AIAA 2011-1846

52nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Conference, 4-7 Apr 2011, Denver, Colorado

A Reduced Spectral Projection Method for Stochastic Finite Element Analysis

Sondipon Adhikari [*] and Abhishek Kundu [†]

Swansea University, Swansea, United Kingdom

The stochastic finite element analysis of elliptic type partial differential equations are considered. An alternative approach by projecting the solution of the discretized equation into a finite dimensional orthonormal vector basis is investigated. It is shown that the solution can be obtained using a finite series comprising functions of random variables and orthonormal vectors. These functions, called as the spectral functions, can be expressed in terms of the spectral properties of the deterministic coefficient matrices arising due to the discretization of the governing partial differential equation. An explicit relationship between these functions and polynomial chaos functions has been derived. Based on the projection in the orthonormal vector basis, a Galerkin error minimization approach is proposed. The constants appearing in the Galerkin method are solved from a system of linear equations which has the same dimension as the original discretized equation. A hybrid analytical and simulation based computational approach is proposed to obtain the moments and pdf of the solution. The method is illustrated using a stochastic beam problem. The results are compared with the direct Monte Carlo simulation results for different correlation lengths and strengths of randomness.

I. Introduction

Due to the significant development in computational hardware it is now possible to solve very high resolution models in various computational physics problems, ranging from fluid mechanics to nano-bio mechanics. However, the spatial resolution is not enough to determine the credibility of a numerical model. The physical model as well its parameters are also crucial. Since neither of these may not be exactly known, over the past three decades there has been increasing research activities to model the governing partial differential equations within the framework of stochastic equations. We refer to few recent review papers.[1,2,3] Consider a bounded domain $\mathcal{D} \in \mathbb{R}^d$ with piecewise Lipschitz boundary $\partial\mathcal{D}$, where $d \leq 3$ is the spatial dimension. Further, consider

[*] Professor of Aerospace Engineering, School of Engineering, Swansea University, Singleton Park, Swansea SA2 8PP, UK, Email: S.Adhikari@swansea.ac.uk; AIAA Senior Member.

[†] PhD Student, Aerospace Engineering, Swansea University, Email: a.kundu.577613@swansea.ac.uk;

that $(\Omega, \mathcal{F}, P)$ is a probability space where $\omega \in \Omega$ is a sample point from the sampling space Ω, $\mathcal{F}$ is the complete σ-algebra over the subsets of Ω and P is the probability measure. We consider the stochastic elliptic partial differential equation (PDE)

$$-\nabla \left[a(\mathbf{r}, \omega) \nabla u(\mathbf{r}, \omega)\right] = p(\mathbf{r}); \quad \mathbf{r} \text{ in } \mathcal{D} \tag{1}$$

with the associated Dirichlet condition

$$u(\mathbf{r}, \omega) = 0; \quad \mathbf{r} \text{ on } \partial\mathcal{D} \tag{2}$$

Here $a : \mathbb{R}^d \times \Omega \to \mathbb{R}$ is a random field,[4] which can be viewed as a set of random variables indexed by $\mathbf{r} \in \mathbb{R}^d$. We assume the random field $a(\mathbf{r}, \omega)$ to be stationary and square integrable. Depending on the physical problem the random field $a(\mathbf{r}, \omega)$ can be used to model different physical quantities. As an example, for a slow flow of an incompressible, viscous fluid through a porus media $a(\mathbf{r}, \omega)$ would be the random field describing the permeability of the medium. The purpose of this paper is to investigate a new solution approach for Eq. (1) after the discretization using the stochastic finite element method.[5,6]

II. Overview of spectral stochastic finite element method

II.A. Discretisation of the stochastic PDE

Consider $a(\mathbf{r}, \omega)$ is a Gaussian random field with a covariance function $C_a : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined in the domain $\mathcal{D}$. Since the covariance function is square bounded, symmetric and positive definite, it can be represented by a spectral decomposition in an infinite dimensional Hilbert space. Using this spectral decomposition, the random process $a(\mathbf{r}, \omega)$ can be expressed (see for example[5,7]) in a generalized fourier type of series known as the Karhunen-Loève expansion

$$a(\mathbf{r}, \omega) = a_0(\mathbf{r}) + \sum_{i=1}^{\infty} \sqrt{\nu_i} \xi_i(\omega) \varphi_i(\mathbf{r}) \tag{3}$$

Here $a_0(\mathbf{r})$ is the mean function, $\xi_i(\omega)$ are uncorrelated standard Gaussian random variables, ν_i and $\varphi_i(\mathbf{r})$ are eigenvalues and eigenfunctions satisfying the integral equation $\int_{\mathcal{D}} C_a(\mathbf{r}_1, \mathbf{r}_2) \varphi_j(\mathbf{r}_1) \mathrm{d}\mathbf{r}_1 = \nu_j \varphi_j(\mathbf{r}_2), \quad \forall\ j = 1, 2, \cdots$. Truncating the series (3) upto the M-th term, substituting $a(\mathbf{r}, \omega)$ in the governing PDE (1) and applying the boundary conditions, the discretized equation can be written as

$$\left[\mathbf{A}_0 + \sum_{i=1}^{M} \xi_i(\omega) \mathbf{A}_i\right] \mathbf{u}(\omega) = \mathbf{f} \tag{4}$$

The necessary technical details to obtain the discrete stochastic algebraic equations from the stochastic partial differential equation (1) has become standard in the literature and therefore omitted here. Excellent references, for example[5,8,9] are available on this topic. In Eq. (4) $\mathbf{A}_0$ is a symmetric nonnegative definite matrix, $\mathbf{A}_i \in \mathbb{R}^{n \times n}; i = 1, 2, \ldots, M$ are symmetric matrices, $\mathbf{u}(\omega) \in \mathbb{R}^n$ is the solution vector and $\mathbf{f} \in \mathbb{R}^n$ in the input vector. We assume that the eigenvalues of $\mathbf{A}_0$ are distinct. For most practical application uncertainties are small compared to the deterministic values. Therefore, we normally have

$$\|\mathbf{A}_0\| \geq \left\| \sum_{i=1}^{M} \xi_i(\omega) \mathbf{A}_i \right\|; \forall\ \omega \in \Omega \tag{5}$$

Here by $\|\bullet\|$ we imply the Frobenius matrix norm,[10] defined as $\|\mathbf{A}\| = \mathrm{Trace}\left(\mathbf{A}\mathbf{A}^T\right)$ for any $\mathbf{A} \in \mathbb{R}^{n\times n}$. The number of terms M in Eq. (4) can be selected based on the 'amount of information' to be retained. This in turn is related to the number of eigenvalues retained, since the eigenvalues, ν_i, in Eq. (3) are arranged in a decreasing order. One of the main aim of a stochastic finite element analysis is to obtain $\mathbf{u}(\omega)$ for $\omega \in \Omega$ from Eq. (4) in an efficient manner and is the main topic of this paper.

II.B. Brief review of the solution techniques

The solution of the set of stochastic linear algebraic equations (4) is a key step in the stochastic finite element analysis. As a result, several methods have been proposed. These methods include, first- and second-order perturbation methods,[11,12] Neumann expansion method,[13,14] Galerkin approach[15] and linear algebra based methods.[16,17,18] More recently very efficient collocation methods have been proposed.[19,20] Another class of problems which have been used widely in the literature is known as the spectral methods (see[1] for a recent review). These methods include the polynomial chaos (PC) expansion,[5] stochastic reduced basis method and,[21] Wiener−Askey chaos expansion.[22,23] According to the polynomial chaos expansion, second-order random variables $u_j(\theta)$ can be represented by the mean-square convergent expansion

$$\begin{aligned} u_j(\omega) &= u_{i_0}^{(j)} h_0 + \sum_{i_1=1}^{\infty} u_{i_1}^{(j)} h_1(\xi_{i_1}(\omega)) \\ &+ \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1} u_{i_1,i_2}^{(j)} h_2(\xi_{i_1}(\omega),\xi_{i_2}(\omega)) + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2} u_{i_1 i_2 i_3}^{(j)} h_3(\xi_{i_1}(\omega),\xi_{i_2}(\omega),\xi_{i_3}(\omega)) \\ &+ \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2}\sum_{i_4=1}^{i_3} u_{i_1 i_2 i_3 i_4}^{(j)} h_4(\xi_{i_1}(\omega),\xi_{i_2}(\omega),\xi_{i_3}(\omega),\xi_{i_4}(\omega)) + \ldots \end{aligned} \tag{6}$$

where $u_{i_1,\ldots,i_p}^{(j)}$ are deterministic constants to be determined and $h_p(\,\xi_{i_1}(\omega),\ldots,\xi_{i_p}(\omega)\,)$ is the p^{th} order Homogeneous Chaos. When $\xi_i(\omega)$ are Gaussian random variables, the functions $h_p(\xi_{i_1}(\omega),\ldots,\xi_{i_p}(\omega)\,)$ are the p^{th} order hermite polynomial so that it becomes othronormal with respect to the Gaussian probability density function. The same idea can be extended to non-Gaussian random variables, provided more generalized functional basis are used.[22,23] When we have a random vector, as in the case of the solution of Eq. (4), then it is natural to 'replace' the constants $u_{i_1,\ldots i_p}^{(j)}$ by vectors $\mathbf{u}_{i_1,\ldots i_p}^{(j)} \in \mathbb{R}^n$. Suppose the series in truncated after P number of terms. The value of P depends on the number of basic random variables M and the order of the PC expansion r as

$$P = \sum_{j=0}^{r} \frac{(M+j-1)!}{j!(M-1)!} \tag{7}$$

After the truncation, there are P number of unknown vectors of dimension n. Then a mean-square error minimization approach can be applied and the unknown vectors can be solved using the Galerkin approach.[5] Since P increases very rapidly with the order of the chaos r and the number of random variables M, the final number of unknown constants Pn becomes very large. As a results several methods have been developed (see for example[21,24,25]) to reduce the computational cost. Here we investigate the possibility of an alternative approach, where instead of projecting the solution in the space of orthonormal polynomials, the solution is projected in an orthonormal vector basis.

III. Spectral decomposition in the vector space

Following the spectral stochastic finite element method, or otherwise, an approximation to the solution of Eq. (4) can be expressed as a linear combination of functions of random variables and deterministic vectors. Recently Nouy[26,27] discussed the possibility of an optimal spectral decomposition. The aim is to use small number of terms to reduce the computation without loosing the accuracy. Here an orthonormal vector basis is considered.

III.A. Expansion in the orthonormal vector basis

In order to propose the approach, some basic results in linear algebra are necessary. We refer to Luenberger[28] for further technical details.

Definition 1. (Linearly independent vectors) A set of vectors $\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_n\}$ is linearly independent if the expression $\sum_{k=1}^{n} \alpha_k \boldsymbol{\phi}_k = \mathbf{0}$ if and only if $\alpha_k = 0$ for all $k = 1, 2, \ldots, n$.

A finite set S of linearly independent vectors is said to be a *complete basis* for the space X if S generates X. A vector space having a finite basis is said to be finite dimensional. Here we consider finite dimensional basis vectors. Suppose H is a Hilbert space which is a subset of $\mathbb{R}^n$. For two vectors $\mathbf{u}, \mathbf{v} \in H \in \mathbb{R}^n$ we define the inner product norm as $(\mathbf{u}, \mathbf{v}) = \mathbf{u}^T\mathbf{v}$. Finite dimensional basis vectors have the following spanning property.

Remark 1. (The spanning property) Suppose $\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_n\}$ is a complete basis in the Hilbert space H. Then for every nonzero $\mathbf{u} \in H$, it is possible to choose $\alpha_1, \alpha_2, \ldots, \alpha_n \neq 0$ uniquely such that $\mathbf{u} = \alpha_1\boldsymbol{\phi}_1 + \alpha_2\boldsymbol{\phi}_2 + \ldots \alpha_n\boldsymbol{\phi}_n$.

