high regularity of the local upscaled quantities with respect to the random variables. The second method is a reduced-order method that explores the low rank property of the multiscale basis functions within each coarse grid patch. Our complexity analysis shows that, compared with the standard FEM on a fine grid, the StoMsFEM can achieve computational savings on the order of  $(H/h)^d / (\log(H/h))^k$ , where  $H/h$  is the ratio between the coarse and the fine grid sizes,  $d$  is the physical dimension, and  $k$  is the local stochastic dimension. Several numerical examples are presented to demonstrate the accuracy and effectiveness of the proposed methods. In the high contrast example, we observe a factor of 2000 speed-up.

**Key words.** uncertainty quantification, random PDEs, MsFEM, polynomial chaos expansion, stochastic collocation

**AMS subject classifications.** 65H35, 65N35, 65N12, 65N15, 65C30

**DOI.** 10.1137/16M1077611

**1. Introduction.** Many problems arising from various physical and engineering applications have multiple scale features and uncertainties. For example, to simulate flow in heterogeneous porous media, the permeability field is often characterized as a multiple-scale random medium. The parametrization of a multiscale random medium requires a large number of random variables, leading to a high dimensional random partial differential equation (PDE), which is challenging to solve numerically. Similarly, in shallow water modeling, the basin topography can contain multiple scales and high dimensional uncertainties. Moreover, these problems are typically solved for many source terms and boundary conditions. These problems can be formulated using an input-output relation, as is typically done in reduced-order modeling. In the case of flow in porous media, the input space consists of the random permeability field, source terms, and/or boundary conditions. The output space depends on the quantities of interest and may consist of the mean of coarse-grid solutions or some other statistical quantities with respect to the solution. In many applications, the dimension of the output space is typically smaller than that of the input space. The main objective of this paper is to design an efficient reduced-order method that takes advantage of the effective low dimensional solution space for problems with multiple scales and large uncertainties.

The direct simulation consists of two steps. First, we generate a large number of

---

\*Received by the editors May 31, 2016; accepted for publication (in revised form) November 22, 2016; published electronically April 6, 2017.

<http://www.siam.org/journals/mms/15-2/M107761.html>

**Funding:** This research was partially supported by NSF grants DMS-1318377 and DMS-1613861.

<sup>†</sup>Applied and Computational Math, Caltech, Pasadena, CA 91125 (hou@cms.caltech.edu, pzzhang@cms.caltech.edu).

<sup>‡</sup>Mathematics Department, UW-Madison, Madison, WI 53705 (qinli@math.wisc.edu).

samples of the random coefficient and numerically solve the corresponding deterministic PDEs. Second, we apply an appropriate stochastic method (e.g., Monte Carlo, stochastic collocation) to compute the statistical quantities of interest. Because of the presence of small scales in the physical space and high dimensional uncertainties in the stochastic space, the direct simulations of these problems are prohibitively expensive. We need to develop an efficient model reduction method by obtaining a low dimensional parametrization of the solution in both the physical space and the stochastic space. In this paper, when we refer to “stochastic space” we mean the space of the parametrized random variables, and “stochastic dimension” means the number of the parametrized random variables.

There are a number of multiscale methods that use multiscale bases to represent the multiscale solutions in the physical space; see, e.g., [28, 30, 51, 45, 1, 46, 17, 47, 41, 10, 44]. Naive application of these multiscale methods to each sample of a multiscale random PDE provides little computational savings, because a low dimensional representation needs to be recomputed for every sample. There have been many research activities that explore the low dimensional representation of the solutions of the corresponding random PDEs in the stochastic space. In particular, the generalized polynomial chaos (gPC) methods [23, 54, 3, 19, 2, 43, 42, 53] have received a lot of attention in the last decade. These methods are very effective when the stochastic dimension is small. However, their performance deteriorates dramatically when the stochastic dimension increases due to the *curse of dimensionality* [12, 13].

It is important to point out that for the problems with high stochastic dimensions their solutions typically have multiple scales in the spatial domain. For example, a random permeability field with a short correlation length has a high stochastic dimension, and at the same time it has multiple spatial scales ranging from the size of the physical domain to the correlation length of the random permeability field. If we use a traditional method to solve these deterministic problems, we need to use a fine-grid mesh that is finer than the correlation length to obtain accurate numerical solutions. Many existing stochastic methods that are used to solve high dimensional stochastic PDEs use standard FEMs with linear nodal basis; see, e.g., [43, 42, 5, 25, 15, 50, 16, 55, 12, 13]. The computational cost of these methods could be very expensive for every sample. The total computational cost can be tremendous since we need to simulate many sample solutions.

In this paper, we propose a reduced-order method that performs model reduction in the physical space for all samples simultaneously by using a local parametrization of the random coefficients. Our method can significantly speed up the existing *non-intrusive* stochastic methods. By “nonintrusive stochastic methods” we mean those methods that can call a deterministic PDE solver as a blackbox, e.g., Monte Carlo, multilevel Monte Carlo [24, 4, 11], (sparse-grid) stochastic collocation [2, 43, 42, 53], least-squares methods [15, 50], and compressed sensing methods [16, 55]. Our method is based on the following observation: most deterministic model reduction methods require solving only local problems (e.g., [28, 30, 47, 41, 44]), and the local problems often have much lower stochastic dimensions. To be more specific, the random coefficients restricted to a local subdomain can be parametrized by a much smaller number of parameters, which depends only on the ratio between the subdomain size and the correlation length of the random coefficients. Therefore, the local upscaling (equivalent to deterministic model reduction in our paper) results in problems with low stochastic dimension locally in the physical space and can be efficiently precomputed by the gPC-like methods in the offline stage. Based on this observation, we propose a stochastic multiscale finite element method (StoMsFEM) to solve the random PDEs

that have high stochastic dimension globally but low stochastic dimension locally. This method inherits almost all the advantages of the deterministic model reduction methods but removes the limitation that the model reduction process needs to be recomputed for every sample. In this paper, we use the following elliptic equation with heterogeneous random coefficients as an example to illustrate the main idea of our approach:

$$(1.1) \quad \begin{cases} -\nabla_x \cdot (\kappa(x, \omega) \nabla_x u(x, \omega)) = b(x), & x \in D, \omega \in \Omega, \\ u(x, \omega) = 0, & x \in \partial D, \end{cases} \quad P\text{-almost surely.}$$

Here,  $D \in \mathbb{R}^d$  is a bounded spatial domain, and  $(\Omega, \mathcal{F}, P)$  is a probability space. The random coefficient  $\kappa(x, \omega)$  is of high stochastic dimension and has multiscale features. We assume that  $\kappa(x, \omega)$  is a symmetric positive definite matrix satisfying  $\lambda_{\min} \geq \alpha > 0$ , for every  $x \in D$  and  $\omega \in \Omega$ , where  $\lambda_{\min}$  is the smallest eigenvalue of  $\kappa(x, \omega)$ . For such coefficients, the solutions are only Hölder continuous. If  $\kappa(x, \omega)$  has multiple scales, the solution will have multiscale features as well. For simplicity, we assume that the forcing function  $b(x)$  is deterministic.

Our StoMsFEM method consists of three steps; the first two are in the offline stage, and the third is in the online stage. In the first step, we parametrize the random coefficient  $\kappa(x, \omega)$  by exploring the locally low dimensional property of the random media. This can be done by using several approaches, including the local Karhnen–Loève (KL) expansion of the random coefficient, sparse principal component analysis (sparse PCA) [57, 14, 52, 48, 38], and the intrinsic sparse mode decomposition [31]. In the second step, we apply a deterministic local upscaling method to obtain a *parametric upscaled system*. We provide two methods for doing the parametric upscaling: a random interpolation method and a reduced basis method. The random interpolation method takes advantage of the fact that the local upscaled coefficients are analytic functions of the local stochastic parameters, and builds an interpolation scheme for each upscaled coefficient at the coarse-grid level. The random interpolation method can be viewed as a local reduced-order method in the stochastic space. The reduced basis method makes use of the low rank property of the solutions for the local upscaling problems and prepares a small set of spatial basis functions for each local upscaling problem. The reduced basis method can be viewed as a local reduced-order method in the physical space. In the online stage, for each sample of the random parameters, we either interpolate the upscaled coefficients in the random interpolation setting or solve the small reduced-order systems to obtain the upscaled coefficients. A numerical coarse-grid solution for this sample can be obtained by solving the upscaled system.

We have performed a careful computational complexity analysis of our method. The computational cost of the StoMsFEM consists of the offline and online costs. The offline cost is equivalent to solving the random PDE for  $N_{\text{off}}$  samples on the fine grid. In the online stage, the computational cost consists of solving the upscaled system  $N_{\text{on}}$  times. Our complexity analysis shows that  $N_{\text{off}} \ll N_{\text{on}}$  and that the offline computational cost of the StoMsFEM is negligible compared with the online cost. Moreover, we show that the ratio between the online cost for the StoMsFEM and the cost of the standard FEM on the fine grid is of the order  $(h/H)^d(\log(H/h))^k$ . Here  $H/h$  is the ratio between the coarse and the fine grid sizes,  $d$  is the physical dimension, and  $k$  is the local stochastic dimension. Therefore, the StoMsFEM gives a speed-up of order  $(H/h)^d/(\log(H/h))^k$  over the standard FEM method on a fine grid for a single query problem. We have applied the StoMsFEM to solve several random elliptic PDEs with varying degrees of difficulty to demonstrate the accuracy and

computational savings of the StoMsFEM. In the high contrast example, we observe a factor of 2000 speed-up over the naive application of the MsFEM.

We remark that the MsFEM achieves computational savings only for multiple queries. For a multiquery problem, the StoMsFEM can reuse the parametric upscaled system that we obtain in the offline stage, and thus there is no offline cost for additional source terms or boundary conditions. The computational savings for the StoMsFEM are even more significant in a multiquery setting.

There are several other methods that share a philosophy similar to that of StoMsFEM. In [17], GMsFEM has been applied to solve parametric PDEs with multiple scales. GMsFEM assumes that the coefficients are already parametrized, while in StoMsFEM we need to first build a locally low dimensional parametrization of the random coefficients. In addition, the StoMsFEM can be implemented with any locally upscaling method including GMsFEM. In [7], the authors also observed that the random inputs have low stochastic dimension locally and used the local KL expansion to parametrize the random coefficients. They proposed combining the deterministic domain decomposition method (DDM) with the local gPC expansions and Monte Carlo sampling to achieve computational savings. A major difference between the StoMsFEM and the method proposed in [7] is that their method does not deal with the multiscale feature, which contributes to the stochastic high dimensionality. As a result, they have not explored the low dimensional structure in the physical space. We will compare the StoMsFEM with these methods as we get into the details of StoMsFEM.

The rest of the paper is organized as follows. In section 2, we give a brief review of several locally low dimensional parametrization methods and the MsFEM. In section 3, we present our StoMsFEM method, which uses either the random interpolation method or the reduced basis method. We also perform complexity analysis for our method. In section 4, we show how to combine our methodology with existing stochastic methods, e.g., the Monte Carlo (MC), multilevel Monte Carlo (MLMC), and sparse-grid stochastic collocation (SC) methods. In section 5, we demonstrate the accuracy and efficiency of our method through several numerical examples. Finally, some concluding remarks are given in section 6.

**2. Preliminaries.** The StoMsFEM that we propose relies on two building blocks: locally low dimensional parametrization of a random field and deterministic local upscaling methods. Although these methods are not the focus of the current paper, we give a brief review below for completeness.

**2.1. Locally low dimensional parametrization.** Typically for a random coefficient  $\kappa(x, \omega)$  in (1.1), its parametrization is not known, but rather, its mean and covariance are given:

$$(2.1) \quad \bar{\kappa}(x) = \mathbb{E}[\kappa(x, \omega)], \quad \text{Cov}(x, y) = \mathbb{E}[(\kappa(x, \omega) - \bar{\kappa}(x))(\kappa(y, \omega) - \bar{\kappa}(y))].$$

In order to solve the random PDE, one first needs to parametrize the random coefficient with some random parameters. The KL expansion [35, 40] is the most popular method in parametrizing the random media. However, the eigenfunctions of the covariance function (also called the KL modes) are global in nature. As a result, the local stochastic dimension is the same as the global stochastic dimension. In this section, we will briefly review a few methods to get a locally low dimensional parametrization, including the local KL expansion, the intrinsic sparse mode decomposition, and the sparse PCA approach. For a detailed description and comparison of these methods, please refer to the companion paper [31].

**2.1.1. Local KL expansion.** The local KL expansion is a natural way to construct a locally low dimensional parametrization of the random medium (also used in [7]). Let  $D$  be divided into a set of nonoverlapping subdomains  $\{P_m\}_{m=1}^M$ , called patches:

$$(2.2) \quad \overline{P} = \bigcup_{m=1}^M \overline{P_m}, \quad P_m \cap P_n = \emptyset \quad \text{for } m \neq n.$$

Let  $\text{Cov}_m : P_m \times P_m \rightarrow \mathbb{R}$  be the global covariance function  $\text{Cov}(x, y)$  restricted to the  $m$ th patch:

$$(2.3) \quad \text{Cov}_m(x, y) = \text{Cov}(x, y), \quad x, y \in P_m.$$

Similar to the standard KL expansion, we can define a local KL expansion as follows.

**DEFINITION 2.1** (local KL expansion of  $\kappa(x, \omega)$ ). *Perform KL expansion in each subdomain  $P_m$ :*

$$(2.4) \quad \begin{aligned} \int_{P_m} \text{Cov}_m(x, y) f_{k,m}(y) dy &= \lambda_{k,m} f_{k,m}(x), \\ \xi_{k,m}(\omega) &= \frac{1}{\sqrt{\lambda_{k,m}}} \int_{P_m} (\kappa(x, \omega) - \bar{\kappa}(x)) f_{k,m}(x) dx. \end{aligned}$$

Arrange  $\lambda_{k,m}$  in a descending order, and truncate the expansion at the  $K_m$ th mode. Then, we obtain a local parametrization as follows:

$$(2.5) \quad \kappa(x, \omega) \approx \mathbb{E}[\kappa(x, \cdot)] + \sum_{k=1}^{K_m} \sqrt{\lambda_{k,m}} \xi_{k,m} f_{k,m}(x), \quad x \in P_m.$$

**2.1.2. Intrinsic sparse mode decomposition.** In [31] the authors proposed the intrinsic sparse mode decomposition (ISMD), which decomposes a symmetric positive semidefinite matrix into several sparse rank-one components. We assume that the covariance matrix,  $\text{Cov}$ , can be decomposed into a finite number of sparse modes; i.e.,  $\text{Cov} = \sum_{k=1}^K g_k g_k^T$ . ISMD looks for a patchwise sparse decomposition by minimizing the total local dimension, i.e.,

$$(2.6) \quad \boxed{\min_{g_1, \dots, g_K} \sum_{m=1}^M d_m \quad \text{subject to} \quad \text{Cov} = \sum_{k=1}^K g_k g_k^T},$$

where  $d_m$  is the number of nontrivial modes among  $\{g_k\}_{k=1}^K$  on the local patch  $P_m$ , defined as

$$d_m = \#\{k : g_k|_{P_m} \neq \mathbf{0}\}.$$

Under certain nondegenerate assumptions on the covariance  $\text{Cov}$  and the partition  $\mathcal{P}$ , we proved that the ISMD produces exactly one minimizer of the minimization problem (2.6); see Theorem 3.5 in [31]. After projecting the random field  $\kappa(x, \omega)$  onto the sparse modes  $\{g_k\}_{k=1}^K$ , we get a parametrization with  $K$  random parameters:

$$(2.7) \quad \kappa(x, \omega) = \bar{\kappa}(x) + \sum_{k=1}^K g_k(x) \xi_k(\omega),$$

where the random variables  $\{\xi_k\}_{k=1}^K$  are normalized (with center zero and variance one) and uncorrelated. Moreover, the parametrization (2.7) achieves the minimal total local stochastic dimension, as desired.

It is worth mentioning that there are several other methods that are able to achieve locally low dimensional parametrization, for example, the sparse PCA [57, 14, 52, 39] and sparse operator compression [32, 33].

When applying the proposed StoMsFEM, the most important factor in choosing a parametrization method is the global stochastic method, which we will discuss in section 4. If one wants to use the MC-type methods, we recommend using the local KL expansion. This is the typical case because the StoMsFEM is aiming at stochastically high dimensional problems, with which the gPC-type methods would have difficulties. If the global stochastic dimension is within the range of the (sparse-grid) SC method, and if one wants to use the (sparse-grid) SC method to save computational costs, one should choose the ISMD or the sparse PCA, which would parametrize the random coefficients more effectively. The ISMD is recommended when the random parameters are required to be uncorrelated and a high accuracy parametrization is desired, e.g., for the synthetic porous media in sections 5.1 and 5.2. The sparse PCA and many other matrix factorization methods are good at parametrizing random coefficients whose covariance matrix has continuously decaying eigenvalues, e.g., the Gaussian kernel  $\exp(-|x - y|^2/l^2)$  in section 5.3. We give a more detailed comparison between the ISMD and the sparse PCA in our companion paper [31].

**2.1.3. Nonlinear transformations.** All the parametrization methods above, including the KL expansion, are affine with respect to the random parameters. In some applications, the use of nonlinear transformations may reduce the number of parameters significantly. For example, the following nonlinear transformation has been widely used to parametrize a positive random field  $\kappa(x, \omega)$ :

$$(2.8) \quad \kappa(x, \omega) = \kappa_{\min} + \exp(\beta(x, \omega)).$$

The expression has strict positive lower bound  $\kappa_{\min}$ , and in practice we can apply affine parametrization to  $\beta(x, \omega)$ .

If the random field has both lower and upper bounds ( $\kappa_{\min}$  and  $\kappa_{\max}$ , respectively), the following nonlinear transformation is usually used:

$$(2.9) \quad \kappa(x, \omega) = \frac{\kappa_{\max} + \kappa_{\min}}{2} + \frac{\kappa_{\max} - \kappa_{\min}}{2} \tanh(\beta(x, \omega)).$$

**2.2. Multiscale finite element method.** Model reduction methods based on local upscaling is the other building block in our StoMsFEM. There have been a number of such local upscaling methods for elliptic equations with heterogeneous diffusion coefficients; see, e.g., [28, 51, 30, 45, 1, 46, 17, 47, 41, 10, 44]. In this paper, we will use the multiscale finite element method (MsFEM) developed in [28, 30] for the local upscaling. We point out that, per the user's preference, the MsFEM can be replaced by other local upscaling methods with minor modifications. In the following, we briefly review the MsFEM applied on (1.1) with a specific sample media, denoted as  $\kappa(x, \omega)$ . Note that here the media is fixed and (1.1) is deterministic.

Suppose that the physical domain  $D$  is partitioned into a finite set of compact triangles or quadrilaterals  $\{D_H^m, 1 \leq m \leq M\}$ , which forms a triangulation  $\mathcal{T}_H$  with mesh size  $H$ . We assume that the coarse grid mesh size  $H$  is much larger than the small scale  $\epsilon$  in the rough coefficient  $\kappa(x, \omega)$ ; i.e.,  $H \gg \epsilon$ . In block  $D_H^m$ , we compute

the following cell problem:

$$(2.10) \quad \begin{cases} -\nabla \cdot (\kappa(x, \omega) \nabla \phi^{ml}(x, \omega)) = 0, & x \in D_H^m, \\ \phi^{ml}(x, \omega) = p^{ml}(x), & x \in \partial D_H^m, \end{cases} \quad l = 1, \dots, L.$$

Here  $L$  is the number of nodes on  $D_H^m$ , and  $p^{ml}$  is defined on the boundary  $\partial D_H^m$  playing the role of Dirichlet boundary conditions. In our computation we could choose  $p^{ml}$  as a linear basis (for triangles) or standard bilinear basis (for quadrilaterals) that takes value 1 at node  $l$  and 0 for all the other  $L - 1$  nodes in the patch. In practice, we solve the local cell problem on a fine mesh  $\mathcal{T}_h$  that resolves the small scales in  $\kappa(x, \omega)$ . Thereafter, the local upscaled stiffness matrix and the load vector can be computed as

$$(2.11) \quad \begin{aligned} \mathbf{S}_{ll'}^m(\omega) &= \int_{D_H^m} \kappa(x, \omega) \nabla \phi^{ml}(x, \omega) \cdot \nabla \phi^{ml'}(x, \omega) dx, \\ \mathbf{b}_l^m(\omega) &= \int_{D_H^m} b(x) \phi^{ml}(x, \omega) dx. \end{aligned}$$

The standard assembling procedure can be utilized to assemble  $\mathbf{S}$  and  $\mathbf{b}$  by looping over all the coarse grid elements. After solving the upscaled system

$$(2.12) \quad \mathbf{S}\mathbf{U} = \mathbf{b},$$

we obtain the multiscale solution

$$(2.13) \quad u_H(x, \omega) = \sum_m U_m(\omega) \phi_m(x, \omega).$$

When the boundary conditions  $p^{ml}$  are linear, the following convergence theorem was proved in [29].

**THEOREM 2.2.** *Let  $\kappa^\varepsilon(x) = \kappa(\frac{x}{\varepsilon})$  be a smooth periodic medium, and  $u^\varepsilon(x)$  the solution to (1.1). Denote by  $u_H^\varepsilon(x)$  the multiscale finite element approximation obtained from the space spanned by the multiscale basis with linear boundary conditions. Then we have*

$$\|u^\varepsilon - u_H^\varepsilon\|_{H^1} \leq C(H + \varepsilon) \|f\|_{L^2} + C \sqrt{\frac{\varepsilon}{H}} \|u_0\|_{H^2},$$

where  $u_0$  is the solution to the homogenized equation of (1.1).

**Remark 2.3.** Theorem 2.2 implies that the MsFEM captures the correct homogenized results for small  $\varepsilon$ . However, the method may produce a large error if  $\varepsilon \sim H$ . This is called the resonance error between the coarse grid scale  $H$  and media small-scale parameter  $\varepsilon$ . The oversampling technique proposed in [28] successfully reduces the resonance error, based on the observation that the boundary layer typically gets damped out quickly within a width of  $\varepsilon$ . This suggests that one compute (2.10) in a larger domain of order  $H + \mathcal{O}(\varepsilon)$  in each dimension and utilize the interior information for basis construction. This significantly reduces the resonance error and gives more accurate results, as demonstrated in numerical examples in [28].

We remark that the oversampling technique results in a nonconforming FEM. To reduce the nonconforming FEM error, we use the Petrov–Galerkin MsFEM formulation [30]. In the Petrov–Galerkin MsFEM, the local upscaled quantities are computed

from

$$(2.14) \quad \begin{aligned} \mathbf{S}_{ll'}^m(\omega) &= \int_{D_H^m} \kappa(x, \omega) \nabla \phi_t^{ml}(x, \omega) \cdot \nabla \phi_t^{ml'}(x) dx, \\ \mathbf{b}_l^m &= \int_{D_H^m} b(x) \phi_t^{ml}(x) dx, \end{aligned}$$

with  $\{\phi_t^{ml}(x)\}_{l=1}^L$ , the test functions, being linear (triangular grid) or bilinear (quadrilateral grid) locally. One added benefit of using the Petrov–Galerkin MsFEM is that the test functions are deterministic, and thus the local load vector  $\mathbf{b}^m$  is independent of random samples.

It is important to point out that there is some computational overhead in computing cell problems (2.10) and assembling the upscaled system (2.11). However, the upscaled stiffness matrix  $\mathbf{S}$  can be reused for different source terms  $b(x)$ , and the MsFEM achieves no computational savings except in the multiquery setting. This is true for most model reduction methods since there is an overhead in the offline stage in constructing the reduced-order models. In the random setting, with a single source term in (1.1), if we naively apply the MsFEM for every sample media, the reduced models need to be recomputed for every sample coefficient, which leads to no computational savings. In the multiquery case, since random samples may be different for different queries, we are still not able to gain the full power of the MsFEM.

We introduce our StoMsFEM below. In this new method, we prepare the upscaled quantities (2.11) for all samples simultaneously in the offline stage. In the online stage, for every sample we only need to assemble and solve the upscaled system (2.12). The computational savings are achieved even in the single-query setting.

**3. Stochastic multiscale finite element method.** The StoMsFEM consists of two stages: the offline stage, in which we construct reduced-order models and prepare upscaled quantities, and the online stage, in which we sample media and compute the upscaled system. We discuss the offline preparation stage in this section and leave the online global computation to the next section.

The offline stage also consists of two steps. In the first step, we parametrize the random coefficients  $\kappa(x, \omega)$  using a local parameterization method, as reviewed in section 2.1. More discussion is found in the companion paper [31]. In this paper, we assume that the random coefficient  $\kappa(x, \omega)$  has already been parametrized, denoted as  $\kappa(x, \xi)$ , where  $\xi \equiv [\xi_1, \xi_2, \dots, \xi_K]^T$  is the collection of  $K$  random variables. Note that we do *not* assume the affine structure with respect to  $\xi$  in the parametrization of  $\kappa(x, \xi)$ . Our StoMsFEM works for any parametrization with locally low stochastic dimensions.

With this parametrization, the cell problem (2.10) becomes a parametrized PDE:

$$(3.1) \quad \begin{cases} -\nabla \cdot (\kappa(x, \xi_m) \nabla \phi^{ml}(x, \xi_m)) = 0, & x \in D_H^m, \\ \phi^{ml}(x, \xi_m) = p^{ml}(x), & x \in \partial D_H^m, \end{cases} \quad \forall \xi_m \in \Gamma_m, \quad l = 1, \dots, L.$$

Here  $\xi_m \in \Gamma_m \subset \mathbb{R}^{K_m}$  are the local effective parameters, with  $\Gamma_m$  being its range, and  $K_m$  is the local stochastic dimension on patch  $D_H^m$ . Following the same assembling procedure (2.11), the local upscaled quantities  $\mathbf{S}^m$  and  $\mathbf{b}^m$  are functions of the local parameters, i.e.,  $\mathbf{S}^m(\xi_m)$  and  $\mathbf{b}^m(\xi_m)$ .

In the second step, we construct reduced-order models for the upscaled quantities  $\mathbf{S}^m(\xi_m)$  and  $\mathbf{b}^m(\xi_m)$ . Two methods are proposed, i.e., the random interpolation

method and the reduced basis method. The random interpolation method makes use of the fact that  $\phi^{ml}(x, \xi_m)$  (and thus  $\mathbf{S}^m(\xi_m)$  and  $\mathbf{b}^m(\xi_m)$ ) are smooth with respect to  $\xi_m$ , while the reduced basis method makes use of the low dimensional structure of  $\phi^{ml}(x, \xi_m)$  in the physical space. The random interpolation method works for all parametrized local coefficients  $\kappa(x, \xi_m)$  that are smooth with respect to the local parameters  $\xi_m$ , while the reduced method basis is recommended only when  $\kappa(x, \xi_m)$  is affine with respect to  $\xi_m$ .

**3.1. Random interpolation method.** In the random interpolation method, we propose to compute the cell problem (3.1) and the local upscaled quantities  $\mathbf{S}^m(\xi_m)$  and  $\mathbf{b}^m(\xi_m)$  on multiple *deterministic* collocation points and use them to build interpolants of the upscaled quantities in terms of the parameters  $\xi_m$ . These interpolants allow us to efficiently approximate the upscaled quantities for any given sample in the online stage.

We take patch  $D_H^m$  as an example to illustrate the idea. If  $\kappa(x, \xi_m)$  is smooth with respect to  $\xi_m$ , the local multiscale basis functions  $\phi^{ml}(x, \xi_m)$  are smooth with respect to  $\xi_m$ ; see [3, 2, 13, 27]. Assume that the range for all these random variables lies in the interval  $[-1, 1]$  (other bounded ranges can be rescaled accordingly), and in the simplest scenario, we sample  $\nu + 1$  collocation points along each dimension; the entire collocation set is the tensor product

$$\Gamma_{1d} = \{-1 \leq \xi_0 < \xi_1 < \cdots < \xi_\nu \leq 1\}, \quad \Gamma_{m,c} = \prod_{k=1}^{K_m} \Gamma_{1d} \in \Gamma_m. \quad ^1$$

The “c” in the subscript stands for collocation. In total  $N_c = (\nu + 1)^{K_m}$  collocation points are sampled. If the joint distribution of  $\xi_m$  is known, we can choose the one dimensional (1d) collocation nodes  $\Gamma_{1d}$  in the same way as for stochastic collocation methods [2]. For example, if elements of  $\xi_m$  are i.i.d. copies of the random variable that is uniformly distributed on  $[-1, 1]$ , then  $\Gamma_{1d}$  should be the zeros of the Legendre polynomials. If the joint distribution of  $\xi_m$  is unknown, we can simply use the Chebyshev nodes.

For each collocation point  $\xi_{m,c} \in \Gamma_{m,c}$ , we solve the cell problem (3.1) for the local basis functions, denoted as  $\phi^{ml}(x, \xi_{m,c})$ . We then assemble the local stiffness matrix and the local load vector, denoted as  $\mathbf{S}^m(\xi_{m,c})$  and  $\mathbf{b}^m(\xi_{m,c})$ , according to (2.11). Here  $\mathbf{S}^m(\xi_{m,c})$  is an  $L \times L$  matrix and  $\mathbf{b}^m(\xi_{m,c})$  is an  $L$  dimensional vector. We then construct the interpolants of  $\mathbf{S}^m$  and  $\mathbf{b}^m$  in terms of the random variables  $\xi_m$ . Such interpolants are constructed for each element of  $\mathbf{S}^m$  and  $\mathbf{b}^m$ . For example, for each element in  $\mathbf{S}^m$  and  $\mathbf{b}^m$ , we construct the Lagrange polynomial approximation, denoted as  $\hat{\mathbf{S}}^m$  and  $\hat{\mathbf{b}}^m$ .

The interpolants  $\hat{\mathbf{S}}^m$  and  $\hat{\mathbf{b}}^m$  will be used to obtain the approximation for every sample in the online stage. We will discuss how to determine which sample set to solve in the online stage in section 4.1.

*Remark 3.1.* Several remarks are in order:

- When the local stochastic dimensions are moderate, sparse grids and dimensional adaptive grids [22, 36] can be used to reduce the number of collocation points. We use these techniques in our numerical examples.

---

<sup>1</sup>To simplify the notation, we choose the same number of collocation points along different directions. However, in general we use different numbers of collocation points along different directions.

- To determine the collocation points  $\Gamma_m$ , we do not need to know the density of the parameters  $\xi_m$ . We need only know their ranges. This is very different from the standard gPC method, which requires the joint density function of the random parameters. It is possible that the range of some random parameter is an unbounded domain, in which case we truncate its range to a bounded domain that is large enough to cover the parameter range with high probability. For parameter configurations that fall outside this bounded domain, we directly compute  $\phi^{ml}(x, \xi_m)$ ,  $\mathbf{S}^m(\xi_m)$ , and  $\mathbf{b}^m(\xi_m)$  from (3.1) and (2.11).

Finally, we summarize the offline and online implementation of random interpolation method in Algorithms 1 and 2.