Definition 2. (Orthonormal vectors) A set of vectors $\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_n\}$ is said to a be orthonormal set in H if $(\boldsymbol{\phi}_j, \boldsymbol{\phi}_k) = \delta_{jk}$ for any $j, k \leq n, j \neq k$ where δ_{jk} is Kroneker's delta function.

Remark 2. An orthonormal set of nonzero vectors is a linearly independent set.

Orthonormal set of vectors are preferred in a Hilbert space over any other set of linearly independent vectors. They can be created from any set of independent vectors using the Gram-Schmidt orthonormalization procedure. Here we generate the orthonormal set from the eigenvectors of a symmetric positive definite matrix.

The motivation of the proposed approach comes form the fact that any vector in H can be expanded in a *finite* basis. Therefore, fixing a value of ω, say $\omega = \omega_1$, the solution of Eq. (4) $\mathbf{u}(\omega_1)$ can be expanded in a complete basis according to Remark 1 as $\mathbf{u}(\omega_1) = \alpha_1^{(1)}\boldsymbol{\phi}_1 + \alpha_2^{(1)}\boldsymbol{\phi}_2 + \ldots \alpha_n^{(1)}\boldsymbol{\phi}_n$. Repeating this for $\omega_1, \omega_2, \ldots$ eventually the whole sample-space can be covered and it would be possible to expand $\mathbf{u}(\omega), \forall\omega \in \Omega$ as a linear combination of $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_n$. Based on this idea, one of the main contribution of this paper is the following result:

Theorem 1. *There exist a finite set of functions $\Gamma_k : (\mathbb{R}^m \times \Omega) \to (\mathbb{R} \times \Omega)$ and an orthonormal basis $\boldsymbol{\phi}_k \in \mathbb{R}^n$ for $k = 1, 2, \ldots, n$ such that the series*

$$\hat{\mathbf{u}}(\omega) = \sum_{k=1}^{n} \Gamma_k(\boldsymbol{\xi}(\omega))\boldsymbol{\phi}_k \tag{8}$$

converges to the exact solution of the discretized stochastic finite element equation (4) with probability 1.

Proof. The first step is to generate a complete orthonormal basis. We use the eigenvectors $\boldsymbol{\phi}_k \in \mathbb{R}^n$ of the matrix $\mathbf{A}_0$ such that

$$\mathbf{A}_0\boldsymbol{\phi}_k = \lambda_{0_k}\boldsymbol{\phi}_k; \quad k = 1, 2, \ldots n \tag{9}$$

We assume that the eigenvalues are distinct so that $\boldsymbol{\phi}_k$ for $k = 1, 2, \ldots n$ forms a complete orthonormal basis. Note that any in principle orthonormal basis can be sued. This choice is selected due to the analytical simplicity as will be seen later. For notational convenience, define the matrix of eigenvalues and eigenvectors

$$\boldsymbol{\Lambda}_0 = \operatorname{diag}\left[\lambda_{0_1}, \lambda_{0_2}, \ldots, \lambda_{0_n}\right] \in \mathbb{R}^{n\times n} \quad \text{and} \quad \boldsymbol{\Phi} = \left[\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_n\right] \in \mathbb{R}^{n\times n} \tag{10}$$

Eigenvalues are ordered in the ascending order so that $\lambda_{0_1} < \lambda_{0_2} < \ldots < \lambda_{0_n}$. Since $\boldsymbol{\Phi}$ is an orthogonal matrix we have $\boldsymbol{\Phi}^{-1} = \boldsymbol{\Phi}^T$ so that the following identities can be easily established

$$\boldsymbol{\Phi}^T\mathbf{A}_0\boldsymbol{\Phi} = \boldsymbol{\Lambda}_0; \quad \mathbf{A}_0 = \boldsymbol{\Phi}^{-T}\boldsymbol{\Lambda}_0\boldsymbol{\Phi}^{-1} \quad \text{and} \quad \mathbf{A}_0^{-1} = \boldsymbol{\Phi}\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Phi}^T \tag{11}$$

We also introduce the transformations

$$\widetilde{\mathbf{A}}_i = \boldsymbol{\Phi}^T\mathbf{A}_i\boldsymbol{\Phi} \in \mathbb{R}^{n\times n}; i = 0, 1, 2, \ldots, M \tag{12}$$

Note that $\widetilde{\mathbf{A}}_0 = \boldsymbol{\Lambda}_0$, a diagonal matrix and

$$\mathbf{A}_i = \boldsymbol{\Phi}^{-T}\widetilde{\mathbf{A}}_i\boldsymbol{\Phi}^{-1} \in \mathbb{R}^{n\times n}; i = 1, 2, \ldots, M \tag{13}$$

Suppose the solution of Eq. (4) is given by

$$\hat{\mathbf{u}}(\omega) = \left[\mathbf{A}_0 + \sum_{i=1}^{M}\xi_i(\omega)\mathbf{A}_i\right]^{-1}\mathbf{f} \tag{14}$$

Using Eqs. (10)–(13) and the orthonormality of $\boldsymbol{\Phi}$ one has

$$\hat{\mathbf{u}}(\omega) = \left[\boldsymbol{\Phi}^{-T}\boldsymbol{\Lambda}_0\boldsymbol{\Phi}^{-1} + \sum_{i=1}^{M}\xi_i(\omega)\boldsymbol{\Phi}^{-T}\widetilde{\mathbf{A}}_i\boldsymbol{\Phi}^{-1}\right]^{-1}\mathbf{f} = \boldsymbol{\Phi}\boldsymbol{\Psi}\left(\boldsymbol{\xi}(\omega)\right)\boldsymbol{\Phi}^T\mathbf{f} \tag{15}$$

where

$$\boldsymbol{\Psi}\left(\boldsymbol{\xi}(\omega)\right) = \left[\boldsymbol{\Lambda}_0 + \sum_{i=1}^{M}\xi_i(\omega)\widetilde{\mathbf{A}}_i\right]^{-1} \tag{16}$$

and the M-dimensional random vector

$$\boldsymbol{\xi}(\omega) = \left\{\xi_1(\omega), \xi_2(\omega), \ldots, \xi_M(\omega)\right\}^T \tag{17}$$

Now we separate the diagonal and off-diagonal terms of the $\widetilde{\mathbf{A}}_i$ matrices as

$$\widetilde{\mathbf{A}}_i = \boldsymbol{\Lambda}_i + \boldsymbol{\Delta}_i, \quad i = 1, 2, \ldots, M \tag{18}$$

Here the diagonal matrix

$$\boldsymbol{\Lambda}_i = \operatorname{diag}\left[\widetilde{\mathbf{A}}\right] = \operatorname{diag}\left[\lambda_{i_1}, \lambda_{i_2}, \ldots, \lambda_{i_n}\right] \in \mathbb{R}^{n\times n} \tag{19}$$

and the matrix containing only the off-diagonal elements $\boldsymbol{\Delta}_i = \widetilde{\mathbf{A}}_i - \boldsymbol{\Lambda}_i$ is such that $\mathrm{Trace}\left(\boldsymbol{\Delta}_i\right) = 0$. Using these, from Eq. (16) one has

$$\boldsymbol{\Psi}\left(\boldsymbol{\xi}(\omega)\right) = \left[\underbrace{\boldsymbol{\Lambda}_0 + \sum_{i=1}^{M} \xi_i(\omega)\boldsymbol{\Lambda}_i}_{\boldsymbol{\Lambda}(\boldsymbol{\xi}(\omega))} + \underbrace{\sum_{i=1}^{M} \xi_i(\omega)\boldsymbol{\Delta}_i}_{\boldsymbol{\Delta}(\boldsymbol{\xi}(\omega))}\right]^{-1} \tag{20}$$

where $\boldsymbol{\Lambda}\left(\boldsymbol{\xi}(\omega)\right) \in \mathbb{R}^{n\times n}$ is a diagonal matrix and $\boldsymbol{\Delta}\left(\boldsymbol{\xi}(\omega)\right)$ is an off-diagonal only matrix. Since $\boldsymbol{\Phi}$ is an orthogonal matrix we have $\left\|\widetilde{\mathbf{A}}_i\right\| = \left\|\boldsymbol{\Phi}^T\mathbf{A}_i\boldsymbol{\Phi}\right\| = \left\|\mathbf{A}_i\right\|$ Therefore, from Eq. (5) one has $\left\|\boldsymbol{\Lambda}_0\right\| \geq \left\|\sum_{i=1}^{M} \xi_i(\omega)\widetilde{\mathbf{A}}_i\right\|$. Using this and following Eqs. (18) and (20), we have

$$\left\|\boldsymbol{\Lambda}\left(\boldsymbol{\xi}(\omega)\right)\right\| \geq \left\|\boldsymbol{\Delta}\left(\boldsymbol{\xi}(\omega)\right)\right\| \tag{21}$$

Because $\boldsymbol{\Lambda}\left(\boldsymbol{\xi}(\omega)\right)$ is a diagonal matrix, the norm of the inverse is the same as the inverse of the norm, that is

$$\left\|\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right)\right\| = \frac{1}{\left\|\boldsymbol{\Lambda}\left(\boldsymbol{\xi}(\omega)\right)\right\|} \tag{22}$$

Using the Cauchy-Schwarz inequality and the relationship in (21) we have

$$\left\|\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right)\boldsymbol{\Delta}\left(\boldsymbol{\xi}(\omega)\right)\right\| \leq \left\|\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right)\right\| \left\|\boldsymbol{\Delta}\left(\boldsymbol{\xi}(\omega)\right)\right\| = \frac{\left\|\boldsymbol{\Delta}\left(\boldsymbol{\xi}(\omega)\right)\right\|}{\left\|\boldsymbol{\Lambda}\left(\boldsymbol{\xi}(\omega)\right)\right\|} \leq 1 \tag{23}$$

We rewrite Eq. (20) as

$$\boldsymbol{\Psi}\left(\boldsymbol{\xi}(\omega)\right) = \left[\boldsymbol{\Lambda}\left(\boldsymbol{\xi}(\omega)\right)\left[\mathbf{I}_n + \boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right)\boldsymbol{\Delta}\left(\boldsymbol{\xi}(\omega)\right)\right]\right]^{-1} \tag{24}$$

Due to the inequality relationship in Eq. (23) the eigenvalues of $\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right)\boldsymbol{\Delta}\left(\boldsymbol{\xi}(\omega)\right)$ are less than 1 and the above expression can be represented using a Neumann type of matrix series[13] as

$$\boldsymbol{\Psi}\left(\boldsymbol{\xi}(\omega)\right) = \sum_{s=0}^{\infty}(-1)^s\left[\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right)\boldsymbol{\Delta}\left(\boldsymbol{\xi}(\omega)\right)\right]^s\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right) \tag{25}$$

Taking an arbitrary r-th element of $\hat{\mathbf{u}}(\omega)$, Eq. (15) can be rearranged to have

$$\hat{u}_r(\omega) = \sum_{k=1}^{n}\Phi_{rk}\left(\sum_{j=1}^{n}\Psi_{kj}\left(\boldsymbol{\xi}(\omega)\right)\left(\boldsymbol{\phi}_j^T\mathbf{f}\right)\right) \tag{26}$$

Defining

$$\Gamma_k\left(\boldsymbol{\xi}(\omega)\right) = \sum_{j=1}^{n}\Psi_{kj}\left(\boldsymbol{\xi}(\omega)\right)\left(\boldsymbol{\phi}_j^T\mathbf{f}\right) \tag{27}$$

and collecting all the elements in Eq. (26) for $r = 1, 2, \ldots, n$ one has

$$\hat{\mathbf{u}}(\omega) = \sum_{k=1}^{n}\Gamma_k\left(\boldsymbol{\xi}(\omega)\right)\boldsymbol{\phi}_k \tag{28}$$

Now assume the series in Eq. (25) is truncated after m-th term. We define the truncated function

$$\mathbf{\Psi}^{(m)}\left(\boldsymbol{\xi}(\omega)\right)=\sum_{s=0}^{m}(-1)^s\left[\mathbf{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right)\mathbf{\Delta}\left(\boldsymbol{\xi}(\omega)\right)\right]^s\mathbf{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right) \tag{29}$$

From this one can obtain a sequence for different m

$$\hat{\mathbf{u}}^{(m)}(\omega)=\sum_{k=1}^{n}\Gamma_k^{(m)}\left(\boldsymbol{\xi}(\omega)\right)\boldsymbol{\phi}_k; \quad m=1,2,3,\ldots \tag{30}$$

Since $\omega\in\Omega$ is arbitrary, comparing (4) and (14) we observe that $\hat{\mathbf{u}}^{(m)}(\omega)$ is the solution of Eq. (4) for every ω when $m\to\infty$. This implies that

$$\text{Prob}\left\{\omega\in\Omega:\lim_{m\to\infty}\hat{\mathbf{u}}^{(m)}(\omega)=\hat{\mathbf{u}}(\omega)\right\}=1 \tag{31}$$

Therefore, $\hat{\mathbf{u}}(\omega)$ is the solution of Eq. (4) in probability 1 and this proves the theorem. □

Remark 3. (Convergence) The series in (30) approaches to the exact solution of the governing Eq. (4) for every $\omega\in\Omega$ for $m\to\infty$. For this reason it converges in probability 1. The convergence in probability 1 is a stronger convergence than, for example, mean-square convergence often used in the stochastic finite element analysis. Since the convergence in probability 1 automatically implies the mean-square convergence, the series is in Eq. (28) is also a mean-square convergent series.

Definition 3. The functions $\Gamma_k\left(\boldsymbol{\xi}(\omega)\right), k=1,2,\ldots n$ are called the spectral functions as they are expressed in terms of the spectral properties of the coefficient matrices of the governing discretized equation.

III.B. Mathematical relationship with the polynomial chaos expansion

The original polynomial chaos expansion (6) proposed by Wiener[29] for scalar functions is optimal because there is no nontrivial vector basis in $\mathbb{R}^1$. This fact is not necessarily true when the same series is directly extended to finite dimensional vectors. We therefore make difference between the original scalar polynomial chaos and vector polynomial chaos. The 'vector version' of the polynomial chaos expansion (6) have been used widely in literature. After the finite truncation, concisely such an expression can be written as

$$\hat{\mathbf{u}}(\omega)=\sum_{k=1}^{P}H_k(\boldsymbol{\xi}(\omega))\mathbf{u}_k \tag{32}$$

where in general $P\gg n$ according to Eq. (7). Equation (32) originated from projecting a random function in an othonormal functional basis. Since $\hat{\mathbf{u}}(\omega)\in\mathbb{R}^n$, an alternative view of Eq. (32) could be taken as the projection in a vector basis in $\mathbb{R}^n$. According to Remark 1, using the spanning property of a complete basis in $\mathbb{R}^n$ it is *always* possible to project $\hat{\mathbf{u}}(\omega)$ in a finite dimensional vector basis for any $\omega\in\Omega$. Therefore, in a vector polynomial chaos expansion (32), all $\mathbf{u}_k$ for $k>n$ must be linearly dependent. Based on the idea of linearly dependency, our main result in this direction is the following:

Theorem 2. *There exist a finite set of functions $\widetilde{\Gamma}_k : (\mathbb{R}^m \times \Omega) \to (\mathbb{R} \times \Omega)$ and an orthonormal basis $\boldsymbol{\phi}_k \in \mathbb{R}^n$ for $k = 1, 2, \ldots, n$ such that a vector polynomial chaos expansion can be expressed by*

$$\hat{\mathbf{u}}(\omega) = \sum_{k=1}^{n} \widetilde{\Gamma}_k(\boldsymbol{\xi}(\omega))\boldsymbol{\phi}_k \tag{33}$$

Proof. The orthonormal basis $\boldsymbol{\phi}_k \in \mathbb{R}^n$ has been identified in Eq. (9). Therefore, we only have to prove the existence of the functions $\widetilde{\Gamma}_k(\boldsymbol{\xi}(\omega))$ from the vector polynomial chaos expansion. Stacking u_j for all $j = 1, 2, \ldots, n$ in a vector, the scalar polynomial chaos expansion (6) can be expressed for a vector valued function as

$$\begin{aligned}\mathbf{u}(\omega) &= \mathbf{u}_{i_0} h_0 + \sum_{i_1=1}^{\infty} \mathbf{u}_{i_1} h_1(\xi_{i_1}(\omega)) \\ &+ \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1} \mathbf{u}_{i_1,i_2} h_2(\xi_{i_1}(\omega), \xi_{i_2}(\omega)) + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2} \mathbf{u}_{i_1 i_2 i_3} h_3(\xi_{i_1}(\omega), \xi_{i_2}(\omega), \xi_{i_3}(\omega)) \\ &+ \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2}\sum_{i_4=1}^{i_3} \mathbf{u}_{i_1 i_2 i_3 i_4}\, h_4(\xi_{i_1}(\omega), \xi_{i_2}(\omega), \xi_{i_3}(\omega), \xi_{i_4}(\omega)) + \ldots,\end{aligned} \tag{34}$$

where $\mathbf{u}_{i_1,\ldots,i_p} \in \mathbb{R}^n$ are deterministic vectors to be determined. Using the spanning property of the orthonormal basis $\boldsymbol{\phi}_k \in \mathbb{R}^n$ in Remark 1, each of the $\mathbf{u}_{i_1,\ldots,i_p}$ can be uniquely expressed as

$$\mathbf{u}_{i_1,\ldots,i_p} = \alpha^{(1)}_{i_1,\ldots,i_p}\boldsymbol{\phi}_1 + \alpha^{(2)}_{i_1,\ldots,i_p}\boldsymbol{\phi}_2 + \ldots + \alpha^{(n)}_{i_1,\ldots,i_p}\boldsymbol{\phi}_n \tag{35}$$

Substituting this in Eq. (34) and collecting all the coefficients associated with each orthonormal vector $\boldsymbol{\phi}_k$ the theorem is proved where

$$\begin{aligned}\widetilde{\Gamma}_k(\boldsymbol{\xi}(\omega)) &= \alpha^{(k)}_{i_0} h_0 + \sum_{i_1=1}^{\infty} \alpha^{(k)}_{i_1} h_1(\xi_{i_1}(\omega)) \\ &+ \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1} \alpha^{(k)}_{i_1,i_2} h_2(\xi_{i_1}(\omega), \xi_{i_2}(\omega)) + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2} \alpha^{(k)}_{i_1 i_2 i_3} h_3(\xi_{i_1}(\omega), \xi_{i_2}(\omega), \xi_{i_3}(\omega)) \\ &+ \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2}\sum_{i_4=1}^{i_3} \alpha^{(k)}_{i_1 i_2 i_3 i_4}\, h_4(\xi_{i_1}(\omega), \xi_{i_2}(\omega), \xi_{i_3}(\omega), \xi_{i_4}(\omega)) + \ldots,\end{aligned} \tag{36}$$

□

The fact that linearly depended vectors can be 'grouped' in this way may offer computational advantage as the number of truly independent vectors to be determined is only n. We have the following result in the regard:

Corollary 1. *The spectral functions in Eq. (8) and the polynomial chaos functions in Eq. (36) are equal, that is* $\Gamma_k(\boldsymbol{\xi}(\omega)) = \widetilde{\Gamma}_k(\boldsymbol{\xi}(\omega)), \forall k = 1, 2, \ldots, n$.

Proof. From Theorem 1 and Theorem 2 we have

$$\hat{\mathbf{u}}(\omega) = \sum_{k=1}^{n} \Gamma_k(\boldsymbol{\xi}(\omega))\boldsymbol{\phi}_k = \sum_{k=1}^{n} \widetilde{\Gamma}_k(\boldsymbol{\xi}(\omega))\boldsymbol{\phi}_k \tag{37}$$

Rearranging the last two terms in the preceding expression, one obtains

$$\sum_{k=1}^{n} \left[\Gamma_k(\boldsymbol{\xi}(\omega)) - \widetilde{\Gamma}_k(\boldsymbol{\xi}(\omega))\right] \boldsymbol{\phi}_k = \mathbf{0} \tag{38}$$

Since the set $\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_n\}$ is linearly independent, according to Definition Definition 1 their linear combination can be $\mathbf{0}$ if and only if the coefficients are 0 for all k, that is, iff

$$\left[\Gamma_k(\boldsymbol{\xi}(\omega)) - \widetilde{\Gamma}_k(\boldsymbol{\xi}(\omega))\right] = 0 \quad \text{or} \quad \Gamma_k(\boldsymbol{\xi}(\omega)) = \widetilde{\Gamma}_k(\boldsymbol{\xi}(\omega)); k = 1, 2, \ldots, n \tag{39}$$

□

Remark 4. Suppose the polynomial chaos expansion is truncated after P terms. From Eqs. (36) and (39) one observes that each of the spectral functions in effect groups together linear combination of P polynomial chaoses. Although both are expressed via infinite series, the difference is that the spectral functions $\Gamma_k(\boldsymbol{\xi}(\omega))$ are completely defined in closed-form by the spectral properties of the governing matrices, while each of the polynomial chaos function $\widetilde{\Gamma}_k(\boldsymbol{\xi}(\omega))$ in (36) contain P unknown constants $\alpha^{(k)}_{i_1,\ldots,i_p}$. Therefore, for all $k = 1, 2, \ldots, n$, there are in total nP number of unknown coefficients to be determined if polynomial chaos function $\widetilde{\Gamma}_k(\boldsymbol{\xi}(\omega))$ are to be used. The explicit relationship proved in Corollary Corollary 1 shows that this computational expense may be avoided if the proposed spectral functions $\Gamma_k(\boldsymbol{\xi}(\omega))$ are used as they are a-priori known. The spectral function approach can be viewed as a method of eliminating the need to compute large number of linearly dependent vectors appearing in a vector polynomial chaos expansion.

IV. Properties of the spectral functions

In the previous subsection we have proved that the proposed spectral functions are effectively linear combinations of infinite number of polynomial chaoses (Hermite polynomials). Since these functions are known in terms of the spectral properties of the system matrices, it eliminates the need for determining the huge number of unknown constants necessary in a vector polynomial chaos based approach. In this section we investigate some properties of the spectral functions.

IV.A. Order of the spectral functions

The main challenge from the computational point of view is that each of the spectral function is represented in terms of an infinite series. Clearly this series has to be truncated for practical purposes. In this section we discuss some important properties of these functions. From the series expansion in Eq. (25) we have

$$\begin{aligned}\boldsymbol{\Psi}\left(\boldsymbol{\xi}(\omega)\right) = \boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right) - \boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right)\boldsymbol{\Delta}\left(\boldsymbol{\xi}(\omega)\right)\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right)\\ + \boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right)\boldsymbol{\Delta}\left(\boldsymbol{\xi}(\omega)\right)\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right)\boldsymbol{\Delta}\left(\boldsymbol{\xi}(\omega)\right)\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right) + \ldots\end{aligned} \tag{40}$$

Since $\boldsymbol{\Lambda}\left(\boldsymbol{\xi}(\omega)\right)$ is a diagonal matrix, its inverse is simply a diagonal matrix containing the inverse of each of the diagonal elements. Also recall that the diagonal of $\boldsymbol{\Delta}\left(\boldsymbol{\xi}(\omega)\right)$ contains only zeros. For further analytical results, truncating the series upto different terms, we define spectral functions of different order.