---

**Algorithm 1** The offline stage of random interpolation method.

---

- 1: Partition the physical domain  $D$  into coarse grid blocks  $\mathcal{T}_H$ .
  - 2: **for** each patch  $D_H^m$  **do**
  - 3:     Determine the set of interpolation nodes  $\Gamma_{m,c}$ .
  - 4:     Solve local multiscale basis  $\{\phi^{ml}(x, \xi_{m,c})\}_{c=1}^{N_c}$  according to (2.10).
  - 5:     Assemble the local upscaled quantities  $\mathbf{S}^m$  and  $\mathbf{b}^m$  according to (2.11) or (2.14).
  - 6:     Build the Lagrange polynomial interpolants for  $\mathbf{S}^m$  and  $\mathbf{b}^m$ .
  - 7: **end for**
- 

**Algorithm 2** The online stage of random interpolation method for a specific configuration  $\xi$ .

---

- 1: **for** each patch  $D_H^k$  **do**
  - 2:     Determine the values of local variables  $\xi_m$ .
  - 3:     Interpolate the local upscaled stiffness matrix  $\mathbf{S}^m$  and local upscaled loading vector  $\mathbf{b}^m$ .
  - 4: **end for**
  - 5: Assemble and solve the upscaled system (2.12) to obtain  $\hat{u}_H(x, \xi)$ .
- 

**3.1.1. Accuracy of the random interpolation.** In this subsection, we estimate the interpolation error  $\hat{\mathbf{S}}^m - \mathbf{S}^m$ . If the media  $\kappa(x, \xi_m)$  smoothly depend on  $\xi_m$ , we can prove that the solution to the cell problem (3.1) is also smooth with respect to  $\xi_m$ . So are  $\mathbf{S}^m$  and  $\mathbf{b}^m$ . The strong regularity of  $\mathbf{S}^m(\xi_m)$  and  $\mathbf{b}^m(\xi_m)$ , combined with the high-order approximation method, is the key for the success of the random interpolation method.

It is worth mentioning that the regularity problem has been well studied in the literature; see [3, 2, 13, 27]. In particular, with a small modification of Lemma 3.2 in [2], we have the following lemma.

LEMMA 3.2. *Let  $\phi^{ml}$  be the multiscale basis function in the  $m$ th cell problem (3.1). We use the notation  $\xi_m = [\xi_{m,1}, \xi_{m,2}, \dots, \xi_{m,K_m}]$  for the list of the random parameters effective in patch  $D_H^m$ . We assume that  $\xi_m \in \Gamma_m \equiv [-1, 1]^{K_m}$ . If the local parametrization on patch  $D_H^m$ , i.e.,  $\kappa(x, \xi_m)$ , is infinitely differentiable with respect to  $\xi_m$  and there exists some  $c_k > 0$  such that*

$$(3.2) \quad \left\| \frac{\partial_{\xi_{m,k}}^n \kappa(\xi_m)}{\kappa(\xi_m)} \right\|_{L^\infty(D_H^m)} \leq c_k^n n! \quad \forall n \geq 0,$$

then we have for every  $\xi_m \in \Gamma_m$  and  $n \geq 0$

$$(3.3) \quad \left\| \sqrt{\kappa(\xi_m)} \nabla \partial_{\xi_{m,k}}^n \phi^{ml}(\xi_m) \right\|_{L^\infty(\Gamma_m; L^2(D_H^m))} \leq C_\phi \sqrt{\kappa_{\max}} n! (2c_k)^n,$$

where  $C_\phi$  depends only on the local domain  $D_H^m$  and the deterministic boundary condition  $p^{ml}$  for  $\phi^{ml}$ .

Assumption (3.2) holds true for most parametrization methods. For example, for a linear parametrization

$$\kappa(x, \xi_m) = \bar{\kappa}(x) + \sum_{k=1}^{K_m} g_{m,k}(x) \xi_{m,k},$$

provided that such expansion guarantees  $\kappa(x, \xi_m) \geq \kappa_{\min}$  for almost every  $x \in D$  and  $\xi_m \in \Gamma_m$ , we can take  $c_k = \|g_k\|_{L^\infty(D)}/\kappa_{\min}$ . When a linear parametrization is combined with the exponential transformation (2.8), i.e.,

$$\kappa(x, \xi_m) = \kappa_{\min} + \exp \left( \bar{\beta}(x) + \sum_{k=1}^{K_m} g_{m,k}(x) \xi_{m,k} \right),$$

we can take  $c_k = \|g_k\|_{L^\infty(D)}$ .

The regularity of  $\phi^{ml}(\xi_m)$  in Lemma 3.2 could be extended to that of  $S^m$  in a straightforward manner.

**THEOREM 3.3.** *Under the same assumptions as in Lemma 3.2, for any  $l, l' \in \{1, 2, \dots, L\}$ ,  $S_{ll'}^m$  as a function of  $\xi_{m,k}$ ,  $S_{ll'}^m : [-1, 1] \rightarrow C([-1, 1]^{K_m-1})$  admits an analytic extension to the complex domain*

$$(3.4) \quad \Sigma([-1, 1]; \tau_k) \equiv \{z \in \mathbb{C}, \text{dist}(z, [-1, 1]) \leq \tau_k\}$$

with  $0 < \tau_k < 1/(2c_k)$ .

Finally, we have the following theorem, which guarantees the accuracy of the random interpolation method.

**THEOREM 3.4.** *Under the same assumptions as in Lemma 3.2, for any  $l, l' \in \{1, 2, \dots, L\}$  there exist positive constants  $r_k$ ,  $k = 1, \dots, K_m$ , and  $C_s$ , independent of  $\nu \equiv (\nu_1, \nu_2, \dots, \nu_{K_m})$ , such that*

$$\|S_{ll'}^m - \hat{S}_{ll'}^m\|_{C(\Gamma_m)} \leq C_s \prod_{k=1}^{K_m} \left( \frac{2}{\pi} \log(\nu_k + 1) + 1 \right) \sum_{k=1}^{K_m} \exp(-r_k \nu_k),$$

where  $\hat{S}_{ll'}^m$  is the Chebyshev interpolation of  $S_{ll'}^m$  with  $\nu_k + 1$  collocation points in the  $\xi_{m,k}$  direction,  $r_k = \log [\tau_k (1 + \sqrt{1 + 1/\tau_k^2})]$ , and  $\tau_k$  any positive constant that is strictly smaller than the distance between the real line segment  $[-1, 1]$  and the nearest singularity in the complex plane, as defined in Theorem 3.3.

The proof is the same as that of Theorem 4.1 in [2]. The only difference is that we are considering the interpolation in the continuous function space. Therefore, we have the Lebesgue constant of the Chebyshev interpolation, i.e.,  $\frac{2}{\pi} \log(\nu_k + 1) + 1$ , which appears in our error estimation. The regularity and approximation accuracy of  $b_m$  can be analyzed similarly.

*Remark 3.5.* Several remarks are in order:

- The estimates above are based on simple energy estimates and are far from being sharp. More dedicated analysis has been carried in [13] for improved results and could be easily carried over. Detailed regularity analysis is not the focus of the paper, and we omit it.
- In the case when the random variables have infinite range (e.g., the Gaussian variable), we can sample the collocation points in a range that is large enough to cover the random variable with high probability. For example, denote by  $\Gamma_{m,0}$  the finite domain that is large enough to cover a very large portion of  $\Gamma_m$  such that  $P(\xi_m \notin \Gamma_{m,0}) < \varepsilon$ ; then we can build an interpolant of  $\mathbf{S}^m$  such that  $\mathbf{S}^m - \widehat{\mathbf{S}}^m$  is pointwise small on  $\Gamma_{m,0}$ . For  $\xi_m \notin \Gamma_{m,0}$  we directly compute  $\mathbf{S}^m(\xi_m)$  and  $\mathbf{b}^m(\xi_m)$  from (3.1) and (2.11). The computed upscaled quantities, denoted as  $\widehat{\mathbf{S}}^m$  and  $\widehat{\mathbf{b}}^m$ , give a very accurate approximation of the true values  $\mathbf{S}^m$  and  $\mathbf{b}^m$ .

**3.1.2. Complexity analysis.** We summarize the computational cost in this subsection. Without loss of generality, we assume that the diameter of the physical domain is 1. The coarse mesh size is denoted by  $H$ , and thus the number of coarse grid elements is  $M \sim 1/H^d$ . In each coarse grid element, we use a fine mesh of size  $h$  to solve the cell problem. Given a sample  $\kappa(x, \omega)$ , we assume that the computational cost to solve a deterministic PDE (1.1) is

$$(3.5) \quad \mu = (1/h)^{d\gamma}, \quad \gamma \geq 1,$$

where  $\gamma = 1$  corresponds to the multigrid method (neglecting the logarithmic factor). In the same manner, we assume that the computational cost of solving an upscaled system (2.12) is  $M^\gamma \sim (1/H)^{d\gamma}$ , and that the computational cost of solving a local cell problem is  $(\eta H/h)^{d\gamma}$ , where  $\eta$  is the oversampling ratio. We also assume that the number of random variables in each coarse grid element is about the same, denoted by  $K_m$ . By an appropriate choice of  $H$ ,  $K_m$  could be as small as 2 or 3. The number of offline collocation points is denoted as  $N_c = (\nu + 1)^{K_m}$  for each coarse grid element.

The computational cost of the random interpolation method consists of the offline cost and the online cost. In the offline stage (see Algorithm 1) we need to solve  $L - 1$  local cell problems and assemble local stiffness matrices for each collocation point on each coarse-grid element.

- For each collocation point within a coarse-grid element, we need to construct  $L - 1$  basis functions by solving an elliptic equation with a total number of  $(\eta H/h)^d$  fine grid points, where  $\eta$  is the oversampling ratio. The computational cost for this step is given by

$$\text{Cost}_{\text{basis}} = (L - 1)MN_c(\eta H/h)^{d\gamma}.$$

The oversampling ratio  $\eta$  is typically taken to be  $\eta = 2$ .

- The second step in the offline stage is to assemble the local stiffness matrix. Such a procedure is performed on every collocation point within each coarse-grid element, and the computational cost is

$$\text{Cost}_{\text{assemble}} = \frac{(L + 1)L}{2}MN_c(H/h)^d.$$

Here  $(L + 1)L/2$  is due to the fact that each local stiffness matrix has  $(L + 1)L/2$  different entries (other  $(L - 1)L/2$  elements are determined by symmetry), and the factor  $(H/h)^d$  comes from evaluating the  $l_2$  inner product

defined in (2.11) over a coarse-grid element by using  $(H/h)^d$  many fine grid points.

In total, we have

$$(3.6) \quad \frac{\text{Cost}_{\text{offline}}}{\mu} \sim (L-1)N_c H^{d(\gamma-1)} + \frac{(L+1)L}{2} N_c h^{d(\gamma-1)}.$$

In the extreme case of  $\gamma = 1$ , we have  $\frac{\text{Cost}_{\text{offline}}}{\mu} \sim O(N_c)$ . If  $\gamma > 1$ ,  $\frac{\text{Cost}_{\text{offline}}}{\mu}$  is even smaller. This is to say that, compared with the multigrid method on the fine grid ( $\gamma = 1$ ), the offline computational cost is the same as solving the original equation (1.1) for about  $N_c$  times. Here  $N_c = (\nu+1)^{K_m}$  is the number of local collocation points and is much smaller than the number of samples  $N_{\text{on}}$  that is required to solve in the online stage. Therefore, the computational overhead of the random interpolation method is quite reasonable.

In the online stage, we need to interpolate the stiffness matrix and then solve the coarse grid system. If  $N_{\text{on}}$  samples are computed, the computational cost is

$$\text{Cost}_{\text{online}} = N_{\text{on}} \left( \frac{(L+1)L}{2} M N_c + M^\gamma \right).$$

Here the first term comes from the stiffness matrix interpolation, and the second is to solve the upscaled linear system (2.12). For every sample in the online stage, we have

$$(3.7) \quad \frac{\text{Cost}_{\text{online}}}{\mu} \sim \frac{(L+1)L}{2} N_c H^{d(\gamma-1)} (h/H)^{\gamma d} + (h/H)^{\gamma d},$$

which is of the order  $N_c(h/H)^d$  in the extreme case  $\gamma = 1$  and is much smaller if  $\gamma > 1$ . The computational savings  $(h/H)^{\gamma d}$  comes from the usage of MsFEM. However, since we need to do interpolation to get the upscaled system, we have to pay a factor of  $N_c$  as the interpolation cost.

Since the cell problem is solved on the fine mesh  $\mathcal{T}_h$ , there is  $\mathcal{O}(h)$  error in the upscaled system due to this spatial discretization. Therefore, as long as the interpolation error in the stochastic space is smaller than  $\mathcal{O}(h)$ , it will not influence the accuracy of the computed upscaled system. Due to the exponential decay of the interpolation error (see Theorem 3.4), it is sufficient to choose the degree of interpolation polynomial  $\nu \sim \log(H/h)$ . Therefore, the online cost would be

$$(3.8) \quad \frac{\text{Cost}_{\text{online}}}{\mu} \sim \frac{(\log(H/h))^{K_m}}{(H/h)^{\gamma d}},$$

which implies that we obtain a significant computational savings in the online stage if the local dimension  $K_m$  is small. However, the computational savings quickly decreases as  $K_m$  increases. In this case, one should use a sparse-grid interpolation instead. Finally, the total computational cost for this random interpolation method is

$$(3.9) \quad \frac{\text{Cost}_{\text{StoMsFEM}}}{\mu} = N_{\text{off}} + R N_{\text{on}},$$

where  $N_{\text{off}} = \mathcal{O}(N_c)$  is the effective number of samples we solve in the offline stage, and  $R = \mathcal{O}(N_c(h/H)^{\gamma d})$  is the online computational savings achieved by the random interpolation method.

**3.2. The reduced basis method.** Besides exploring the regularity of the upscaled quantities  $\mathbf{S}^m$ , as is done in the random interpolation method, another idea is to make use of the low rank property of multiscale basis  $\phi^m(x, \omega)$ . This leads to the design of the reduced basis method. As in the last section, we suppress the superindex  $l$  in  $\phi^{ml}$  in (3.1) when no confusion arises.

**3.2.1. Reduced basis construction via KL expansion.** To obtain the reduced basis, we apply the KL expansion to  $\phi^m(x, \omega) \in L^2(\Omega; H_0^1(D_H^m; \bar{\kappa}))$ ,

$$(3.10) \quad \phi^m(x, \omega) = \bar{\phi}^m(x) + \sum_{q=1}^{\infty} \sqrt{\lambda_q} \tau_q(\omega) \zeta^{m,q}(x), \quad \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots,$$

where  $\bar{\phi}^m(x) = \int \phi^m(x, \omega) dP(\omega)$  is the mean of  $\phi^m(x, \omega)$ . Notice that this KL expansion is performed in the Hilbert space  $L^2(\Omega; H_0^1(D_H^m; \bar{\kappa}))$  (which is isometric to  $L^2(\Omega) \times H_0^1(D_H^m; \bar{\kappa})$ ) to guarantee that the reduced basis method is accurate in  $H_0^1(D_H^m; \bar{\kappa})$ .<sup>2</sup> Thanks to the local low stochastic dimensionality, the energies  $\{\lambda_q\}_{q=1}^{\infty}$  in (3.10) decay exponentially fast. We take the first  $Q$  KL modes  $\{\zeta^{m,q}\}_{q=1}^Q$  as the reduced basis functions and expand the solution to (2.10) as

$$(3.11) \quad \phi_{rb}^m(x, \omega) = \bar{\phi}^m(x) + \sum_{q=1}^Q c_q^m(\omega) \zeta^{m,q}(x).$$

Using the Galerkin method, we solve the following linear system to obtain the coefficients  $\mathbf{c}^m$ :

$$(3.12) \quad \mathbf{A}^m(\omega) \mathbf{c}^m(\omega) = \mathbf{F}^m(\omega),$$

where  $\mathbf{A}^m(\omega)$  is a  $Q \times Q$  symmetric positive definite matrix with entries  $\mathbf{A}_{qq'}^m(\omega) = \kappa(\zeta^{m,q}, \zeta^{m,q'}; \omega)$ ,  $1 \leq q, q' \leq Q$ , and  $\mathbf{F}^m(\omega)$  is the load vector with entries  $\mathbf{F}_q^m(\omega) = -\kappa(\bar{\phi}^m, \zeta^{m,q}; \omega)$ ,  $1 \leq q \leq Q$ . Since the number of the reduced basis  $Q$  is much smaller than the number of fine grid points,  $(H/h)^d$ , (3.12) can be solved very efficiently. Finally we use  $\{\phi_{rb}^{ml}\}$  to build an approximation of the local stiffness matrix  $\mathbf{S}^m$ , denoted  $\widehat{\mathbf{S}}^m$ , as in (2.11).