Definition 4. The first-order spectral functions $\Gamma_k^{(1)}(\boldsymbol{\xi}(\omega)), k = 1, 2, \ldots, n$ are obtained by retaining one term in the series (40).

Retaining one term in (40) we have

$$\boldsymbol{\Psi}^{(1)}\left(\boldsymbol{\xi}(\omega)\right) = \boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right) \quad \text{or} \quad \Psi_{kj}^{(1)}\left(\boldsymbol{\xi}(\omega)\right) = \frac{\delta_{kj}}{\lambda_{0_k} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_k}} \tag{41}$$

Using the definition of the spectral function in Eq. (27), the first-order spectral functions can be explicitly obtained as

$$\Gamma_k^{(1)}\left(\boldsymbol{\xi}(\omega)\right) = \sum_{j=1}^{n} \Psi_{kj}^{(1)}\left(\boldsymbol{\xi}(\omega)\right)\left(\boldsymbol{\phi}_j^T \mathbf{f}\right) = \frac{\boldsymbol{\phi}_k^T \mathbf{f}}{\lambda_{0_k} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_k}} \tag{42}$$

From this expression it is clear that $\Gamma_k^{(1)}\left(\boldsymbol{\xi}(\omega)\right)$ are non-Gaussian random variables even if $\xi_i(\omega)$ are Gaussian random variables. Since we assumed that all eigenvalues λ_{0_k} are distinct, every $\Gamma_k^{(1)}\left(\boldsymbol{\xi}(\omega)\right)$ in Eq. (42) are different for different values of k.

Definition 5. The second-order spectral functions $\Gamma_k^{(2)}(\boldsymbol{\xi}(\omega)), k = 1, 2, \ldots, n$ are obtained by retaining two terms in the series (40).

Retaining two terms in (40) we have

$$\boldsymbol{\Psi}^{(2)}\left(\boldsymbol{\xi}(\omega)\right) = \boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right) - \boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right)\boldsymbol{\Delta}\left(\boldsymbol{\xi}(\omega)\right)\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right) \quad \text{or} \tag{43}$$

$$\Psi_{kj}^{(2)}\left(\boldsymbol{\xi}(\omega)\right) = \frac{\delta_{kj}}{\lambda_{0_k} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_k}} - \frac{\sum_{i=1}^{M} \xi_i(\omega)\Delta_{i_{kj}}}{\left(\lambda_{0_k} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_k}\right)\left(\lambda_{0_j} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_j}\right)} \tag{44}$$

Using the definition of the spectral function in Eq. (27), the second-order spectral functions can be obtained in closed-form as

$$\Gamma_k^{(2)}\left(\boldsymbol{\xi}(\omega)\right) = \frac{\boldsymbol{\phi}_k^T \mathbf{f}}{\lambda_{0_k} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_k}} - \sum_{j=1}^{n} \frac{\left(\boldsymbol{\phi}_j^T \mathbf{f}\right)\sum_{i=1}^{M} \xi_i(\omega)\Delta_{i_{kj}}}{\left(\lambda_{0_k} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_k}\right)\left(\lambda_{0_j} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_j}\right)} \tag{45}$$

The second-order function can be viewed as adding corrections to the first-order expression derived in Eq. (42). This is again a non-Gaussian random variable.

Definition 6. The vector of spectral functions of order s can be obtained by retaining s terms in the series (40) and can be expressed as

$$\boldsymbol{\Gamma}^{(s)}(\boldsymbol{\xi}(\omega)) = \left[\mathbf{I}_n - \mathbf{R}(\boldsymbol{\xi}(\omega)) + \mathbf{R}(\boldsymbol{\xi}(\omega))^2 - \mathbf{R}(\boldsymbol{\xi}(\omega))^3 \ldots s^{\text{th term}}\right]\boldsymbol{\Gamma}^{(1)}(\boldsymbol{\xi}(\omega)) \tag{46}$$

where $\mathbf{I}_n$ is the n-dimensional identity matrix and $\mathbf{R}$ is defined as $\mathbf{R}(\boldsymbol{\xi}(\omega)) = [\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\xi}(\omega)\right)][\boldsymbol{\Delta}\left(\boldsymbol{\xi}(\omega)\right)]$. Different terms of this series can be obtained recursively from the previous term.[13]

IV.B. Relationship with the algebraic solution

In order to obtain further insight into these functions, we look into the functional nature of the solution $\mathbf{u}(\omega)$ in terms of the random variables $\xi_i(\omega)$. We have the following result in this regard:

Theorem 3. *If all $\mathbf{A}_i \in \mathbb{R}^{n\times n}$ are matrices of rank n, then the elements of $\mathbf{u}(\omega)$ are the ratio of polynomials of the form*

$$\frac{p^{(n-1)}(\xi_1(\omega), \xi_2(\omega), \dots, \xi_M(\omega))}{p^{(n)}(\xi_1(\omega), \xi_2(\omega), \dots, \xi_M(\omega))} \tag{47}$$

where $p^{(n)}(\xi_1(\omega), \xi_2(\omega), \dots, \xi_M(\omega))$ is an n-th order complete multivariate polynomial of variables $\xi_1(\omega), \xi_2(\omega), \dots, \xi_M(\omega)$.

Proof. Suppose we denote

$$\mathbf{A}(\omega) = \left[\mathbf{A}_0 + \sum_{i=1}^{M} \xi_i(\omega)\mathbf{A}_i\right] \in \mathbb{R}^{n\times n} \tag{48}$$

so that

$$\mathbf{u}(\omega) = \mathbf{A}^{-1}(\omega)\mathbf{f} \tag{49}$$

From the definition of the matrix inverse (omitting ω for notational convenience) we have

$$\mathbf{A}^{-1} = \frac{\mathrm{Adj}(\mathbf{A})}{\det(\mathbf{A})} = \frac{\mathbf{C}_a^T}{\det(\mathbf{A})} \tag{50}$$

where $\mathbf{C}_a$ is the matrix of cofactors. The determinant of $\mathbf{A}$ contains a maximum of n number of products of A_{kj} and their linear combinations. Note from Eq. (48) that

$$A_{kj}(\omega) = A_{0_{kj}} + \sum_{i=1}^{M} \xi_i(\omega)\mathbf{A}_{i_{kj}} \tag{51}$$

Since all the matrices are of full rank, the determinant contains a maximum of n number of products of linear combination of random variables in Eq. (51). On the other hand, each entries of the matrix of cofactors, contains a maximum of $(n-1)$ number of products of linear combination of random variables in Eq. (51). From Eqs. (49) and (50) it follows that

$$\mathbf{u}(\omega) = \frac{\mathbf{C}_a^T\mathbf{f}}{\det(\mathbf{A})} \tag{52}$$

Therefore, the numerator of each element of the solution vector contains linear combinations of the elements of the cofactor matrix, which are complete polynomials of order $(n-1)$. Thus, the theorem is proved from the ratio in (52). □

The result derived in this theorem is important because the solution methods proposed for stochastic finite element analysis essentially aim to approximate the ratio of the polynomials given in Eq. (47). Our next result shows the nature of the spectral functions in this regard.

Theorem 4. *The linear combination of the spectral functions has the same functional form in* $(\xi_1(\omega), \xi_2(\omega), \ldots, \xi_M(\omega))$ *as the elements of the solution vector, that is,*

$$\hat{u}_r(\omega) \equiv \frac{p_r^{(n-1)}(\xi_1(\omega), \xi_2(\omega), \ldots, \xi_M(\omega))}{p_r^{(n)}(\xi_1(\omega), \xi_2(\omega), \ldots, \xi_M(\omega))}, \quad \forall r = 1, 2, \ldots, n \tag{53}$$

Proof. We give the proof for the first and second order spectral functions. The extension to the general case can be proved by induction.

When first-order spectral functions (42) are considered, we have

$$\hat{u}_r^{(1)}(\omega) = \sum_{k=1}^{n} \Gamma_k^{(1)}\left(\boldsymbol{\xi}(\omega)\right)\phi_{rk} = \sum_{k=1}^{n} \frac{\boldsymbol{\phi}_k^T \mathbf{f}}{\lambda_{0_k} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_k}} \phi_{rk} \tag{54}$$

All $(\lambda_{0_k} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_k})$ are different for different k because it is assumed that all eigenvalues λ_{0_k} are distinct. Carrying out the above summation one has n number of products of $(\lambda_{0_k} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_k})$ in the denominator and n sums of $(n-1)$ number of products of $(\lambda_{0_k} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_k})$ in the numerator, that is,

$$\hat{u}_r^{(1)}(\omega) = \frac{\sum_{k=1}^{n} (\boldsymbol{\phi}_k^T \mathbf{f}) \phi_{rk} \prod_{j=1 \neq k}^{n-1} \left(\lambda_{0_j} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_j}\right)}{\prod_{k=1}^{n-1} \left(\lambda_{0_j} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_j}\right)} \tag{55}$$

This proves the theorem.

For the second-order spectral functions (45)

$$\begin{aligned}
\hat{u}_r^{(2)}(\omega) &= \sum_{k=1}^{n} \Gamma_k^{(1)}\left(\boldsymbol{\xi}(\omega)\right)\phi_{rk} \\
&= \sum_{k=1}^{n} \left[\frac{\boldsymbol{\phi}_k^T \mathbf{f}}{\lambda_{0_k} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_k}} - \sum_{j=1}^{n} \frac{\left(\boldsymbol{\phi}_j^T \mathbf{f}\right) \sum_{i=1}^{M} \xi_i(\omega)\Delta_{i_{kj}}}{\left(\lambda_{0_k} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_k}\right)\left(\lambda_{0_j} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_j}\right)} \right] \phi_{rk} \\
&= \hat{u}_r^{(1)}(\omega) - \sum_{k=1}^{n}\sum_{j=1}^{n} \frac{\left(\boldsymbol{\phi}_j^T \mathbf{f}\right) \sum_{i=1}^{M} \xi_i(\omega)\Delta_{i_{kj}}}{\left(\lambda_{0_k} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_k}\right)\left(\lambda_{0_j} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_j}\right)} \phi_{rk}
\end{aligned} \tag{56}$$

Carrying out the above summation proves the theorem. □

This theorem proves that the nature of the solution has the same mathematical form of the exact solution, that is, the ratio of two polynomials in $(\xi_1(\omega), \xi_2(\omega), \ldots, \xi_M(\omega))$ where the numerator has a lower order compared to the denominator. This is in contrast with other methods such as the perturbation methods, classical Neumann series or polynomial chaos expansions, which are in effect power series in $(\xi_1(\omega), \xi_2(\omega), \ldots, \xi_M(\omega))$ (i.e., no polynomials in the denominator). Although the functional form of the proposed approach and the exact solutions are the same, Theorem 4 by no means imply that the polynomials are the same polynomials when truncated spectral functions are used. Next we propose a Galerkin approach to minimize the error arising due to the truncation of the spectral functions.

V. Model reduction using a reduced number of basis

So far all n number of spectral functions have been used. Based on the decay of the spectrum of operator $\mathbf{A}_0$, we proposed a proper orthogonal decomposition (POD)-like novel model reduction approach in this section. Suppose the eigenvalues of the deterministic system matrix $\mathbf{A}_0$ are arranged in an increasing order such that

$$\lambda_{0_1} < \lambda_{0_2} < \ldots < \lambda_{0_n} \tag{57}$$

From the expression of the spectral functions observe that the eigenvalues appear in the denominator:

$$\Gamma_k^{(1)}\left(\boldsymbol{\xi}(\omega)\right) = \frac{\boldsymbol{\phi}_k^T \mathbf{f}}{\lambda_{0_k} + \sum_{i=1}^{M} \xi_i(\omega)\lambda_{i_k}} \tag{58}$$

The numerator $(\boldsymbol{\phi}_k^T \mathbf{f})$ is the projection of the force on the deformation mode. For concentrated forcing, the significant mode can be identified. If $\mathbf{f}$ is distributed in nature, then the numerator does not change significantly with k. Therefore, for any given sample one has $|\Gamma_k^{(1)}\left(\boldsymbol{\xi}(\omega)\right)| \geq |\Gamma_{k+r}^{(1)}\left(\boldsymbol{\xi}(\omega)\right)|$ for every $k, r > 1$. The same can be established for higher order spectral functions using the general expression (46).

The series (70) can be truncated based on the magnitude of the eigenvalues as the higher terms becomes smaller. Therefore one could only retain the dominant terms in the series. A similar model reduction technique has been widely used within the proper orthogonal decomposition method where the eigenvalues of a symmetric positive definite matrix (the covariance matrix of a snapshot the system response) are used. One can select a small value ϵ such that $\lambda_{0_1}/\lambda_{0_p} < \epsilon$ for some value of p. Based on this discussion we have the following proposition.

Proposition 1. *(reduced orthonormal basis) Suppose there exist an ϵ and $p < n$ such that $\lambda_{0_1}/\lambda_{0_p} < \epsilon$. Then the solution of the discretized stochastic finite element equation (4) can be expressed by the series representation*

$$\hat{\mathbf{u}}(\omega) = \sum_{k=1}^{p} c_k \widehat{\Gamma}_k(\boldsymbol{\xi}(\omega))\boldsymbol{\phi}_k \tag{59}$$

such that the error is minimized in a least-square sense. c_k, $\widehat{\Gamma}_k(\boldsymbol{\xi}(\omega))$ and $\boldsymbol{\phi}_k$ can be obtained following the procedure described in the previous section by letting the indices j, k upto p in Eqs. (72) and (73).

The mean-square convergence of the series (59) can be improved in two ways, namely, (a) by increasing the number of terms p, or (b) by increasing the order of the spectral functions $\widehat{\Gamma}_k(\boldsymbol{\xi}(\omega))$.

VI. Error minimization in the Hilbert space: The Galerkin approach

In Theorem 1 we proved the existence of the spectral functions such that a projection in an orthonormal basis converges to the exact solution in probability 1. The spectral functions are expressed in terms of a convergent infinite series. Approximate first and second order spectral functions by truncating the infinite series have been derived. We have also proved that they have the same functional form as the exact solution of Eq. (4). Additionally, in the previous section the idea of reduced basis was proposed. This motivates us to use the reduced number of spectral functions functions as ‘trial functions’ to construct an approximate solution. The stochastic subspace projection scheme has been utilized to compute the undetermined coefficients of the reduced basis approximation. In particular, the focus is on the Bubnov-Galerkin scheme which involves the formulation of a stochastic residual of the form

$$\boldsymbol{\varepsilon}(\omega) = \mathbf{A}(\omega)\boldsymbol{\Psi}(\omega)\mathbf{C}(\omega) - \mathbf{f} \in \mathbb{R}^n \tag{60}$$

where $\boldsymbol{\varepsilon}(\omega)$ is the residual error, $\mathbf{A}(\omega)$ is a general stochastic system matrix and $\boldsymbol{\Psi}(\omega)$ the basis functions in the reduced subspace. $\mathbf{C}(\omega) \in \mathbb{R}^p$ is the vector of the random undetermined coefficients which have to be evaluated to project the solution in the reduced subspace of dimension p. The Bubnov-Galerkin scheme applied to deterministic problems is equivalent to enforcing the condition of orthogonality between the residue and the basis functions used to represent the solution in the reduced subspace, which can be expressed as $\boldsymbol{\varepsilon}(\omega) \perp \boldsymbol{\Psi}(\omega)$. Using this orthogonal projection scheme, the undetermined coefficients can be expressed as

$$\tilde{\mathbf{A}}(\omega)\mathbf{C}(\omega) = \tilde{\mathbf{f}} \in \mathbb{R}^p \tag{61}$$

where $\tilde{\mathbf{A}}(\omega) = \boldsymbol{\Psi}^*(\omega)\mathbf{A}(\omega)\boldsymbol{\Psi}(\omega) \in \mathbb{R}^{p\times p}$ and $\tilde{\mathbf{f}}(\omega) = \boldsymbol{\Psi}^*(\omega)\mathbf{f} \in \mathbb{R}^p$ reduced random coefficient matrix and the rhs, respectively.

Eq. (61) can be solved for any realization of ω and following this it can be said that (see reference[30]) $P[\boldsymbol{\Psi}^*(\omega)\boldsymbol{\varepsilon}(\omega) = 0] = 1$. In the computational mechanics literature it is found that a solution to Eq. (61) can also be obtained by treating the coefficients as deterministic scalars and solving the following system of equations[5]

$$\left\langle \tilde{\mathbf{A}}(\omega) \right\rangle \mathbf{C} = \left\langle \tilde{\mathbf{f}}(\omega) \right\rangle \in \mathbb{R}^p \tag{62}$$

which is derived using the measure of the norm defined in the Hilbert space of random variables. Eq. (61) is expected to produce more accurate results compared to Eq. (62), however, the former is computationally more expensive. It has been shown[30] that the Bubnov-Galerkin scheme in Eq. (62) is a zero-order approximation of a Neumann expansion of the exact scheme given in Eq. (61). In the present study, we give a comparison of the above two approaches in Subsection VIII.C to study their accuracy.

It may be also interesting to look into the relationship between the approach of minimization of the squared norm of the residual and the Bubnov-Galerkin scheme of orthogonalization of the basis functions to the residual vector. The squared residual vector can be written as

$$\boldsymbol{\varepsilon}(\omega)^T\boldsymbol{\varepsilon}(\omega) = \left[\mathbf{A}(\omega)\hat{\mathbf{u}}(\omega) - \mathbf{f}\right]^T \left[\mathbf{A}(\omega)\hat{\mathbf{u}}(\omega) - \mathbf{f}\right] \tag{63}$$

where $\hat{\mathbf{u}}(\omega)$ is the solution vector expressed in terms of a reduced basis. Here the right hand side can be expressed as $\mathbf{f} = \mathbf{A}(\omega)\mathbf{u}(\omega)$ where $\mathbf{u}(\omega)$ is the actual solution of the system. Hence Eq. (63) becomes

$$\|\boldsymbol{\varepsilon}(\omega)\| = \left[\hat{\mathbf{u}}^T(\omega) - \mathbf{u}^T(\omega)\right] \mathbf{A}^T(\omega)\mathbf{A}(\omega) \left[\hat{\mathbf{u}}(\omega) - \mathbf{u}(\omega)\right] \tag{64}$$

where $\|\boldsymbol{\varepsilon}(\omega)\| = \boldsymbol{\varepsilon}(\omega)^T\boldsymbol{\varepsilon}(\omega)$. The system response can be expressed as $\hat{\mathbf{u}}(\omega) = \sum_{k=1}^{p} c_k(\omega)\boldsymbol{\psi}_k(\omega)$, where $\boldsymbol{\psi}_k(\omega)$ are the stochastic basis vectors used to represent the response in the reduced subspace. Now, minimizing the norm of residual with respect to the coefficient $c_k(\omega)$ we have

$$\frac{\partial \|\boldsymbol{\varepsilon}(\omega)\|}{\partial c_k(\omega)} = 0 \tag{65}$$

which gives

$$\boldsymbol{\psi}_k^T(\omega)\left[\mathbf{A}^T(\omega)\mathbf{A}(\omega)[\hat{\mathbf{u}}(\omega)-\mathbf{u}(\omega)]\right]=0 \quad \text{or}$$
$$\boldsymbol{\psi}_k^T(\omega)\left[\mathbf{A}^T(\omega)\mathbf{A}(\omega)\boldsymbol{\varepsilon}(\omega)\right]=0. \tag{66}$$

The symmetry property of the system matrix $\mathbf{A}(\omega)$ has been utilized in the above derivation. Thus we see that minimizing the norm of the residual renders the basis vectors $\mathbf{A}^2$-orthogonal to the residual vector.

However, if we choose the objective function as $[\boldsymbol{\varepsilon}^T(\omega)\mathbf{A}^{-T}(\omega)\boldsymbol{\varepsilon}(\omega)]$ and minimize it with respect to the coefficient $c_k(\omega)$, we have

$$\frac{\partial(\boldsymbol{\varepsilon}^T(\omega)\mathbf{A}^{-T}(\omega)\boldsymbol{\varepsilon}(\omega))}{\partial c_k(\omega)}=0 \tag{67}$$

which gives

$$\frac{\partial}{\partial c_k(\omega)}\left[(\mathbf{A}(\omega)\hat{\mathbf{u}}(\omega)-\mathbf{f})^T\mathbf{A}^{-T}(\omega)\left(\mathbf{A}(\omega)\hat{\mathbf{u}}(\omega)-\mathbf{f}\right)\right]=0 \quad \text{or}$$
$$\frac{\partial}{\partial c_k(\omega)}\left[\left(\sum_{k=1}^{p}c_k(\omega)\boldsymbol{\psi}_k(\omega)-\mathbf{u}(\omega)\right)^T\mathbf{A}^T(\omega)\mathbf{A}^{-T}(\omega)\boldsymbol{\varepsilon}(\omega)\right]=0 \tag{68}$$

by utilizing the symmetry of the system matrix $\mathbf{A}(\omega)$. Hence from the above equation we obtain the orthogonality relationship of

$$\boldsymbol{\psi}_k^T(\omega)\boldsymbol{\varepsilon}(\omega)=0 \quad \forall\, k=1,2,\ldots,p \tag{69}$$

Thus we see that the scheme of orthogonalization of the basis vectors to the residual is obtained from an **A**-norm residual optimization approach in contrast to the direct minimization of the norm of the residual vector. The study if the effect of this on the coefficients of the basis vectors would be interesting and comparison of the results obtained via these schemes remains to be explored further.

Here we adopt the idea of Galerkin approach in the Hilbert space where the error is made orthogonal to the spectral functions. We recall that the Hilbert space $L^2\left(\mathbb{R}^n,\mathcal{H}\right)$ is empowered with the inner product norm $\langle\mathbf{u}(\omega),\mathbf{v}(\omega)\rangle=\int_\Omega P(\mathrm{d}\omega)\mathbf{u}^T(\omega)\mathbf{v}(\omega)$. This norm is used in our next result.