To perform the KL expansion (3.10), we apply the stochastic collocation method to estimate the mean and covariance. Due to the locally low dimensionality, the SC method requires many fewer samples than the MC method does, and thus it accelerates the offline computation significantly. We have the following theorem, which guarantees the accuracy and efficiency of the reduced basis method.

**THEOREM 3.6.** Suppose we take the first  $Q$  KL modes  $\{\zeta^{m,q}(x)\}_{q=1}^Q$  as the reduced basis functions and use the Galerkin method to obtain the reduced basis solution of (2.10), denoted as  $\phi_{rb}^m(x, \omega)$  in (3.11). Assume that

1. there exists  $C_1 > 0$  and  $\beta > 1$  such that  $\lambda_j \leq C_1 \beta^{-j}$ , and
2. there exists a constant  $C_2$  such that  $\kappa(x, \omega) \leq C_2 \bar{\kappa}(x)$  for all realizations  $\omega \in \Omega$ .

Then we have for any  $\varepsilon > 0$  the following:

(a)

$$(3.13) \quad P \left[ \|\phi^m(x, \omega) - \phi_{rb}^m(x, \omega)\|_{H_0^1(D_H^m; \bar{\kappa})} \geq \varepsilon \right] \leq \frac{C_1 C_2^2 \beta^{-Q}}{(\beta - 1) \varepsilon^2};$$

---

<sup>2</sup> $H_0^1(D_H^m; \bar{\kappa})$  is the function space  $\{v \in H_0^1(D_H^m) : \int_{D_H^m} \bar{\kappa} |\nabla v|^2 dx \leq \infty\}$  with inner product  $\langle u, v \rangle = \int_{D_H^m} \bar{\kappa} \nabla u \cdot \nabla v dx$ .

(b) for any  $l, l' \in \{1, 2, \dots, L\}$

$$(3.14) \quad P \left[ |\mathbf{S}_{ll'}^m(\omega) - \widehat{\mathbf{S}}_{ll'}^m(\omega)| \geq 2C_\phi \sqrt{\kappa_{max}} \varepsilon \right] \leq \frac{2C_1 C_2^2 \beta^{-Q}}{(\beta - 1) \varepsilon^2}.$$

Here,  $H_0^1(D; \kappa)$  is the Hilbert space with norm  $\int_D \kappa(x, \omega) |\nabla f(x)|^2 dx$ ;  $C_\phi$  is a constant that only depends on the local domain  $D_H^m$  and the deterministic boundary condition  $p^{ml}$  for  $\phi^{ml}$ .

We point out that the first assumption in Theorem 3.6 holds true in general. Moreover, under the same assumptions in Lemma 3.2, we can prove that the smaller  $c_k$  is, the bigger  $\beta$  is. This exponential decay is also observed in our numerical examples; see Figure 7. The second assumption is also valid in general. We will demonstrate that this assumption is satisfied in our numerical example with a high contrast random medium; see section 5.2. We will not present the proof of Theorem 3.6 in this paper, but refer the interested reader to [56].

Theorem 3.6 guarantees that for any prespecified  $\varepsilon > 0$  and  $\delta > 0$ , with only  $Q = \mathcal{O}(\log(1/\varepsilon) + \log(1/\delta))$  reduced basis functions, our reduced basis approximation  $\widehat{\mathbf{S}}^m$  is  $\mathcal{O}(\varepsilon)$  accurate with probability at least  $1 - \delta$ . There are two ways to deal with the rare event when our approximation is not guaranteed to be accurate. In the first approach, with an appropriate a posteriori error estimate (see [49]), we are able to efficiently detect the small-probability failure samples and recompute these samples directly to make sure they are accurate. In the second approach, we do not care about this rare event at all because this small probability error only introduces a small error in estimating statistical properties of the solution  $u_H$ . In this paper, we take the second approach; see also [34, 6].

*Remark 3.7.* In the case when the distribution of local parameters  $\xi_m$  is unknown, we need to choose an auxiliary distribution, its density denoted as  $\bar{\rho}_m(\xi_m)$ , to do the KL expansion (3.10). In practice, one can just use the uniform distribution for bounded variables. For unbounded variables, one can choose a sufficient large square domain, which covers the range of local parameters  $\xi_m$  with high probability, and then use the uniform distribution on the square domain. Theoretically, we can prove that when there exists a constant  $C > 0$  such that  $\rho_m / \bar{\rho}_m \leq C$ , our reduced basis approximation  $\widehat{\mathbf{S}}^m$  is still  $\mathcal{O}(\varepsilon)$  accurate with probability at least  $1 - \delta$  with only  $Q = \mathcal{O}(\log(1/\varepsilon) + \log(1/\delta))$  reduced basis functions. In practice, one can simply take  $Q = 3K_m$ , where  $K_m$  is the local stochastic dimension.

**3.2.2. Exploring the affine structure of the coefficient for further speed-up.** In the online stage, the reduced basis method described above still evaluates  $\phi^m(x, \omega)$  in a pointwise manner according to (3.10) and assembles local stiffness matrix (2.11) on the fine grid. Even in the case when  $\phi^m(x, \xi_m)$  is prepared by gPC expansions, as in [7], the evaluation of  $\phi^m(x, \omega)$  and numerical integration are still performed on the fine grid, which offers little computational savings compared with the multigrid method. To make the upscaling step more efficient, we assume that the parametrization of  $\kappa(x, \xi_m)$  is affine with respect to the parameters  $\xi_m$ . With this assumption, we can precompute the essential part of the stiffness matrix in the offline stage, which leads to considerable savings in assembling the stiffness matrix for each sample in the online stage. Specifically, we assume that the local random coefficient

$\kappa(x, \xi_m)$  can be expressed as follows:

$$(3.15) \quad \kappa(x, \xi_m) = \sum_{k=1}^{K_m} \xi_{m,k} \kappa_{m,k}(x), \quad x \in D_H^m.$$

By applying the affine structure of the coefficient, we obtain

$$(3.16) \quad \mathbf{A}^m(\omega) = \sum_{k=1}^{K_m} \xi_{m,k}(\omega) \mathbf{A}_k^m, \quad \mathbf{F}^m(\omega) = \sum_{k=1}^{K_m} \xi_{m,k}(\omega) \mathbf{F}_k^m,$$

where the deterministic coefficients  $\mathbf{A}_k^m$  and  $\mathbf{F}_k^m$  are given by  $\mathbf{A}_{k,qq'}^m = \kappa_{m,k}(\zeta^{m,q}, \zeta^{m,q'})$ ,  $1 \leq q, q' \leq Q$ , and  $\mathbf{F}_{k,q}^m = -\kappa_{m,k}(\bar{\phi}^m, \zeta^{m,q})$ ,  $1 \leq q \leq Q$ . We can precompute  $\mathbf{A}_k^m$  and  $\mathbf{F}_k^m$  and efficiently assemble the stiffness matrix  $\mathbf{A}^m(\omega)$  and load vector  $\mathbf{F}^m(\omega)$  for each sample. We remark that the affine structure also simplifies the assembling of local stiff matrix  $\mathbf{S}^m$  and loading vector  $\mathbf{b}^m$ .

Finally, we summarize the offline and online implementations of the random interpolation method in Algorithms 3 and 4.

---

**Algorithm 3** The offline stage of the reduced basis method.

---

- 1: Partition the physical domain  $D$  into coarse grid blocks  $\mathcal{T}_H$ .
  - 2: **for** each patch  $D_H^m$  **do**
  - 3:     Solve local multiscale basis (2.10) with affine coefficient (3.15) by the SC method to obtain samples  $\{\phi^{ml}(x, \xi_{m,c})\}_{c=1}^{N_c}$ .
  - 4:     Apply a KL expansion to get reduced basis  $\{\zeta^{ml,q}\}_{q=1}^Q$ .
  - 5: **end for**
- 

---

**Algorithm 4** The online stage of the reduced basis method for specific parameter configuration  $\xi$ .

---

- 1: **for** each patch  $D_H^k$  **do**
  - 2:     Determine the values of local variables  $\xi_m$ .
  - 3:     Assemble and solve local reduced systems (3.12).
  - 4:     Assemble the local upscaled stiffness matrix  $\hat{\mathbf{S}}^m$  and loading vector  $\hat{\mathbf{b}}^m$ .
  - 5: **end for**
  - 6: Assemble and solve the upscaled system (2.12) to obtain  $\hat{u}_H(x, \xi)$ .
- 

**3.2.3. Complexity analysis.** Using the same notation and assumptions as in section 3.1.2, we analyze the computational cost of the StoMsFEM with local reduced basis in this section. As for the random interpolation method, the computational cost consists of offline and online parts.

The offline cost consists of three parts: obtaining samples for local cell problems (2.10), performing the KL expansion to get reduced basis (3.10), and assembling upscaled quantities. At the coarse-grid level, we have about  $M$  quadrilateral coarse-grid elements, and on each element we solve  $N_c$  samples to do the KL expansion (3.10). The cost of obtaining these solution samples is  $MN_c(\eta H/h)^{d\gamma}$ . The cost of obtaining the first  $Q$  KL modes in (3.10) is of order  $N_f N_c \log(Q) + Q^2(N_f + N_c)$ ; see [26, 8]. Since  $N_f = (\eta H/h)^d$ , this part of the cost is  $\mathcal{O}((N_c \log(Q) + Q^2)(\eta H/h)^d + Q^2 N_c)$ . Finally, the cost of assembling the upscaled stiffness matrix and the loading vector is

about  $(\sum_{m=1}^M K_m Q^2)(\eta H/h)^d$ . Assuming that  $Q = \mathcal{O}(K_m)$ ,  $\eta = \mathcal{O}(1)$ , and that all the local dimensions are about the same, we have

$$\text{Cost}_{\text{offline}} \sim N_c H^{d(\gamma-1)} h^{-d\gamma} + (N_c \log(K_m) + K_m^2) h^{-d} + K_m^2 N_c H^{-d} + K_m^3 h^{-d}.$$

Therefore, the ratio between  $\text{Cost}_{\text{offline}}$  and  $\mu$  is bounded by

$$(3.17) \quad \frac{\text{Cost}_{\text{offline}}}{\mu} \lesssim N_c H^{d(\gamma-1)} + (N_c \log(K_m) + K_m^2 + K_m^3) h^{d(\gamma-1)} + K_m^2 N_c H^{d(\gamma-1)} (H/h)^{-d\gamma}.$$

For  $\gamma = 1$ , we can see that the ratio is of the order  $(N_c \log(K_m) + K_m^3)$ .

In the online stage, for a given configuration of the random parameters  $\xi$ , the cost also consists of three parts: assembling and solving the reduced basis system (3.12), assembling the local upscaled stiffness matrix and loading vector, and finally globally assembling and solving the upscaled system (2.12). For each sample, the computational cost is

$$\text{Cost}_{\text{online}} \sim M(K_m Q^2 + Q^2) + MK_m Q^2 + M + M^\gamma.$$

Here the first term comes from assembling and solving the reduced basis system (3.12); the second term is for assembling the local upscaled stiffness matrix and loading vector; and the third and forth terms are from globally assembling and solving the upscaled system (2.12). In practice, we observe that  $Q = \mathcal{O}(K_m)$ , and then the ratio between  $\text{Cost}_{\text{online}}$  and  $\mu$  is given by

$$(3.18) \quad \frac{\text{Cost}_{\text{online}}}{\mu} \lesssim \frac{1 + K_m^3 H^{d(\gamma-1)}}{(H/h)^{d\gamma}}.$$

For  $\gamma = 1$ , the ratio is about  $K_m^3 (h/H)^d$ , where  $(h/H)^d$  comes from the usage of the MsFEM and the factor  $K_m^3$  comes from assembling and solving the reduced basis systems. Again, the computational savings is more significant for  $\gamma > 1$ . Finally, when we solve  $N_{\text{on}}$  samples in the online stage, the total computational cost for this reduced basis method is

$$(3.19) \quad \frac{\text{Cost}_{\text{StoMsFEM}}}{\mu} = N_{\text{off}} + RN_{\text{on}},$$

where  $N_{\text{off}} = \mathcal{O}(N_c \log(K_m) + K_m^3)$  is the effective number of samples that we solve in the offline stage, and  $R = \mathcal{O}(K_m^3 (h/H)^{\gamma d})$  is the online computational savings achieved by the reduced basis method.

**4. Global stochastic methods.** The StoMsFEM is designed to compute an approximate solution  $\hat{u}_H(x, \xi)$  for every parameter configuration  $\xi$  efficiently, as described in Algorithms 2 and 4. It is straightforward to combine the StoMsFEM with any nonintrusive stochastic method, which determines the sample set to solve in the online stage, to finally estimate the statistical properties of the coarse-grid solution. In subsection 4.1, we combine StoMsFEM with the MC method and the (sparse-grid) SC method. In subsection 4.2, we show that to achieve the same level of estimation error, compared with the standard FEM on a fine grid, StoMsFEM indeed offers significant computational savings by optimally balancing the spatial discretization error from MsFEM and the stochastic sampling error from the global stochastic methods.

#### 4.1. Global stochastic methods.

**4.1.1. Global Monte Carlo method.** The MC method estimates statistical properties by ensemble average, i.e.,

$$(4.1) \quad \mathbb{E}[f(\omega)] \approx \mathcal{M}[f(\omega)] \equiv \frac{1}{N_{\text{on}}} \sum_{\omega_i \in \mathcal{S}} f(\omega_i),$$

where  $\omega_i$  is the  $i$ th sample and  $N_{\text{on}}$  is the total number of independent samples  $\mathcal{S}$ . This can be used to approximate the mean value or the variance of  $u_H(x, \omega)$  as

$$\begin{aligned} \mathbb{E}[u_H(x, \omega)] &\approx \mathcal{M}[u_H(x, \omega)], \\ \text{var}[u_H(x, \omega)] &\approx \mathcal{M}[(u_H(x, \omega))^2] - \{\mathcal{M}[u_H(x, \omega)]\}^2. \end{aligned}$$

For sample  $\omega_i$ , Algorithm 2 or Algorithm 4 can be applied to compute the solution  $\hat{u}_H(x, \omega_i)$  on the coarse mesh  $\mathcal{T}_H$ .

Due to the probabilistic nature of the MC estimator (4.1), we use the mean-squared error (MSE) to quantify its performance. For example, to estimate  $\mathbb{E}[u(x, \omega)]$  by  $\mathcal{M}[\hat{u}_H]$ , simple calculation shows that the MSE can be written as

$$(4.2) \quad \mathbb{E}_{\mathcal{S}} [\|\mathcal{M}[\hat{u}_H] - \mathbb{E}[u(x, \omega)]\|_2^2] = \|\mathbb{E}[\hat{u}_H(x, \omega)] - \mathbb{E}[u(x, \omega)]\|_2^2 + \frac{1}{N_{\text{on}}} \int_D \text{var}[\hat{u}_H(x, \omega)] dx.$$

Here,  $\mathbb{E}_{\mathcal{S}}$  is the expectation taken w.r.t. to the random ensemble  $\mathcal{S}$ , the first part is the spatial discretization error introduced by StoMsFEM, and the second part is the sampling error introduced by the MC method.