Theorem 5. *There exist a set of finite functions $\widehat{\Gamma}_k:(\mathbb{R}^m\times\Omega)\to(\mathbb{R}\times\Omega)$, constants $c_k\in\mathbb{R}$ and orthonormal vectors $\boldsymbol{\phi}_k\in\mathbb{R}^n$ for $k=1,2,\ldots,p$ such that the series*

$$\hat{\mathbf{u}}(\omega)=\sum_{k=1}^{p}c_k\widehat{\Gamma}_k(\boldsymbol{\xi}(\omega))\boldsymbol{\phi}_k \tag{70}$$

converges to the exact solution of the discretized stochastic finite element equation (4) in the mean-square sense provided the vector $\mathbf{c}=\{c_1,c_2,\ldots,c_p\}^T$ satisfies the following $p\times p$ algebraic equations:

$$\mathbf{S}\,\mathbf{c}=\mathbf{b} \tag{71}$$

with

$$S_{jk} = \sum_{i=0}^{M} \widetilde{A}_{i_{jk}} D_{ijk}; \quad \forall\, j,k = 1,2,\ldots,p \tag{72}$$

where

$$\widetilde{A}_{i_{jk}} = \boldsymbol{\phi}_j^T \mathbf{A}_i \boldsymbol{\phi}_k, D_{ijk} = \mathrm{E}\left[\xi_i(\omega)\widehat{\Gamma}_j(\boldsymbol{\xi}(\omega))\widehat{\Gamma}_k(\boldsymbol{\xi}(\omega))\right], b_j = \mathrm{E}\left[\widehat{\Gamma}_j(\boldsymbol{\xi}(\omega))\right]\left(\boldsymbol{\phi}_j^T \mathbf{f}\right). \tag{73}$$

Proof. The functions $\widehat{\Gamma}_k(\boldsymbol{\xi}(\omega))$ can be the first-order (42), second-order (45) or any higher-order spectral functions and $\boldsymbol{\phi}_k$ are the eigenvectors introduced earlier in Eq. (9). Substituting the approximate expression of $\hat{\mathbf{u}}(\omega)$ in the governing equation (4), the error vector can be obtained as

$$\boldsymbol{\varepsilon}(\omega) = \left(\sum_{i=0}^{M} \mathbf{A}_i \xi_i(\omega)\right)\left(\sum_{k=1}^{p} c_k \widehat{\Gamma}_k(\boldsymbol{\xi}(\omega))\boldsymbol{\phi}_k\right) - \mathbf{f} \in \mathbb{R}^n \tag{74}$$

where $\xi_0 = 1$ is used to simplify the first summation expression. The expression (70) is viewed as a projection where $\left\{\widehat{\Gamma}_k(\boldsymbol{\xi}(\omega))\boldsymbol{\phi}_k\right\} \in \mathbb{R}^n$ are the basis functions and c_k are the unknown constants to be determined. Using the Galerkin approach we make the error orthogonal to the basis functions, that is, mathematically

$$\boldsymbol{\varepsilon}(\omega) \perp \left(\widehat{\Gamma}_j(\boldsymbol{\xi}(\omega))\boldsymbol{\phi}_j\right) \quad \text{or} \quad \left\langle \widehat{\Gamma}_j(\boldsymbol{\xi}(\omega))\boldsymbol{\phi}_j, \boldsymbol{\varepsilon}(\omega)\right\rangle = 0 \quad \forall\, j = 1,2,\ldots,p \tag{75}$$

Imposing this condition and using the expression of $\boldsymbol{\varepsilon}(\omega)$ from Eq. (74) one has

$$\mathrm{E}\left[\widehat{\Gamma}_j(\boldsymbol{\xi}(\omega))\boldsymbol{\phi}_j^T\left(\sum_{i=0}^{M} \mathbf{A}_i \xi_i(\omega)\right)\left(\sum_{k=1}^{p} c_k \widehat{\Gamma}_k(\boldsymbol{\xi}(\omega))\boldsymbol{\phi}_k\right) - \widehat{\Gamma}_j(\boldsymbol{\xi}(\omega))\boldsymbol{\phi}_j^T \mathbf{f}\right] = 0 \;\forall\, j \tag{76}$$

Interchanging the $\mathrm{E}[\bullet]$ and summation operations, this can be simplified to

$$\sum_{k=1}^{p}\left(\sum_{i=0}^{M}\left(\boldsymbol{\phi}_j^T \mathbf{A}_i \boldsymbol{\phi}_k\right)\mathrm{E}\left[\xi_i(\omega)\widehat{\Gamma}_j(\boldsymbol{\xi}(\omega))\widehat{\Gamma}_k(\boldsymbol{\xi}(\omega))\right]\right) c_k = \mathrm{E}\left[\widehat{\Gamma}_j(\boldsymbol{\xi}(\omega))\right]\left(\boldsymbol{\phi}_j^T \mathbf{f}\right) \tag{77}$$

$$\text{or} \quad \sum_{k=1}^{p}\left(\sum_{i=0}^{M} \widetilde{A}_{i_{jk}} D_{ijk}\right) c_k = b_j \tag{78}$$

where $\widetilde{A}_{i_{jk}}$, D_{ijk} and b_j are as defined in Eq. (73). This completes the proof. □

Remark 5. (Comparison with the classical spectral SFEM) We compare this Galerkin approach with the classical spectral stochastic finite element approach for further insight. The number of equations to be solved for the unknown coefficients in Eq. (71) is p. The computational reduction from Theorem 5 is arising from the fact that each of the spectral functions groups P number of polynomial chaos functions exploiting the linear dependence of the associated vectors as proved in Theorem 2 (see Eq. (36)). As a result, there are only p unknown constants, as opposed to pP unknown constants arising in the polynomial chaos expansion. However, the spectral functions $\widehat{\Gamma}_k(\boldsymbol{\xi}(\omega))$ are highly non-Gaussian in nature and do not in general enjoy any orthogonality properties like the Hermite polynomials or any other orthogonal polynomials[22,31,23] with respect to the underlying probability measure. The coefficient matrix **S** and the vector **b** in Eq. (71) should be obtained numerically using the Monte Carlo simulation or other numerical integration technique. In the classical PC expansion, however, the coefficient matrix and the associated vector are obtained exactly in closed-form. In addition, the coefficient matrix is a sparse matrix where the matrix **S** in Eq. (71) is a fully populated matrix.

It can be observed that the matrix **S** in Eq. (71) is symmetric. Therefore, one need to determine $p(p+1)/2$ number of coefficients by numerical methods. Any numerical integration method, such as the Gaussian quadrature method, can be used to obtain the elements of D_{ijk} and b_j in Eq. (73). In this paper Monte Carlo simulation is used. The samples of the spectral functions $\widehat{\Gamma}_k(\boldsymbol{\xi}(\omega))$ can be simulated from Eq. (42) or Eq. (45). These can be used to compute D_{ijk} and b_j from Eq. (73). The simulated spectral functions can also be 'recycled' to obtain the statistics and probability density function (pdf) of the solution. In summary, compared to the classical spectral stochastic finite element method, the proposed Galerkin approach results in a smaller size matrix but requires numerical integration techniques to obtain its entries. The numerical method proposed here therefore can be considered as a hybrid analytical-simulation approach.

VII. Post processing and computational approach

VII.A. Moments of the solution

For the practical application of the method developed here, the efficient computation of the response moments and pdf is of crucial importance. A simulation based algorithm is proposed in this section. The coefficients c_k in Eq. (59) can be calculated from a reduced set of equations given by (71). The reduced equations can be obtained by letting the indices j, k upto $p < n$ in Eqs. (72) and (73). After obtaining the coefficient vector $\mathbf{c} \in \mathbb{R}^p$, the statistical moments of the solution can be obtained from Eq. (59) using the Monte Carlo simulation. The spectral functions used to obtain the vector **c** itself, can be reused to obtain the statistics and pdf of the solution. The mean vector can be obtained as

$$\bar{\mathbf{u}} = \mathrm{E}\left[\hat{\mathbf{u}}(\omega)\right] = \sum_{k=1}^{p} c_k \mathrm{E}\left[\widehat{\Gamma}_k(\boldsymbol{\xi}(\omega))\right] \boldsymbol{\phi}_k \tag{79}$$

The covariance of the solution vector can be expressed as

$$\boldsymbol{\Sigma}_u = \mathrm{E}\left[\left(\hat{\mathbf{u}}(\omega) - \bar{\mathbf{u}}\right)\left(\hat{\mathbf{u}}(\omega) - \bar{\mathbf{u}}\right)^T\right] = \sum_{k=1}^{p}\sum_{j=1}^{p} c_k c_j \Sigma_{\Gamma_{kj}} \boldsymbol{\phi}_k \boldsymbol{\phi}_j^T \tag{80}$$

where the elements of the covariance matrix of the spectral functions are given by

$$\Sigma_{\Gamma_{kj}} = \mathrm{E}\left[\left(\widehat{\Gamma}_k(\boldsymbol{\xi}(\omega)) - \mathrm{E}\left[\widehat{\Gamma}_k(\boldsymbol{\xi}(\omega))\right]\right)\left(\widehat{\Gamma}_j(\boldsymbol{\xi}(\omega)) - \mathrm{E}\left[\widehat{\Gamma}_j(\boldsymbol{\xi}(\omega))\right]\right)\right] \tag{81}$$

VII.B. Summary of the computational approach

Based on the results derived in the paper, a hybrid reduced simulation-analytical approach is proposed. The method is applicable to elliptic problems with general non-Gaussian random fields. The computational procedure for the solution of the stochastic elliptic PDE (1) can be implemented as follows:

1. Obtain the system matrices $\mathbf{A}_i, i = 0, 1, 2, \ldots, M$ and the forcing vector $\mathbf{f}$ by discretizing the governing stochastic partial differential equation using the well established stochastic finite element methodologies.

2. Solve the eigenvalue problem associated with the mean matrix $\mathbf{A}_0$
$$\mathbf{A}_0\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Lambda}_0 \tag{82}$$

3. Select a small value of ϵ, say $\epsilon = 10^{-3}$. Obtain the number of the reduced orthonormal basis p such that $\lambda_{0_1}/\lambda_{0_p} < \epsilon$.

4. Create the reduced matrix of eigenvalues and eigenvectors
$$\mathbf{\Lambda}_{0_p} = \text{diag}\left[\lambda_{0_1}, \lambda_{0_2}, \ldots, \lambda_{0_p}\right] \in \mathbb{R}^{p\times p} \quad \text{and} \quad \mathbf{\Phi}_p = \left[\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_p\right] \in \mathbb{R}^{n\times p} \tag{83}$$

5. Calculate the transformed matrices and vector
$$\widetilde{\mathbf{A}}_i = \mathbf{\Phi}_p^T\mathbf{A}_i\mathbf{\Phi}_p \in \mathbb{R}^{p\times p}; i = 1, 2, \ldots, M \quad \text{and} \quad \widetilde{\mathbf{f}} = \mathbf{\Phi}_p^T\mathbf{f} \tag{84}$$
and separate the diagonal and off diagonal terms as $\widetilde{\mathbf{A}}_i = \mathbf{\Lambda}_i + \mathbf{\Delta}_i$.

6. Select a number of samples, say N_{samp}. Generate the samples of (in general non-Gaussian) random variables $\xi_i(\omega), i = 1, 2, \ldots, M$.

7. Obtain the inverse of the diagonal matrix $\mathbf{\Lambda}\left(\boldsymbol{\xi}(\omega)\right)$ defined in Eq. (20) as
$$\mathbf{\Lambda}_I(\omega) = \left[\mathbf{\Lambda}_0^{-1} + \sum_{i=1}^{M}\xi_i(\omega)\mathbf{\Lambda}_i^{-1}\right] \tag{85}$$
and the trace less matrix
$$\mathbf{\Delta}(\omega) = \sum_{i=1}^{M}\xi_i(\omega)\mathbf{\Delta}_i \tag{86}$$

8. Calculate the first-order spectral function in a vector form as
$$\widehat{\mathbf{\Gamma}}^{(1)}(\omega) = \mathbf{\Lambda}_I\left(\boldsymbol{\xi}(\omega)\right)\widetilde{\mathbf{f}} \in \mathbb{R}^p \tag{87}$$
If higher order spectral function are necessary, then calculate the matrix
$$\mathbf{R}(\boldsymbol{\xi}(\omega)) = \mathbf{\Lambda}_I\left(\boldsymbol{\xi}(\omega)\right)\mathbf{\Delta}\left(\boldsymbol{\xi}(\omega)\right) \in \mathbb{R}^{p\times p} \tag{88}$$
From this calculate s-th order spectral function as
$$\mathbf{\Gamma}^{(s)}(\boldsymbol{\xi}(\omega)) = \left[\mathbf{I}_p - \mathbf{R}(\boldsymbol{\xi}(\omega)) + \mathbf{R}(\boldsymbol{\xi}(\omega))^2 - \mathbf{R}(\boldsymbol{\xi}(\omega))^3 \ldots s^{\text{th term}}\right]\mathbf{\Gamma}^{(1)}(\boldsymbol{\xi}(\omega)) \in \mathbb{R}^p \tag{89}$$

9. Calculate the mean vector from the generated samples
$$\overline{\mathbf{\Gamma}} = \mathrm{E}\left[\mathbf{\Gamma}(\omega)\right] \in \mathbb{R}^p \tag{90}$$

10. Calculate the following $(1+M)$ matrices from the samples of $\mathbf{\Gamma}(\omega)$

$$\mathbf{D}_0 = \mathrm{E}\left[\mathbf{\Gamma}(\omega)\mathbf{\Gamma}^T(\omega)\right] \in \mathbb{R}^{p\times p} \tag{91}$$

$$\text{and} \quad \mathbf{D}_i = \mathrm{E}\left[\mathbf{\Gamma}(\omega)\xi_i(\omega)\mathbf{\Gamma}^T(\omega)\right] \in \mathbb{R}^{p\times p}, \forall\, i = 1,2,\ldots,M \tag{92}$$

11. Following Eq. (70), form the coefficient matrix **S** and the vector **b** as

$$\mathbf{S} = \mathbf{\Lambda}_{0_p} \odot \mathbf{D}_0 + \sum_{i=1}^{M} \widetilde{\mathbf{A}}_i \odot \mathbf{D}_i \in \mathbb{R}^{p\times p} \quad \text{and} \quad \mathbf{b} = \widetilde{\mathbf{f}} \odot \overline{\mathbf{\Gamma}} \in \mathbb{R}^p \tag{93}$$

where $\odot$ implies element to element multiplication (as in MATLAB$^{\mathrm{TM}}$ dot notation).

12. Obtain the coefficient vector

$$\mathbf{c} = \mathbf{S}^{-1}\mathbf{b} \in \mathbb{R}^p \tag{94}$$

13. Calculate the mean of the solution

$$\bar{\mathbf{u}} = \sum_{k=1}^{p} c_k \overline{\Gamma}_k \boldsymbol{\phi}_k \tag{95}$$

14. Obtain the covariance matrix of the spectral functions as

$$\mathbf{\Sigma}_\Gamma = \mathbf{D}_0 - \overline{\mathbf{\Gamma}}\overline{\mathbf{\Gamma}}^T \tag{96}$$

From this calculate the covariance of the solution using Eq. (80).

Remark 6. (The computational complexity) The main computational cost of the proposed method depends on (a) the solution of the matrix eigenvalue problem (82) with reduced number of eigenvalues, (b) the generation of the $\mathbf{D}_i$ matrices in Eqs. (91) and (92), and (c) the calculation of the coefficient vector in Eq. (94). Both the matrix inversion and the matrix eigenvalue problem scales in $O(p^3)$ in the worse case. The calculation of the $\mathbf{D}_i$ matrices in Eqs. (91) and (92) scales linearly with M and $p(p+1)/2$ with p. Therefore, this cost scales with $O((M+1)\,p(p+1)/2)$. The overall cost is $2O(p^3) + O((M+1)\,p(p+1)/2)$. For large M and p, asymptotically the computational cost becomes $C_s = O(Mp^2) + O(p^3)$. The important point to note here that the proposed approach scales linearly with the number of random variables M. For comparison, in the classical PC expansion one needs to solve a matrix equation of dimension Pn, which in the worse case scales with $(O(Pn)^3)$. Since $P \gg M$ and $n > p$, we have $(O(P^3n^3) \gg O(Mp^2) + O(p^3)$.

VIII. Numerical example

In this section we apply the computational method to a beam with stochastic bending modulus. We assume that the bending modulus is a homogeneous stationary Gaussian random field of the form

$$EI(x,\omega) = EI_0(1 + a(x,\omega)) \tag{97}$$

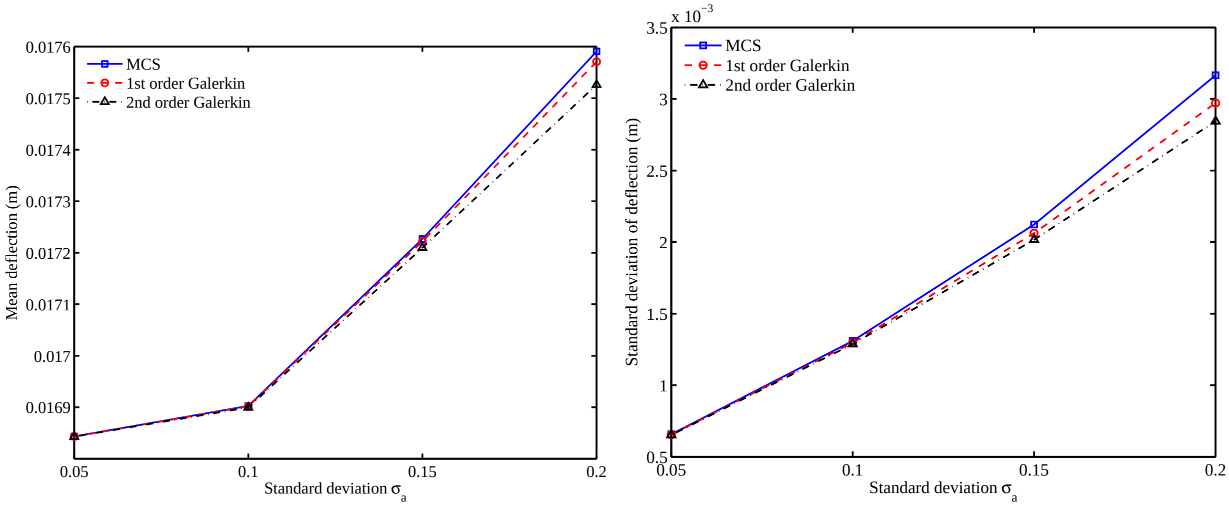
where x is the coordinate along the length of the beam, EI_0 is the estimate of the mean bending modulus, $a(x,\omega)$ is a zero mean stationary Gaussian random field. The autocorrelation function of this random field is assumed to be

$$C_a(x_1,x_2) = \sigma_a^2 e^{-(|x_1-x_2|)/\mu_a} \tag{98}$$

where μ_a is the correlation length and σ_a is the standard deviation. We use the baseline parameters as the length $L = 1$m, cross-section 40.06×2.05 mm^2 and Young's modulus $E = 69 \times 10^9$ Pa (that of Aluminum). In study we consider deflection of the tip of the beam due to $P = 0.1$N load. Two correlation lengths are considered in the numerical studies, namely $\mu_a = L/3$ and $\mu_a = L/10$. The number of terms retained (M) in the Karhunen-Loève expansion (3) is selected such that $\nu_M/\nu_1 = 0.05$ in order to retain 95% of the variability. For the two correlation lengths considered, the number of terms M becomes 24 and 67. For the finite element discretization, the beam is divided into 50 elements. Standard four degrees of freedom Euler-Bernoulli beam model is used.[32] After applying the fixed boundary condition at the edge, we obtain the number of degrees of freedom of the model $n = 100$.

VIII.A. Results for larger correlation length

The proposed method is applied with 10,000-sample Monte Carlo Simulation (MCS). The hybrid analytical-simulation method proposed here is compared with the direct MCS obtained by solving Eq. (4) for each sample. In figure 1 the mean and standard deviation of the deflection of the beam are shown for four values of σ_a. We consider $\sigma_a = \{0.05, 0.10, 0.15, 0.20\}$ to simulate increasing uncertainty. This is done to check the accuracy of the proposed method against the direct MCS results. It can be seen that the results from both the first and second-order spectral functions in conjunction with the Galerkin approach produce accurate results for all the four values of σ_a.

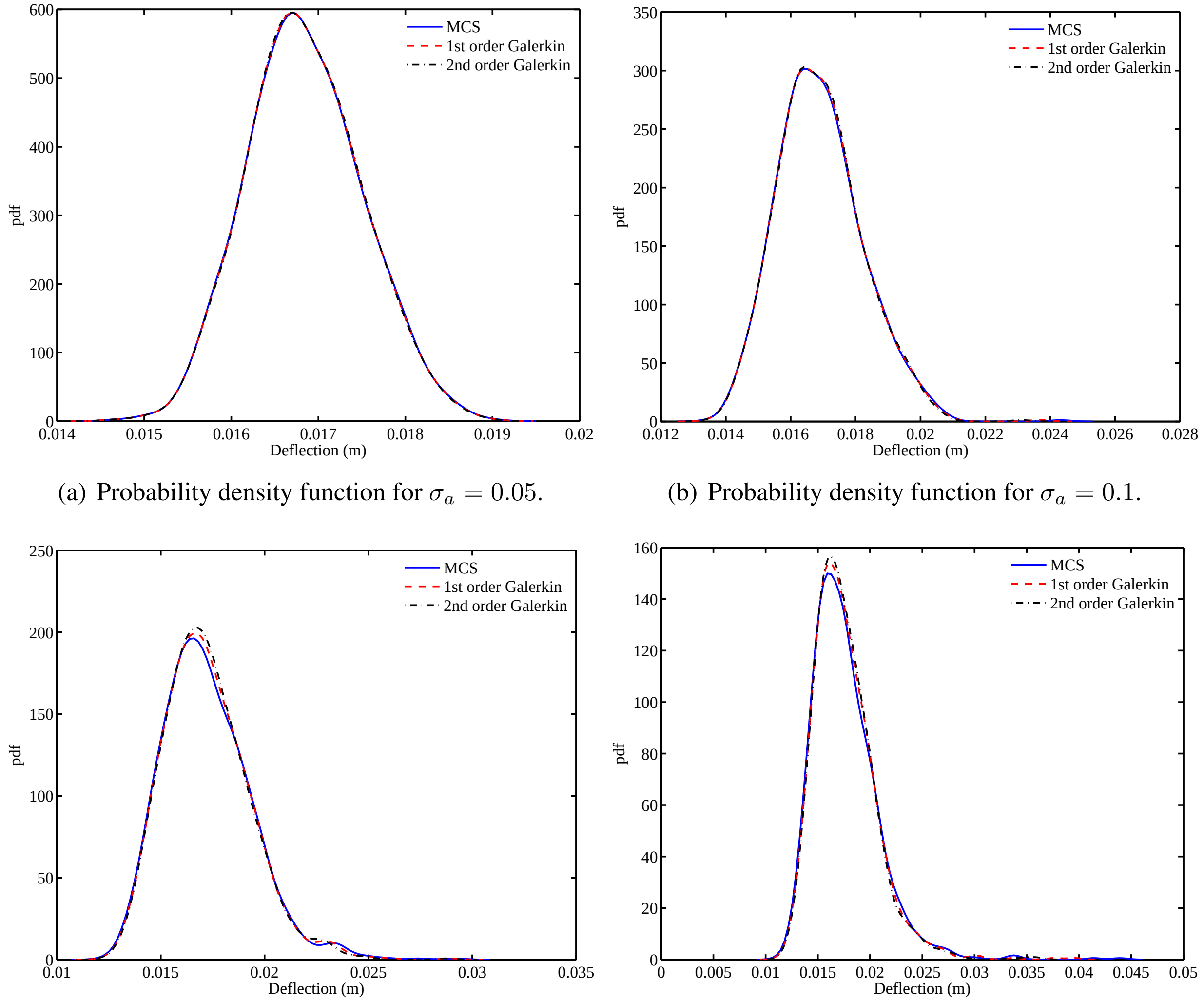


(a) Mean of the deflection.