*Remark 4.1.* We can also consider the following two-level MC estimator [24, 4, 11]:

$$(4.3) \quad \mathcal{M}^{(2)}[\hat{u}_H] = \frac{1}{N_{\text{on},H}} \sum_{i=1}^{N_{\text{on},H}} \hat{u}_H(x, \omega_{i,H}) + \frac{1}{N_{\text{on},h}} \sum_{i=1}^{N_{\text{on},h}} [u_h(x, \omega_{i,h}) - \hat{u}_H(x, \omega_{i,h})],$$

where  $\{\omega_{i,H}\}_{i=1}^{N_{\text{on},H}}$  and  $\{\omega_{i,h}\}_{i=1}^{N_{\text{on},h}}$  are independent samples. Simple calculation shows that its MSE is

$$(4.4) \quad \begin{aligned} \|\mathbb{E}[u_h(x, \omega)] - \mathbb{E}[u(x, \omega)]\|_2^2 &+ \frac{1}{N_{\text{on},H}} \int_D \text{var}[\hat{u}_H(x, \omega)] dx \\ &+ \frac{1}{N_{\text{on},h}} \int_D \text{var}[\hat{u}_H(x, \omega) - u_h(x, \omega)] dx. \end{aligned}$$

Compared with (4.2), the two-level MC estimator is able to reduce its MSE to the order of fine-grid discretization error by properly choosing  $N_{\text{on},H}$  and  $N_{\text{on},h}$ . At the same time, its computational cost, including computing  $N_{\text{on},H} + N_{\text{on},h}$  coarse-grid solutions and  $N_{\text{on},h}$  fine-grid solutions, can be significantly smaller than that of the standard MC method. The complexity analysis and comparison with the standard MC method are provided in the last paragraph of section 4.2.

**4.1.2. Global stochastic collocation methods.** When the density  $\rho(\xi)$  is known, the SC methods [2, 43, 42, 53] may have better convergence rates when approximating the expectation (multivariate integral) in some cases. To illustrate this idea, we assume that the random variables  $\xi$  are independent. In this case, their joint density  $\rho$  factorizes as  $\rho(\xi_1, \dots, \xi_K) = \prod_{k=1}^K \rho_k(\xi_k)$ .

Similar to the MC estimator (4.1), the global SC method solves the parametrized problem (1.1) on a *deterministic* set of collocation points, denoted as  $\text{CN}(K)$ , and then approximates  $\mathbb{E}[f(\xi)]$  by some *deterministic* numerical quadrature rule, i.e.,

$$(4.5) \quad \mathbb{E}[f(\xi)] \approx \mathcal{I}[f(\xi)] \equiv \sum_{\xi \in \text{CN}(K)} w(\xi) f(\xi).$$

This can be used to approximate the mean value or the variance of  $u_H(x, \omega)$  as

$$\begin{aligned} \mathbb{E}[u_H(x, \omega)] &\approx \mathcal{I}[u_H(x, \xi)], \\ \text{var}[u_H(x, \omega)] &\approx \mathcal{I}[(u_H(x, \xi))^2] - \{\mathcal{I}[u_H(x, \xi)]\}^2. \end{aligned}$$

For the standard SC method [2],  $\text{CN}(K)$  is a tensor product grid of all the 1d collocation points. In our case when the global stochastic dimension  $K$  is large, the sparse-grid SC method [43, 42] is preferred, where  $\text{CN}(K)$  is a high dimensional sparse grid.

The locally low dimensionality can offer huge computational savings for the SC method. The key observation is that the global collocation points  $\text{CN}(K)$  repeatedly use the local collocation points. For example, the tensor product grid collocation points  $\prod_{k=1}^K \Gamma_k$ , where  $\Gamma_k$  is the collocation points in the  $\xi_k$  direction, reuse the local collocation points  $\prod_{\xi_k \in \xi_m} \Gamma_k$ , which is the local tensor product grid with the same degree. In the same manner, the local collocation points of a global sparse grid are still a sparse grid of the low dimensional local parameter space.

Therefore, the global (sparse-grid) SC method also contains the offline and online stages. In the offline stage, we take the local interpolation nodes  $\Gamma_m$  to be the local collocation points corresponding to the global (sparse-grid) collocation points  $\text{CN}(K)$ , and run Algorithm 1. Our estimate in (3.6) implies that

$$(4.6) \quad \frac{\text{Cost}_{\text{offline}}}{\mu} \sim (L-1)N_c H^{d(\gamma-1)} + \frac{(L+1)L}{2} N_c h^{d(\gamma-1)},$$

where  $N_c$  is the number of local collocation points. In the extreme case of  $\gamma = 1$ , we have  $\frac{\text{Cost}_{\text{offline}}}{\mu} \sim O(N_c)$ .

The algorithm in the online stage is almost the same as Algorithm 2, but we simply do search-and-plug-in instead of interpolation for every collocation point  $\xi \in \text{CN}(K)$ . Since searching cost is typically negligible, the online computational cost for *each collocation point* only contains assembling and solving the upscaled system (2.12); i.e.,

$$(4.7) \quad \frac{\text{Cost}_{\text{online}}}{\mu} \approx \frac{M^\gamma}{\mu} \sim (h/H)^{\gamma d}.$$

Compared with (3.7), we do not have the interpolation cost when the global (sparse-grid) SC solver is utilized. This is a big difference between the global SC solver and the global MC solver when they are combined with the local random interpolation method.

We still use  $N_{\text{on}}$  to denote the number of samples to solve in the online stage, which is  $|\text{CN}(K)|$  in the global SC solver. Therefore, the total computational cost for the StoMsFEM with the global SC solver is

$$(4.8) \quad \frac{\text{Cost}_{\text{StoMsFEM}}}{\mu} \approx N_c + RN_{\text{on}},$$

where the online savings factor  $R = (h/H)^{\gamma d}$ . On the other hand, the total computational cost for the standard FEM on the fine grid  $\mathcal{T}_h$  with a sparse-grid collocation is  $\mu N_{on}$ . Notice that  $N_c$  is the number of local sparse-grid collocation points, which is nearly negligible compared with the number of global collocation points  $N_{on}$ , thanks to the locally low dimensionality. Therefore, we get a computational savings with nearly a factor of  $R = (h/H)^{\gamma d}$ .

The estimation error of the SC method is determined by the error of the numerical quadrature, i.e.,  $\mathcal{I}[f(\xi)] - \mathbb{E}[f(\xi)]$ . For example, to estimate  $\mathbb{E}[u(x, \omega)]$  by  $\mathcal{I}[\hat{u}_H]$ , the estimation error can be bounded as follows:

$$(4.9) \quad \|\mathcal{I}[\hat{u}_H] - \mathbb{E}[u(x, \xi)]\|_2^2 \leq 2 \|\mathbb{E}[\hat{u}_H(x, \omega)] - \mathbb{E}[u(x, \omega)]\|_2^2 + 2\|\mathcal{I}[\hat{u}_H] - \mathbb{E}[\hat{u}_H(x, \omega)]\|_2^2.$$

Here, the first part is the spatial discretization error introduced by StoMsFEM, and the second part is the sampling error introduced by the SC method.

*Remark 4.2.* In a general multivariate problem, if the random variables  $\xi$  are not independent, the density  $\rho$  does not factorize, i.e.,  $\rho(\xi_1, \dots, \xi_K) \neq \prod_{k=1}^K \rho_k(\xi_k)$ . To this end, we first introduce an auxiliary probability density function  $\hat{\rho}: \mathbb{R}^K \rightarrow \mathbb{R}$  that can be seen as the joint density of  $K$  independent random variables; i.e., it factorizes as  $\hat{\rho}(\xi_1, \dots, \xi_K) = \prod_{k=1}^K \hat{\rho}_k(\xi_k)$  and satisfies  $\frac{\rho(\xi)}{\hat{\rho}(\xi)} \leq C$  for a positive constant  $C$ . For each dimension  $k = 1, 2, \dots, K$ , the 1d collocation nodes  $\mathcal{V}_k^i$  can be the Gaussian abscissas of  $\hat{\rho}_k$  or nested abscissas associated with  $\hat{\rho}_k$ . The auxiliary density  $\hat{\rho}$  should be chosen as close to the true density  $\rho$  as possible, so that the quotient  $\rho/\hat{\rho}$  remains bounded.

**4.2. Global error analysis.** The estimation error of both the MC method and the SC method consists of the spatial discretization error from StoMsFEM and the sampling error from the corresponding global stochastic methods; see (4.2) and (4.9). We should balance these two kinds of errors to achieve the optimal estimate within our budget of computing resources. To further analyze the estimation error, we assume the following estimates:

$$(4.10) \quad \|\mathbb{E}[u_h(x, \omega)] - \mathbb{E}[u(x, \omega)]\|_2^2 \lesssim h^\beta, \quad \|\mathbb{E}[\hat{u}_H(x, \omega)] - \mathbb{E}[u(x, \omega)]\|_2^2 \lesssim H^\beta,$$

$$(4.11) \quad \|\mathcal{I}(u_h) - \mathbb{E}[u_h]\|_2 \lesssim N_{on}^{-\zeta}, \quad \|\mathcal{I}(\hat{u}_H) - \mathbb{E}[\hat{u}_H]\|_2 \lesssim N_{on}^{-\zeta},$$

$$(4.12) \quad \int_D \text{var}[\hat{u}_H(x, \omega)] dx \approx \int_D \text{var}[u_h(x, \omega)] dx \approx \int_D \text{var}[u(x, \omega)] dx = c_1.$$

The rate  $\beta$  in (4.10) characterizes the discretization error from the standard FEM on fine mesh  $\mathcal{T}_h$  and MsFEM on coarse mesh  $\mathcal{T}_H$ , and  $\beta \approx 4$  in our case. The rate  $\zeta$  in (4.11) characterizes the sampling error from the (sparse-grid) SC method, and it is typically very small in our high stochastic dimension case. For some problems with moderate stochastic dimensions,  $\zeta$  can be relatively large. For example, in our high contrast example, the SC with the sparse Clenshaw–Curtis formulas, we observe  $\zeta \approx 5$ ; see Figure 10 below. We assume that  $\int_D \text{var}[u(x, \omega)] dx = \mathcal{O}(1)$  in (4.12). Error analysis of the standard FEM gives  $\|u_h - u\|_{L^2(D \times \Omega)} = \mathcal{O}(h^2)$ . For any successful upscaling method, we expect  $\|u_H - u\|_{L^2(D \times \Omega)} = \mathcal{O}(H^2)$ . For example, Theorem 2.2 validates this for the MsFEM on periodic random coefficients with period  $\epsilon \ll H$ . Other local upscaling methods [47, 41, 44] satisfy this assumption on a much richer set of random coefficients, and the StoMsFEM can be adapted to work with them. Therefore, we have  $\int_D \text{var}[u_h(x, \omega)] - \int_D \text{var}[u(x, \omega)] dx = \mathcal{O}(h^2)$  and  $\int_D \text{var}[u_H(x, \omega)] - \int_D \text{var}[u(x, \omega)] dx = \mathcal{O}(H^2)$ , and thus we validate assumption (4.12).

In (4.10), (4.11), and (4.12), we assume that  $\hat{u}_H - u_H$  is negligible. This is reasonable since we can easily drive the error  $\hat{u}_H - u_H$  below other errors due to its exponential decay implied by Theorem 3.4 in the random interpolation setting and Theorem 3.6 in the reduced basis setting. Our numerical examples also validate this assumption.

Combining the above assumptions and (4.2) and (4.9), we need  $N_{\text{on}} = \mathcal{O}(H^{-\beta})$  for the MC method and  $N_{\text{on}} = \mathcal{O}(H^{-\beta/\zeta})$  for the SC method to achieve  $\mathcal{O}(H^\beta)$  estimation error. Notice that the number of samples required remains the same for the standard FEM on a fine grid when it aims to achieve the same  $\mathcal{O}(H^\beta)$  estimation error. Since in high stochastic dimensional problems the decay rate of the physical discretization error is much faster than that of the sampling error, i.e.,  $\beta$  and  $\beta/\zeta$  is large, the number of samples to be solved in the online stage is huge. For example, the MC method requires about 100,000,000 samples when we take  $H = 0.01$  for a physical domain with  $\mathcal{O}(1)$  size. Compared with this huge number, the effective number of samples  $N_{\text{off}}$  in the StoMsFEM offline stage, which is roughly equal to the number of local interpolation nodes, is negligible. Therefore, to achieve  $\mathcal{O}(H^\beta)$  estimation error, the total computational cost ratio between StoMsFEM and the standard FEM on fine mesh, i.e.,  $N_{\text{off}}/N_{\text{on}} + R$ , is nearly  $R$ . As we derived in the previous sections,  $R$  is  $\mathcal{O}(N_c(h/H)^{\gamma d})$  for the random interpolation method,  $\mathcal{O}(K_m^3(h/H)^{\gamma d})$  for the reduced basis method, and  $\mathcal{O}((h/H)^{\gamma d})$  for the global SC method.

If we want to reduce the estimation error to the level of  $\mathcal{O}(h^\beta)$ , we can combine StoMsFEM with the two-level MC estimator (4.3). Similar to the multilevel Monte Carlo (MLMC) method, we reduce the variance part in (4.4) to  $\mathcal{O}(h^\beta)$  while optimally distributing computing resources to the coarse- and fine-grid computations. If we assume that  $\int_D \text{var}[\hat{u}_H(x, \omega) - u(x, \omega)] dx \lesssim H^\alpha$  characterizes the variance reduction effect of  $\hat{u}_H$ , the ratio of total computation cost between this two-level MC estimator and the MC based on the standard FEM on the fine grid is  $\mathcal{O}((R^2 + H^{\alpha/2})^{1/2})$ , where  $R$  is the cost ratio as before.

**5. Numerical examples.** In this section we demonstrate the accuracy and efficiency of the proposed StoMsFEM. All our computations are performed using 64-bit MATLAB R2015a on an Intel(R) Core(TM) i7-3770 (3.40 GHz).

**5.1. Patch study of a synthetic 2d example.** This synthetic example is adopted from problems with porous media [20] where the medium contains some channels and inclusions:

$$\kappa(x, \omega) = 0.2 + 0.2 \sin(\pi x) \sin(\pi y) + \sum_{k=1}^{20} \kappa_m(x) \xi_m.$$

Here the first two terms give the background of the medium, and in the summation  $\kappa_m(x)$  are the characteristic functions representing the channels/inclusions and  $\xi_m$  are the associated random variables. In our computation we set them uniformly distributed in  $[0, 1]$ . We plot one sample of the medium in Figure 1 and show the mean and the variance of the medium. It is easy to see that the medium contains many small inclusions, making the multiscale treatment necessary. For this 2d problem, we first decompose it into  $16 \times 16$  coarse grid elements. The oversampling ratio is chosen as  $\eta = 2$ , meaning each patch is enlarged by 2 in each dimension for the oversampling. In Figure 2 we plot the number of random variables in each patch. As shown in Figure 2, in each patch there are about 2 to 3 random variables.

In this example, we only show how to apply the random interpolation method

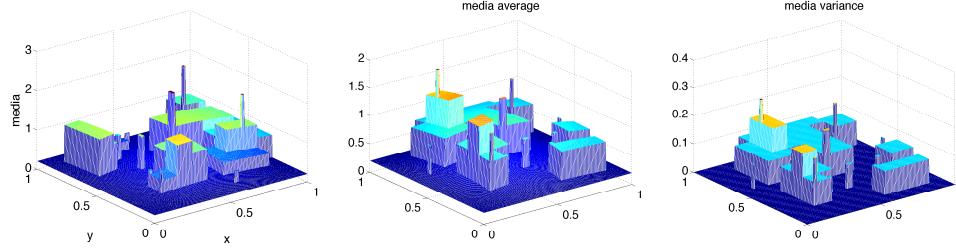


FIG. 1. One sample each of the media (left), the mean (middle), and the variance of the media (right).

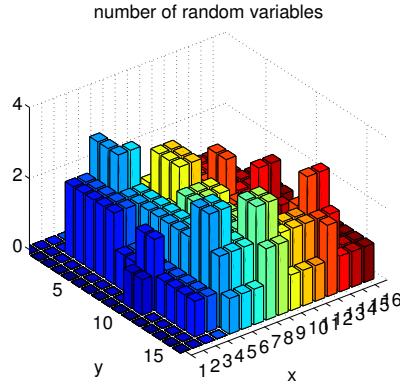


FIG. 2. The number of random variables seen in each patch. The domain is decomposed into  $16 \times 16$  coarse grid elements.

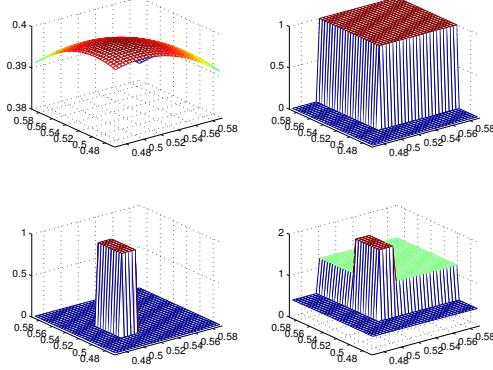


FIG. 3. Here we show the medium confined in patch (9,9). The upper left panel presents the background media, and in the upper right and lower left panels we plot the two physical modes associated with two random variables. The lower right panel shows the sample media confined in this patch.

and the reduced basis method on a local patch and study their performances. We will show the full process of the StoMsFEM on more realistic examples later. Let us pick patch (9,9) for example; two random variables are present in this patch, as shown in Figure 3, and thus the local stiffness matrix and the local load vector are functions of only two random variables.

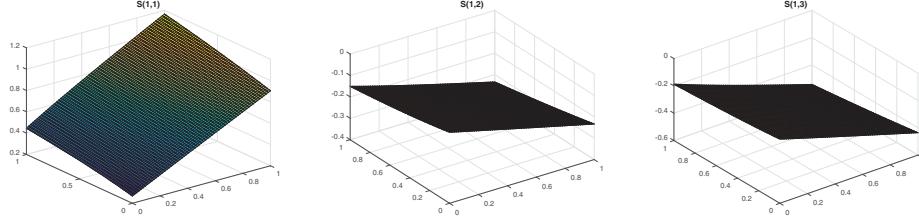


FIG. 4. The dependence on the random variables of the (1, 1) (left), (1, 2) (middle), and (1, 3) (right) entries in the local stiffness matrix, when confined in patch (9, 9). Both random variables are uniformly distributed in  $[0, 1]$ .

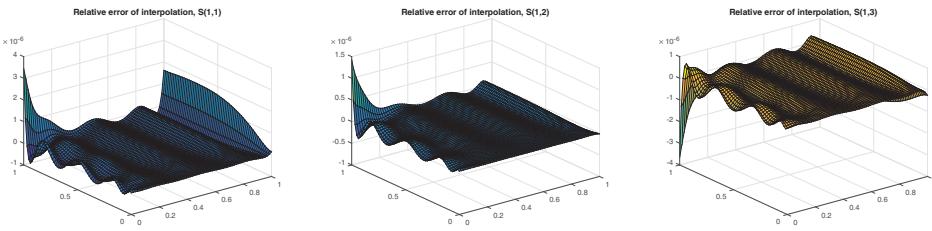


FIG. 5. The relative interpolation error for the (1, 1) (left), (1, 2) (middle), and (1, 3) (right) entries of the local upscaled stiffness matrix is in the order of  $10^{-6}$ . Its contribution to the final estimation error is negligible compared with the spatial discretization error and sampling error.

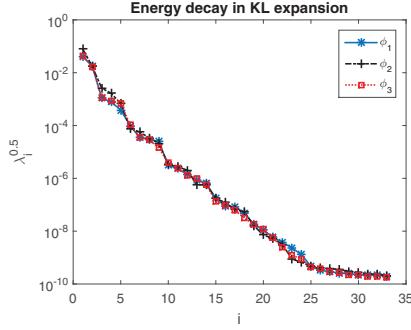


FIG. 6. The fast KL energy decay for the solutions of the local cell problems (2.10), when confined to patch (9, 9). We truncate at  $10^{-6}$  to obtain the reduced basis functions.

In the offline step, we use both the random interpolation method and the reduced basis method to construct approximations for the upscaled local stiffness matrices. In Figure 4 we plot the (1, 1), (1, 2), and (1, 3) entries of the stiffness matrix's dependence on the two random variables. For the random interpolation method, we take nine Chebyshev nodes along each dimension and take their tensor products. In Figure 5 we plot the relative interpolation error for the (1, 1), (1, 2), and (1, 3) entries. The relative interpolation error is on the order of  $10^{-6}$ , whose contribution to the final estimation error is negligible compared with the spatial discretization error and sampling error. For the reduced basis method, we perform the KL expansion of the three basis functions, obtain their reduced basis functions, and precompute the relevant quantities. In Figure 6, we show the fast energy decay in the KL expansion

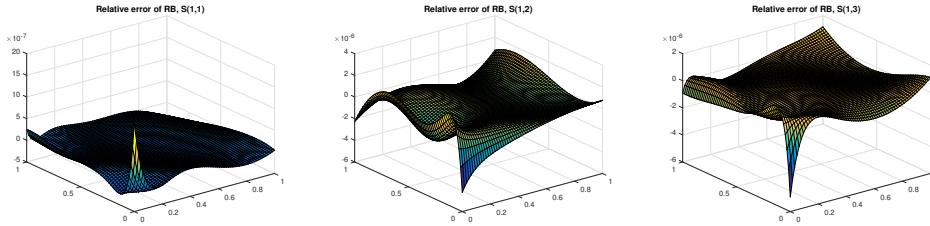


FIG. 7. The relative error for the (1, 1), (1, 2), and (1, 3) entries of the local upscaled stiffness matrix is in the order of  $10^{-6}$ .

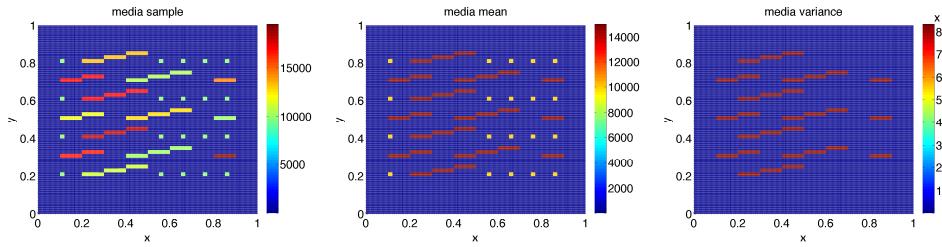


FIG. 8. Left: one sample medium. Middle: medium mean. Right: medium variance. There are 13 high permeability (of order  $10^4$ ) channels in the  $x$ -direction and a few high permeability inclusions. The background permeability is of order 1.

of these basis functions. We truncate the KL expansion at  $\sqrt{\lambda_Q/\lambda_1} < 10^{-6}$ , resulting in 15, 13, and 15 basis functions for  $\phi^1, \phi^2$ , and  $\phi^3$ , respectively. In Figure 7 we see that the relative error to compute the (1, 1), (1, 2), and (1, 3) entries is also on the order of  $10^{-6}$ .

It is worth mentioning that the oversampling domain is four times bigger than the original patch. If the effective region for a random variable falls in the boundary layer region, the associated random variable shows limited impact on the stiffness matrix. This anisotropic property suggests that we can do interpolation on dimension-adaptive grids to reduce the number of interpolation nodes when the local dimension grows. We will use dimensional-adaptive grids in our next two examples. We point out that the local reduced basis approach automatically detects this anisotropic property and always gives the most important basis functions for the local cell problems (2.10).

**5.2. Example with high contrast random medium.** The random medium of this example contains a nonconstant global background, channels with high permeability and localized inclusions. One sample and the statistical properties of the random media are shown in Figure 8. We can see that there are several high permeability channels in the  $x$ -direction and some high permeability inclusions. Utilizing the ISMD presented in section 2.1.2, we parametrize this random medium as

$$(5.1) \quad \kappa(x, y, \omega) = f_0(x, y) + \xi_0(\omega) + \sum_{k=1}^{13} f_k(x, y) \xi_k(\omega).$$

Here,  $\xi_0$  is the global random variable uniformly distributed in  $[0, 1]$  corresponding to the low permeability background, and  $\{\xi_k\}_{k=1}^{13}$  are independent random variables uniformly distributed in  $[10^4, 2 \times 10^4]$  corresponding to the high permeability channels. In this problem, we incorporate the StoMsFEM with both the MC method and

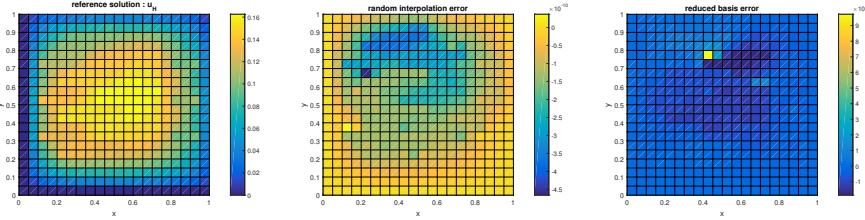


FIG. 9. Left: One sample solution from direct MsFEM. Middle: The absolute error of the approximation by the random interpolation, which is of order  $10^{-10}$ . Right: The absolute error of the approximation by the reduced basis method, which is of order  $10^{-9}$ .

the sparse-grid SC method. This demonstrates that our StoMsFEM can be easily combined with most nonintrusive global stochastic methods.

We use the MsFEM in [28], with oversampling and linear boundary conditions. We point out that the MsFEM is not the best local upscaling method for high contrast coefficients, and it is not guaranteed to have error estimate  $\|u_H - u\|_{H_1}$  small. Local upscaling methods specifically designed for high contrast problems can be found at [9, 18, 46]. In this paper, we focus on the accuracy of the proposed random interpolation and the reduced basis method, i.e.,  $\hat{u}_H - u_H$ , instead of the accuracy of the upscaling method, i.e.,  $u_H - u$ .

In the physical domain  $[0, 1]^2$ , we have a uniform coarse mesh  $\mathcal{T}_H$  with mesh size  $H_x = H_y = 0.05$  and a fine mesh  $\mathcal{T}_h$  with  $h_x = h_y = 0.0025$ . Due to the high contrast permeability, we take a relatively large oversampling ratio  $\eta = 3$ . Thanks to ISMD, the local stochastic dimensions of the parametrization (5.1) are small, typically 2 or 3, on these oversampling local coarse-grid elements. As in the example of section 5.1, we use both the random interpolation method and the reduced basis method to construct approximations for local upscaled stiffness matrices. Due to the locally low dimensionality, we achieve negligible errors when approximating the local upscaled matrix  $\mathbf{S}^m$  with a small number of interpolation nodes or reduced basis functions. For example, when we approximate  $\mathbf{S}_{1,1}^m$  on patch (14,9), the random interpolation method achieves  $\mathcal{O}(10^{-6})$  relative error with Chebyshev interpolation on a  $5 \times 16$  Chebyshev grid, while the reduced basis method also achieves  $\mathcal{O}(10^{-6})$  relative error with only seven reduced basis functions for all  $\phi^{ml}$ 's. The error plots look similar to Figures 5 and 7 in our patch-study example, and we do not show them here.

In the global MC solver, since the basis functions constructed from oversampling are nonconforming, we apply the Petrov–Galerkin MsFEM formulation [30] with the standard bilinear basis on the coarse mesh as test functions.

In the first experiment, we set the source  $b(x) = 1$  and a zero Dirichlet boundary condition. In Figure 9 we show one sample solution directly computed by the MsFEM, i.e.,  $u_H$ , and the absolute error of the StoMsFEM approximating solutions  $\hat{u}_H$ . We can see that  $\hat{u}_H - u_H$  is of the order  $10^{-10}$  for the random interpolation method and  $10^{-9}$  for the reduced basis method, which is negligible compared with the spatial discretization error  $u_H - u$ . Therefore, we can treat the approximating solution  $\hat{u}_H$  as the solution  $u_H$  computed directly by the MsFEM. However, their computational times are very different. Table 1 shows CPU times and the actual computational cost ratio. In our setting, the grid size ratio  $\eta H/h = 60$ , and on average we use  $N_c \approx 50$  local interpolation points. Theoretically, we have  $1/R = \mathcal{O}((\eta H/h)^\gamma / N_c) = \mathcal{O}(72)$  for the random interpolation method. The savings we observe is purely on the order of  $\mathcal{O}((\eta H/h)^\gamma)$ , because the cost to evaluate interpolants is negligible in practice. For

TABLE 1  
*Computational cost for one sample solution (units: s).*

Naive MsFEM	StoMsFEM (random interpolation)	StoMsFEM (reduced basis)
27.38	0.0133 (1/R = 2060)	0.0814 (1/R = 336)

the reduced basis method with  $K_m = 2$ , we have  $1/R = \mathcal{O}((\eta H/h)^\gamma / K_m^3) = \mathcal{O}(450)$  theoretically, which matches what we observed numerically.

To balance the spatial discretization error and sampling error as discussed in section 4.2, we need about  $N_{\text{on}} = \mathcal{O}(H^{-4}) = \mathcal{O}(10^5)$  for MC sampling. Therefore, we estimate the mean and standard deviation of  $\hat{u}_H$  by both the random interpolation method and the reduced basis method on the same set of  $10^5$  independent samples. The difference between these two methods is of the order  $10^{-9}$ , confirming again that the error introduced by the random interpolation and reduced basis method is negligible. To compute these  $10^5$  samples, it takes 1329 seconds for the random interpolation approximation and 8146 seconds for the reduced basis method. If we directly compute these samples by the MsFEM, it would take  $2.7 \times 10^6$  seconds.

In the second experiment, we combine our StoMsFEM with the sparse-grid SC method. Since variables  $\{\xi_k\}_{k=0}^{13}$  are independent and  $u_H$  is smooth with respect to  $\{\xi_k\}_{k=0}^{13}$ , we can implement the sparse-grid SC method to estimate  $\mathbb{E}[u_H]$  and  $\text{var}[u_H]$ . The dimension-adaptive sparse grid integration is performed on the sparse grid toolbox [37, 36], and the dimension-adaptive degree is set to be 0.6. As described in section 4.1.2, we prepare the local upscaled stiffness matrices at the local sparse grid collocation nodes in the offline stage. Since the local dimensions are small, the biggest number of local collocation nodes is only 1073. It takes about 6700 seconds to finish the offline computation. In the online stage, each sample takes only  $9.8 \times 10^{-3}$  seconds because we only look up the precomputed dictionaries to get the local upscaled quantities. The numerical cost ratio  $1/R = 2800$ , which exactly matches our theoretical estimation  $1/R = \mathcal{O}((\eta H/h)^\gamma) = \mathcal{O}(3600)$  for the global (sparse-grid) SC method.

We implement the sparse-grid integration using both the trapezoidal rule and the Clenshaw–Curtis formulas [21]. Taking the sparse-grid integration with the Clenshaw–Curtis formulas with 50433 collocation nodes as the reference  $\mathbb{E}[u_H]$ , we define the quadrature estimation error as

$$(5.2) \quad e_{\text{quad}}(\mathcal{I}[u_H]) := \|\mathcal{I}[u_H] - \mathbb{E}[u_H]\|_2.$$

In Figure 10, we compare the performance of the MC method, the SC method with the sparse-grid trapezoidal rule (piecewise linear), and the sparse Clenshaw–Curtis formulas (Chebyshev). Due to the smoothness of  $u_H$  with respect to  $\xi$ , the SC with the sparse Clenshaw–Curtis formulas has the best convergence rate ( $\approx 2.64$  from linear regression), the SC with the sparse-grid trapezoidal rule has convergence rate about 1.5, and that of the MC method is only about 0.65. Due to the stochastic nature of the MC estimator, we can see that its estimation error oscillates while slowly decreasing.

In Table 2, we list different parts of CPU times for the most accurate SC with the sparse Clenshaw–Curtis formulas in Figure 10, which has about 49805 collocation points. Since the standard FEM on the fine grid takes 27.38 seconds per sample, our SC method based on the sparse representation will have computational savings as long as the total collocation points  $N_{\text{on}} \geq \frac{N_{\text{off}}}{1-R} = 246$ , which is obviously true in our case.

In the third experiment, we reuse the offline computation above to explore the

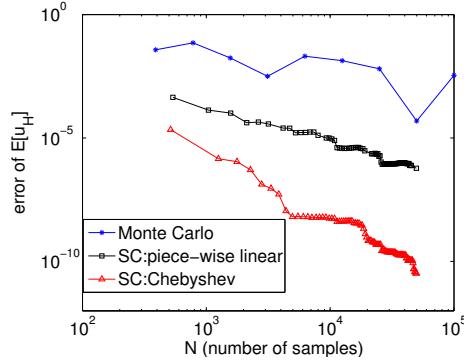


FIG. 10. Comparison between the MC method, the SC with the sparse-grid trapezoidal rule (SC:piecewise linear), and the sparse Clenshaw-Curtis formulas (SC:Chebyshev). It shows that sparse-grid collocation with high order quadrature rules is superior in this example.

TABLE 2  
Computation cost for SC on a sparse grid (units: s).

Offline	Online	Online per sample
6700	488	0.0098 (1/R = 2800)

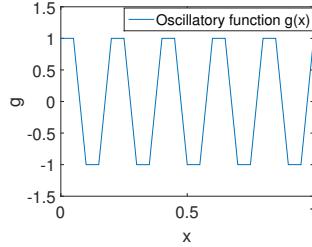


FIG. 11. Oscillatory function applied to the  $x$ -direction or  $y$ -direction.

anisotropic property of this random media. We first set

$$(5.3) \quad \begin{aligned} u(x, y, \omega) |_{x=0.1} &= u(x, y, \omega) |_{x=0.9} = g(y), \\ \mathbf{n} \cdot \nabla u &= 0 \quad \text{on } \partial\mathcal{D}, \end{aligned}$$

where  $g(y)$  is the oscillatory function shown in Figure 11. With zero-source term, we get one solution  $u_H(x, y, \omega)$ .

We then set the same zero-Neumann boundary condition and zero-source term but specify the oscillatory function in the  $y$ -direction,

$$(5.4) \quad u(x, y, \omega) |_{y=0.1} = u(x, y, \omega) |_{y=0.9} = g(x),$$

and get another solution  $u_H(x, y, \omega)$ .

We compute  $10^5$  samples for each example and compare their means in Figure 12. Because the results of the two approximations are visually the same, we show only the results from the random interpolation method. Since high conductivity channels are presented along the  $x$ -direction, the medium behaves as a homogeneous medium in the first setting but shows high conductivity in the second setting. Note that our local

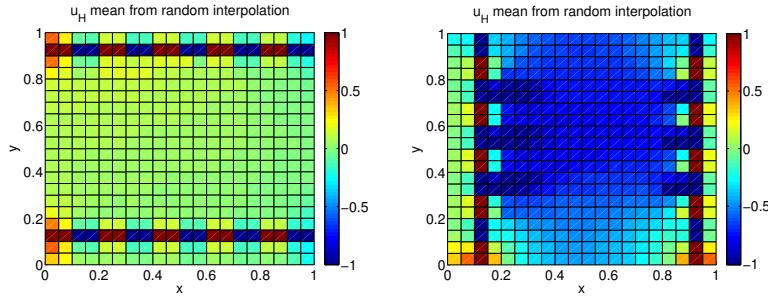


FIG. 12. Left: Oscillatory condition applied in the  $x$ -direction. Right: Oscillatory condition applied in the  $y$ -direction. The medium behaves as a homogeneous medium in the first setting but shows high conductivity in the second setting.

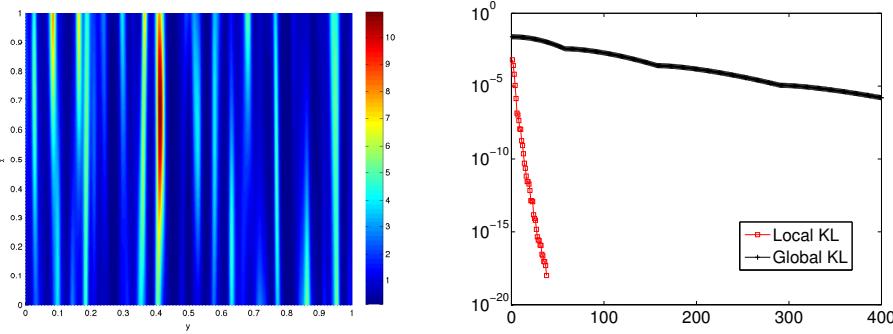


FIG. 13. Left: One sample of the anisotropic random media. Right: The eigenvalues of the global and local KL expansion for covariance function  $\text{Cov}_\beta$ .

upscaled quantities  $\mathbf{S}^m$  are independent of the boundary conditions and the source functions, and thus we can reuse them for different settings.

**5.3. A 2d example with short correlation length.** In this example, we consider the 2d elliptic problem (1.1) in the physical domain  $D = [0, 1]^2$  with the source  $b(x_1, x_2) = 2 + x_1 x_2$  and the zero Dirichlet boundary condition. The random medium  $\kappa(\mathbf{x}, \omega)$  is given as

$$(5.5) \quad \kappa(\mathbf{x}, \omega) = 0.1 + \exp(\beta(\mathbf{x}, \omega)),$$

where  $\beta(\mathbf{x}, \omega)$  is a Gaussian random field with zero mean and a Gaussian covariance function,

$$(5.6) \quad \text{Cov}_\beta(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{|x_1 - y_1|^2}{l_1^2} - \frac{|x_2 - y_2|^2}{l_2^2}\right), \quad l_1 = 1, l_2 = \frac{1}{64}.$$

Here, we have different correlation lengths in  $x_1$  and  $x_2$  directions to model the anisotropic media. A sample of the random media is shown in Figure 13(left). We can clearly see the small scales in the  $x_2$ -direction due to the small correlation length.

We apply our StoMsFEM to solving this elliptic problem on a coarse mesh  $\mathcal{T}_H$  with mesh size  $H_x = H_y = 2^{-6}$ , which does not resolve the fine scales. For the local upscaling method, we use the MsFEM with oversampling ratio  $\eta = 2$  and with the

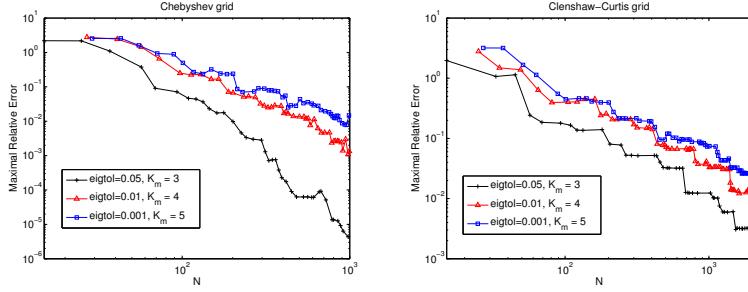


FIG. 14. Interpolation error of  $\mathbf{S}_{1,1}^m$  on sparse grids versus number of interpolation nodes.

oscillatory boundary condition under the Petrov–Galerkin formulation. We solve the local cell problems (2.10) on a fine mesh  $\mathcal{T}_h$  with mesh size  $h_x = h_y = 2^{-11}$ . For the parametrization on every local patch, i.e., every oversampling domain, we use the local KL expansion. Figure 13(right) shows the eigenvalues for the global and local KL expansions of the Gaussian covariance function  $\text{Cov}_\beta(\mathbf{x}, \mathbf{y})$ . It is obvious that the local KL expansion exhibits a much faster eigenvalue decay than the global KL expansion. In fact, to keep about 99% of the total spectrum, the global KL requires 168 terms, whereas the local KL expansion (on the oversampling domain) requires only 4 terms. The stochastic dimensionality of the global KL expansion is 168, which is too high for most gPC-based stochastic methods. On the other hand, the local stochastic dimensionality is only 4, and the random interpolation method works well in this nonaffine parametrization setting.

In the offline stage, we construct the interpolants  $\hat{\mathbf{S}}_m$  for the local upscaled stiffness matrices. Notice that the local parameters have a standard normal distribution, whose support is  $(-\infty, +\infty)$ . To make the interpolation pointwise accurate, we construct interpolants when all the parameters lie in  $[-3, 3]$ . We utilize the polynomial interpolation on the Chebyshev Gauss–Lobatto sparse grid [37, 36]. The interpolation error is estimated by the largest error among  $10^4$  randomly drawn points. The relative interpolation error of  $\mathbf{S}_{1,1}^m$  versus the number of interpolation nodes is shown in Figure 14. We show the results for three local KL expansions, which keep 95%, 99%, and 99.9% of the total spectrum and whose local dimensions are 3, 4, and 5, respectively. We also plot the interpolation error with the piecewise linear interpolation on the Clenshaw–Curtis sparse grid. We can clearly observe that with the same number of interpolation nodes, the high-order polynomial interpolation is more accurate than the piecewise linear interpolation. In Figure 14, we consider only the interpolation error, i.e.,  $\hat{\mathbf{S}}^m(\xi_m) - \mathbf{S}^m(\xi_m)$ . Given a sample of the medium, denoted as  $\kappa(x, \omega)$ , we project it onto the local KL modes, obtain the local parameters  $\xi_m$ , and truncate the small terms in the local KL expansion. The upscaled stiffness matrix  $\mathbf{S}^m(\xi_m)$  is defined based on this truncated local KL expansion. Due to this truncation, we introduce another source of error  $\mathbf{S}^m(\xi_m) - \mathbf{S}^m(\omega)$ , where  $\mathbf{S}^m(\omega)$  is the exact upscaled stiffness matrix based on the sample  $\kappa(x, \omega)$ . This truncation error is plotted in Figure 15 with respect to the number of terms we keep in the local KL expansion. The error is estimated by the largest error among  $10^4$  randomly drawn samples. We notice that the errors decay as more terms in the local KL expansion are retained. We also note that the errors decay more slowly than those in Figure 14, which implies that the predominant contribution in the overall error  $\hat{\mathbf{S}}^m(\xi_m) - \mathbf{S}^m(\omega)$  is from the truncation

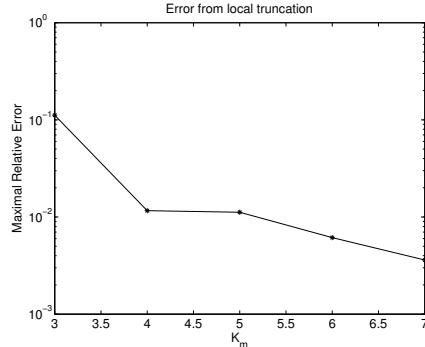


FIG. 15. Error of  $\mathbf{S}_{1,1}^m$  versus the number of terms retained in the local KL expansion.

of the local KL expansion. A theoretical result to estimate the parametrization error can be found in Corollary 2.1 in [3].

We emphasize that the errors introduced by truncation of the local KL expansions will typically dominate the errors induced by random interpolation or reduced basis method when computing  $\mathbf{S}^m(\omega)$ . As we discussed in section 3.1.1,  $\mathbf{S}^m(\xi_m)$  is smooth, and thus the random interpolation method will converge very fast. In addition, since the local dimensions are of order 1, we are able to compute reasonably high-order interpolants. The fast convergence and locally low dimensionality mean that we can easily drive the interpolation error  $\hat{\mathbf{S}}_m(\xi_m) - \mathbf{S}^m(\xi_m)$  below the error induced by the truncation of the local KL expansions, i.e.,  $\mathbf{S}^m(\xi_m) - \mathbf{S}^m(\omega)$ . Consequently, the error arising from the local KL expansion provides the leading contribution to the total error of the StoMsFEM method. Errors introduced by the parametrization, e.g., either the standard KL expansion or the local parametrization methods in section 2.1, should be considered as modeling error, since they are not directly related to the StoMsFEM algorithm. Moreover, the local parametrization methods presented in this paper have smaller parametrization error compared with the popular parametrization by the global KL expansion, because the local parametrization methods allow one to capture a greater percentage of the uncertainties due to the fast eigenvalue decay in the local KL expansion.

In the online stage, we retain four terms in the local KL expansion, which keeps 99% of the total spectrum. The interpolants for  $\mathbf{S}^m(\xi_m)$  have about 1000 interpolation nodes, and the maximal relative error is below 1%, which is smaller than the error induced by the truncated local KL expansion. For every sample, we first generate the media sample  $\kappa(x, \omega)$  by the standard spectral method. Then we project it onto the local KL modes and get the local parameters  $\xi_m$ . If all the local parameters lie in  $[-3, 3]$ , we evaluate the interpolant  $\hat{\mathbf{S}}_m$ . Otherwise, we directly solve the multiscale basis functions on the fine grid and assemble the local stiffness matrix directly from (2.10) and (2.11). In our case, the probability of doing interpolation for a local stiffness matrix is 0.9892, and with a very small probability 0.0108 the multiscale basis functions are required to solve on the fine grid. Finally, we solve the upscaled system (2.12) to get the coarse-grid solution  $\hat{u}_H$ . In Figure 16, we show the solution sample corresponding to the media sample in Figure 13. We find that the error  $\hat{u}_H - u_H$  induced by the interpolation is of order  $10^{-5}$ , while the error  $u_H(x, \xi) - u_H(x, \omega)$  induced by the truncated local KL expansion is of order  $10^{-4}$ . Therefore, we again confirm that the local truncation error is the main contribution of the overall error  $\hat{u}_H(x, \xi) - u_H(x, \omega)$ .

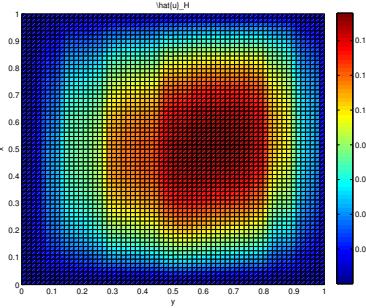


FIG. 16. The sample solution corresponding to the medium sample in Figure 13. There are several layers in the  $y$ -direction (the horizontal direction). The boundaries of these layers are exactly the low permeability strips in the medium sample.

TABLE 3  
Computational cost for one sample solution (units: s).

Naive MsFEM per sample	StoMsFEM (offline)	StoMsFEM (online per sample)
602.6104	389.3256	17.8173 ( $1/R = 34$ )

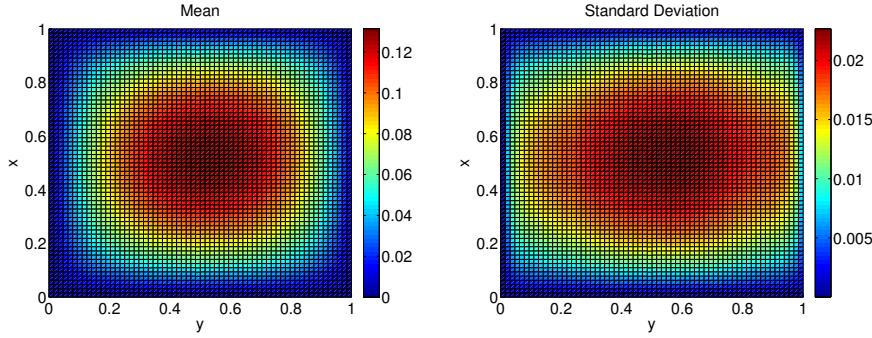


FIG. 17. The statistics of the numerical solutions using  $10^3$  samples. Note that the layered structures in sample solutions have been “homogenized” when we average over samples.

We summarize the computational cost in Table 3. The offline cost for the StoMsFEM is extremely small in this example because the random field  $\kappa(x, \omega)$  is translational invariant and we can construct interpolants for  $S^m(\xi_m)$  on only one local domain. In the online stage, the ratio of the computational cost between the naive application of the MsFEM and the StoMsFEM is about 34. Finally, we use the MC method as the global stochastic method to estimate the statistical properties of  $\hat{u}_H(x, \omega)$ . In Figure 17 we show its mean and standard deviation estimated from  $10^3$  samples. To compute these  $10^3$  samples, it takes 17431 seconds for the StoMsFEM. If we directly compute these samples by naively applying the MsFEM, it would take  $6.026 \times 10^5$  seconds. We remark that to balance the spatial discretization error and stochastic sampling error, we need  $N_{\text{on}} \approx \mathcal{O}(H^{-4}) = \mathcal{O}(10^7)$  samples, and  $10^3$  is far from being enough. This fact shows the necessity of StoMsFEM because the computational savings from StoMsFEM grows closely with the number of samples we solve in the online stage.

*Remark 5.1.* Similar to the two-level MC in Remark 4.1, the error induced by the

local KL expansion can be corrected with a little more computation. We can write the coarse-grid solution as

$$u_H(x, \omega) = \hat{u}_H(x, \xi) + u_H(x, \omega) - \hat{u}_H(x, \xi).$$

From Figure 16, we can see that  $\hat{u}_H$  serves as a good variance reduction for the true solution on the fine grid  $u_H(x, \omega)$ . With a few more samples of  $u_H(x, \omega)$ , we can correct the errors introduced both by the random interpolation and by the local KL expansion.

**6. Conclusions and future work.** We proposed a stochastic multiscale finite element method (StoMsFEM) to solve random elliptic PDEs with a high stochastic dimension. An essential difficulty in solving this type of elliptic random PDE is that we need to generate a huge number of sample solutions to get an acceptable statistical estimation and that the computational cost for every sample solution is already quite high since we need to resolve the small scale features of the solution. The StoMsFEM saves computational costs for every sample by simultaneously upscaling the stochastic solutions in the physical space for all random samples and exploring the low stochastic dimensions of the stochastic solution within each local patch.

Moreover, we proposed two effective methods to achieve this simultaneous local upscaling. The first method is the random interpolation method, which explores the high regularity of the local upscaled quantities with respect to the random variables. The second method is the reduced-order method that explores the low rank property of the multiscale basis functions within each coarse-grid element. For every sample solution, our complexity analysis shows that the cost ratio between the StoMsFEM and the standard FEM on a fine grid is  $R$ , where  $R = \mathcal{O}(N_c(h/H)^{\gamma d})$  for the random interpolation method and  $R = \mathcal{O}(K_m^3(h/H)^{\gamma d})$  for the reduced basis method. In practice, the savings is even more significant due to highly optimized fast numerical interpolation methods. In our high contrast example, we observed a factor of 2000 speed-up by the random interpolation method.

We also analyzed how different kinds of errors contributed to the final statistical estimation error. We showed that the error introduced by the interpolation or the reduced basis method is negligible, and thus we can treat our approximating solution  $\hat{u}_H(x, \xi)$  as the solution  $u_H(x, \xi)$  directly computed from MsFEM. We also showed that StoMsFEM optimally balances the spatial discretization error from MsFEM and the stochastic sampling error from the global stochastic methods. In comparison, the standard FEM on the fine grid wastes a lot of computational resources on resolving the small scale of the solution in order to reduce the spatial discretization error, while the total error is actually dominated by the stochastic sampling error. Therefore, to achieve the same level of estimation error, the StoMsFEM indeed offers a factor of  $R$  computational savings compared with the standard FEM on a fine grid.

In our last numerical example, we discussed the modeling error introduced by the local KL expansion when we parametrize the random medium. We showed that the errors introduced by truncating the local KL expansions will typically dominate the errors induced by the random interpolation or the reduced basis method. In our future work, we plan to combine the local parametrization step and the parametric local upscaling step together, and to optimally balance the modeling error from the local parametrization and that from the parametric local upscaling. We also briefly discussed the two-level Monte Carlo approach to achieving an  $\mathcal{O}(h)$  statistical estimation error; this topic will be further explored in future work.

## REFERENCES

- [1] I. BABUSKA AND R. LIPTON, *Optimal local approximation spaces for generalized finite element methods with application to multiscale problems*, Multiscale Model. Simul., 9 (2011), pp. 373–406, <https://doi.org/10.1137/100791051>.
- [2] I. BABUŠKA, F. NOBILE, AND R. TEMPONE, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM J. Numer. Anal., 45 (2007), pp. 1005–1034, <https://doi.org/10.1137/050645142>.
- [3] I. BABUŠKA, R. TEMPONE, AND G. E. ZOURARIS, *Galerkin finite element approximations of stochastic elliptic partial differential equations*, SIAM J. Numer. Anal., 42 (2004), pp. 800–825, <https://doi.org/10.1137/S0036142902418680>.
- [4] A. BARTH, C. SCHWAB, AND N. ZOLLINGER, *Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients*, Numer. Math., 119 (2011), pp. 123–161.
- [5] M. BIERI AND C. SCHWAB, *Sparse high order FEM for elliptic SPDEs*, Comput. Methods Appl. Mech. Engrg., 198 (2009), pp. 1149–1170.
- [6] X. CHEN, E.-J. PARK, AND D. XIU, *A flexible numerical approach for quantification of epistemic uncertainty*, J. Comput. Phys., 240 (2013), pp. 211–224.
- [7] Y. CHEN, J. JAKEMAN, C. GITTELSON, AND D. XIU, *Local polynomial chaos expansion for linear differential equations with high dimensional random inputs*, SIAM J. Sci. Comput., 37 (2015), pp. A79–A102, <https://doi.org/10.1137/140970100>.
- [8] M. CHENG, T. Y. HOU, M. YAN, AND Z. ZHANG, *A data-driven stochastic method for elliptic PDEs with random coefficients*, SIAM/ASA J. Uncertainty Quant., 1 (2013), pp. 452–493, <https://doi.org/10.1137/130913249>.
- [9] C.-C. CHU, I. GRAHAM, AND T.-Y. HOU, *A new multiscale finite element method for high-contrast elliptic interface problems*, Math. Comput., 79 (2010), pp. 1915–1955.
- [10] M. CI, T. Y. HOU, AND Z. SHI, *A multiscale model reduction method for partial differential equations*, ESAIM Math. Model. Numer. Anal., 48 (2014), pp. 449–474, <http://authors.library.caltech.edu/44693/>.
- [11] K. A. CLIFFE, M. B. GILES, R. SCHEICHL, AND A. L. TECKENTRUP, *Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients*, Comput. Vis. Sci., 14 (2011), pp. 3–15.
- [12] A. COHEN, R. DEVORE, AND C. SCHWAB, *Convergence rates of best  $n$ -term Galerkin approximations for a class of elliptic sPDEs*, Found. Comput. Math., 10 (2010), pp. 615–646, <https://doi.org/10.1007/s10208-010-9072-2>.
- [13] A. COHEN, R. DEVORE, AND C. SCHWAB, *Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's*, Anal. Appl., 9 (2011), pp. 11–47, <https://doi.org/10.1142/S0219530511001728>.
- [14] A. d'ASPREMONT, L. EL GHAOUI, M. I. JORDAN, AND G. R. G. LANCKRIET, *A direct formulation for sparse PCA using semidefinite programming*, SIAM Rev., 49 (2007), pp. 434–448, <https://doi.org/10.1137/050645506>.
- [15] A. DOOSTAN AND G. IACCARINO, *A least-squares approximation of partial differential equations with high-dimensional random inputs*, Journal of Computational Physics, 228 (2009), pp. 4332–4345.
- [16] A. DOOSTAN AND H. OWHADI, *A non-adapted sparse approximation of PDEs with stochastic inputs*, J. Comput. Phys., 230 (2011), pp. 3015–3034.
- [17] Y. EFENDIEV, J. GALVIS, AND T. Y. HOU, *Generalized multiscale finite element methods (GMs-FEM)*, J. Comput. Phys., 251 (2013), pp. 116–135, <https://doi.org/10.1016/j.jcp.2013.04.045>.
- [18] Y. EFENDIEV, J. GALVIS, AND X.-H. WU, *Multiscale finite element methods for high-contrast problems using local spectral basis functions*, J. Comput. Phys., 230 (2011), pp. 937–955.
- [19] P. FRAUNEFELDER, C. SCHWAB, AND R. A. TODOR, *Finite elements for elliptic problems with stochastic coefficients*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 205–228.
- [20] J. GALVIS AND Y. EFENDIEV, *Domain decomposition preconditioners for multiscale flows in high contrast media: Reduced dimension coarse spaces*, Multiscale Model. Simul., 8 (2010), pp. 1621–1644, <https://doi.org/10.1137/100790112>.
- [21] T. GERSTNER AND M. GRIEBEL, *Numerical integration using sparse grids*, Numer. Algorithms, 18 (1998), pp. 209–232.
- [22] T. GERSTNER AND M. GRIEBEL, *Dimension-adaptive tensor-product quadrature*, Computing, 71 (2003), pp. 65–87.
- [23] R. G. GHANEM AND P. D. SPANOS, *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, New York, 1991.
- [24] M. B. GILES, *Multilevel Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617.

- [25] C. GITTTELSON, *An adaptive stochastic Galerkin method for random elliptic operators*, Math. Comput., 82 (2013), pp. 1515–1541.
- [26] N. HALKO, P. G. MARTINSSON, AND J. A. TROPP, *Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions*, SIAM Rev., 53 (2011), pp. 217–288, <https://doi.org/10.1137/090771806>.
- [27] V. H. HOANG AND C. SCHWAB, *N-term Wiener chaos approximation rates for elliptic PDEs with lognormal Gaussian random inputs*, Math. Models Methods Appl. Sci., 24 (2014), pp. 797–826.
- [28] 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, <https://doi.org/10.1006/jcph.1997.5682>.
- [29] T. Y. HOU, X.-H. WU, AND Z. CAI, *Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients*, Math. Comp., 68 (1999), pp. 913–943.
- [30] T. Y. HOU, X.-H. WU, AND Y. ZHANG, *Removing the cell resonance error in the multiscale finite element method via a Petrov-Galerkin formulation*, Comm. Math. Sci., 2 (2004), pp. 185–205, <http://projecteuclid.org/euclid.cms/1109706534>.
- [31] Y. T. HOU, Q. LI, AND P. ZHANG, *A sparse decomposition of low rank symmetric positive semidefinite matrices*, Multiscale Model. Simul., 15 (2017), pp. 410–444, <https://doi.org/10.1137/16M107760X>.
- [32] Y. T. HOU AND P. ZHANG, *Sparse operator compression of elliptic operators—Part I: Second order elliptic operators*, Res. Math. Sci., submitted.
- [33] Y. T. HOU AND P. ZHANG, *Sparse operator compression of elliptic operators—Part II: High order elliptic operators*, Res. Math. Sci., submitted.
- [34] J. JAKEMAN, M. ELDRED, AND D. XIU, *Numerical approach for quantification of epistemic uncertainty*, J. Comput. Phys., 229 (2010), pp. 4648–4663.
- [35] K. KARHUNEN, *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*, Ann. Acad. Sci. Fenn. Math. Phys., Universitat Helsinki, 1947, <http://books.google.com/books?id=bGUUAQAAIAAJ>.
- [36] A. KLIMKE, *Sparse Grid Interpolation Toolbox—User’s Guide*, IANS Report 2007/017, University of Stuttgart, Stuttgart, Germany, 2007.
- [37] A. KLIMKE AND B. WOHLMUTH, *Algorithm 847: spinterp: Piecewise multilinear hierarchical sparse grid interpolation in MATLAB*, ACM Trans. Math. Software, 31 (2005), pp. 561–579.
- [38] R. LAI, J. LU, AND S. OSHER, *Density matrix minimization with  $l_1$  regularization*, Comm. Math. Sci., 13 (2015), pp. 2097–2117.
- [39] R. LAI, J. LU, AND S. OSHER, *Density matrix minimization with  $L_1$  regularization*, Comm. Math. Sci., to appear.
- [40] M. LOÈVE, *Probability Theory I*, Comprehensive Manuals of Surgical Specialties, Springer, New York, 1977.
- [41] A. MÅLQVIST AND D. PETERSEIM, *Localization of elliptic multiscale problems*, Math. Comput., 83 (2014), pp. 2583–2603.
- [42] F. NOBILE, R. TEMPONE, AND C. G. WEBSTER, *An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM J. Numer. Anal., 46 (2008), pp. 2411–2442, <https://doi.org/10.1137/070680540>.
- [43] F. NOBILE, R. TEMPONE, AND C. G. WEBSTER, *A sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM J. Numer. Anal., 46 (2008), pp. 2309–2345, <https://doi.org/10.1137/060663660>.
- [44] H. OWHADI, *Multigrid with rough coefficients and multiresolution operator decomposition from hierarchical information games*, SIAM Rev., 59 (2017), pp. 99–149, <https://doi.org/10.1137/15M1013894>.
- [45] H. OWHADI AND L. ZHANG, *Metric-based upscaling*, Comm. Pure Appl. Math., 60 (2007), pp. 675–723, <https://doi.org/10.1002/cpa.20163>.
- [46] H. OWHADI AND L. ZHANG, *Localized bases for finite-dimensional homogenization approximations with nonseparated scales and high contrast*, Multiscale Model. Simul., 9 (2011), pp. 1373–1398, <https://doi.org/10.1137/100813968>.
- [47] H. OWHADI, L. ZHANG, AND L. BERLYAND, *Polyharmonic homogenization, rough polyharmonic splines and sparse super-localization*, ESAIM: Math. Model. Numer. Anal., 48 (2014), pp. 517–552, <https://doi.org/10.1051/m2an/2013118>.
- [48] V. OZOLIŅŠ, R. LAI, R. CAFLISCH, AND S. OSHER, *Compressed modes for variational problems in mathematics and physics*, Proc. Natl. Acad. Sci. USA, 110 (2013), pp. 18368–18373, <http://www.pnas.org/content/110/46/18368.short>.
- [49] G. ROZZA, D. B. P. HUYNH, AND A. T. PATERA, *Reduced basis approximation and a posteriori*

- error estimation for affinely parametrized elliptic coercive partial differential equations*, Arch. Comput. Methods Engrg., 15 (2008), pp. 229–275.
- [50] I. SRAJ, O. P. L. MAÎTRE, O. M. KNIO, AND I. HOTEIT, *Coordinate transformation and polynomial chaos for the Bayesian inference of a Gaussian process with parametrized prior covariance function*, Comput. Methods Appl. Mech. Engrg., 298 (2016), pp. 205–228.
  - [51] T. STROUBOLIS, K. COPPS, AND I. BABUŠKA, *The generalized finite element method*, Comput. Methods Appl. Mech. Engrg., 190 (2001), pp. 4081–4193.
  - [52] V. Q. VU, J. CHO, J. LEI, AND K. ROHE, *Fantope projection and selection: A near-optimal convex relaxation of sparse PCA*, in Advances in Neural Information Processing Systems, MIT Press, Cambridge, MA, 2013, pp. 2670–2678.
  - [53] D. XIU AND J. S. HESTHAVEN, *High-order collocation methods for differential equations with random inputs*, SIAM J. Sci. Comput., 27 (2005), pp. 1118–1139, <https://doi.org/10.1137/040615201>.
  - [54] D. XIU AND G. E. KARNIADAKIS, *The Wiener–Askey polynomial chaos for stochastic differential equations*, SIAM J. Sci. Comput., 24 (2002), pp. 619–644, <https://doi.org/10.1137/S1064827501387826>.
  - [55] L. YAN, L. GUO, AND D. XIU, *Stochastic collocation algorithms using  $l_1$ -minimization*, Internat. J. Uncertainty Quant., 2 (2012), pp. 279–293.
  - [56] P. ZHANG, *Localized Decomposition of Symmetric Positive Semidefinite Operators and Its Applications*, in preparation, 2017.
  - [57] H. ZOU, T. HASTIE, AND R. TIBSHIRANI, *Sparse principal component analysis*, J. Comput. Graph. Statist., 15 (2004), pp. 265–286.