(b) Standard deviation of the deflection.

Figure 1. The mean and standard deviation of the deflection of the beam. The correlation length of the random field describing the bending rigidity is assumed to be $\mu_a = L/3$. The number of random variable used: $M = 24$.

The probability density function of the deflection is shown figure 2 for the four values of σ_a. As expected, the error corresponding to the second-order spectral function approach is smaller than the first-order approach. In this problem the size of the system $n = 100$ and the number of random variables $M = 24$. If the second-order PC was used, then from Eq. (7) one obtains $P = 324$. This implies that one would need to solve a linear system of equation of size $nP = 100 \times 324 = 32400$. For the proposed Galerkin approach in Theorem 5, only a set of 100 equations are solved to obtain the coefficients. This shows the efficiency of this approach without loosing significant accuracy.



(a) Probability density function for $\sigma_a = 0.05$.

(b) Probability density function for $\sigma_a = 0.1$.

(c) Probability density function for $\sigma_a = 0.15$.

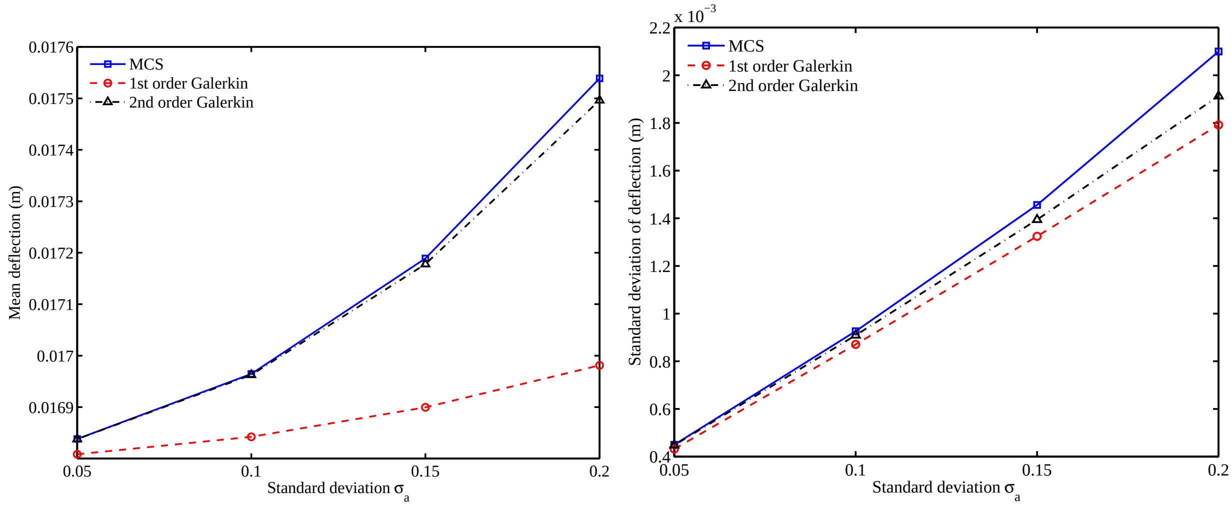
(d) Probability density function for $\sigma_a = 0.2$.

Figure 2. The probability density function of the deflection of the beam. The correlation length of the random field describing the bending rigidity is assumed to be $\mu_a = L/3$. The pdfs are obtained with 10,000 sample MCS and four values of σ_a have been used.

VIII.B. Results for smaller correlation length

When the correlation length becomes smaller, as expected for nano systems, the number of term to be retained in the Karhunen-Loève expansion (3) becomes large. In this case use have used $M = 67$ number of random variables. The method developed in the paper is applied with 10,000-sample Monte Carlo Simulation (MCS). In figure 3 the mean and standard deviation of the deflection of the beam are shown for four values of σ_a. It can be seen that the results from both the first and second-order spectral functions in conjunction with the Galerkin approach produce accurate results for all the four values of σ_a even when the number of random variables are large.

The probability density function of the deflection is shown figure 4 for the four values of σ_a. As expected, the error corresponding to the second-order spectral function approach is smaller than the first-order approach. For this problem the size of the system $n = 100$ and the number of random variables $M = 67$. If the second-order PC was used, then from Eq. (7) one obtains $P = 2345$. This implies that one would need to solve a linear system of equation of size $nP = 100 \times 2345 = 234,500$. For the proposed Galerkin approach in Theorem 5, only a set of 100



(a) Mean of the deflection.

(b) Standard deviation of the deflection.

Figure 3. The mean and standard deviation of the deflection of the beam. The correlation length of the random field describing the bending rigidity is assumed to be $\mu_a = L/10$. The number of random variable used: $M = 67$.

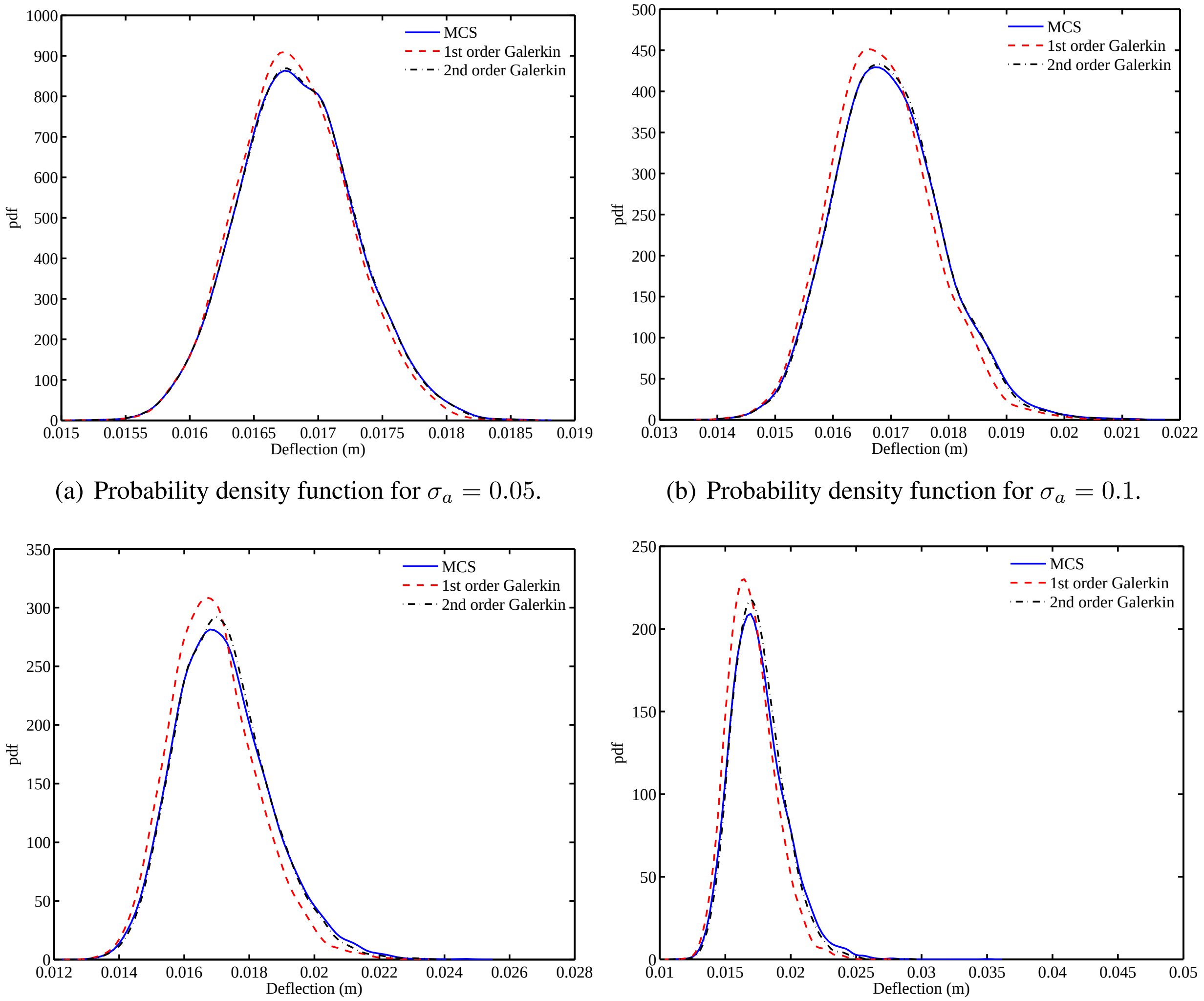
equations are solved to obtain the coefficients. Overall, when such a large number of random variables used, the accuracy is slightly less compared to the previous case where relatively smaller number of random variables are used. The results obtained here show the computational efficiency of the proposed approach even for such large number of random variables without loosing the accuracy significantly.

VIII.C. Comparison of the reduced basis projection schemes

This subsection compares the results obtained with the exact approach of orthogonalizing the residual to the basis vectors (described in Eq. (61)) with that of the Bubnov-Galerkin method given in Eq. (62). Table 1 shows the value of the mean deflection of the tip of the beam for the three approaches and it shows that the response. The comparison has been made for three different cases, namely (a) the Galerkin scheme, (b) the exact evaluation of the coefficients for each sample in the random space, and (c) without applying any error minimization scheme. For the latter case the displacement is calculated as $\hat{\mathbf{u}}(\omega) = \sum_j \widehat{\Gamma}_j \phi_j$. It is seen that the convergence of the mean deflection is quite accurate and no difference is observed for the cases of Galerkin and the exact case. However, it is seen that the values of mean deflection calculated without the application of any adjusting coefficients converges to those obtained by the Galerkin and the exact method.

Table 1. Mean deflection of the tip of the beam for correlation length of $L/10$ and $\sigma_a = 0.05$.

Order	Galerkin	Exact case	No minimization
1	0.0794	0.0792	0.1862
2	0.0792	0.0792	0.0796
3	0.0792	0.0792	0.0795
4	0.0792	0.0792	0.0792



(a) Probability density function for $\sigma_a = 0.05$.

(b) Probability density function for $\sigma_a = 0.1$.

(c) Probability density function for $\sigma_a = 0.15$.

(d) Probability density function for $\sigma_a = 0.2$.

Figure 4. The probability density function of the deflection of the beam. The correlation length of the random field describing the bending rigidity is assumed to be $\mu_a = L/10$. The pdfs are obtained with 10,000 sample MCS as before.

table 2 lists the percentage values of standard deviation calculated as follows

$$\Theta = 100 \left(\frac{\sigma_a - \sigma_{mcs}}{\sigma_{mcs}} \right) \tag{99}$$

where σ_a is the standard deviation of the stochastic system response for a particular order of spectral functions and σ_{mcs} is the same obtained with direct Monte-Carlo simulation. The comparison has again been made for the aforementioned three different cases. The table shows that for system response calculated with higher order of spectral functions the percentage error of the standard deviation converges to the value (0.1483) obtained with the exact approach. This is indicative of the fact that calculation of the values of the undetermined coefficients for each random sample yields better results than those obtained with the Galerkin (or without Galerkin) approach.

Fig. 5 shows the plot of the percentage error of the standard deviation for different degrees of variability (σ_a) of the random parameter (bending stiffness) of the Euler-Bernoulli beam. The figure shows that the error in the standard deviation of the response increases with higher values

Table 2. Percentage error of the measure of standard deviation of the mean deflection of the tip of the beam with respect to the standard deviation of the direct MCS for correlation length of $L/10$ and $\sigma_a = 0.05$.

Order	Galerkin	Exact case	No minimization
1	0.4259	0.1483	0.5299
2	0.1808	0.1483	0.1814
3	0.1504	0.1483	0.1507
4	0.1485	0.1483	0.1485

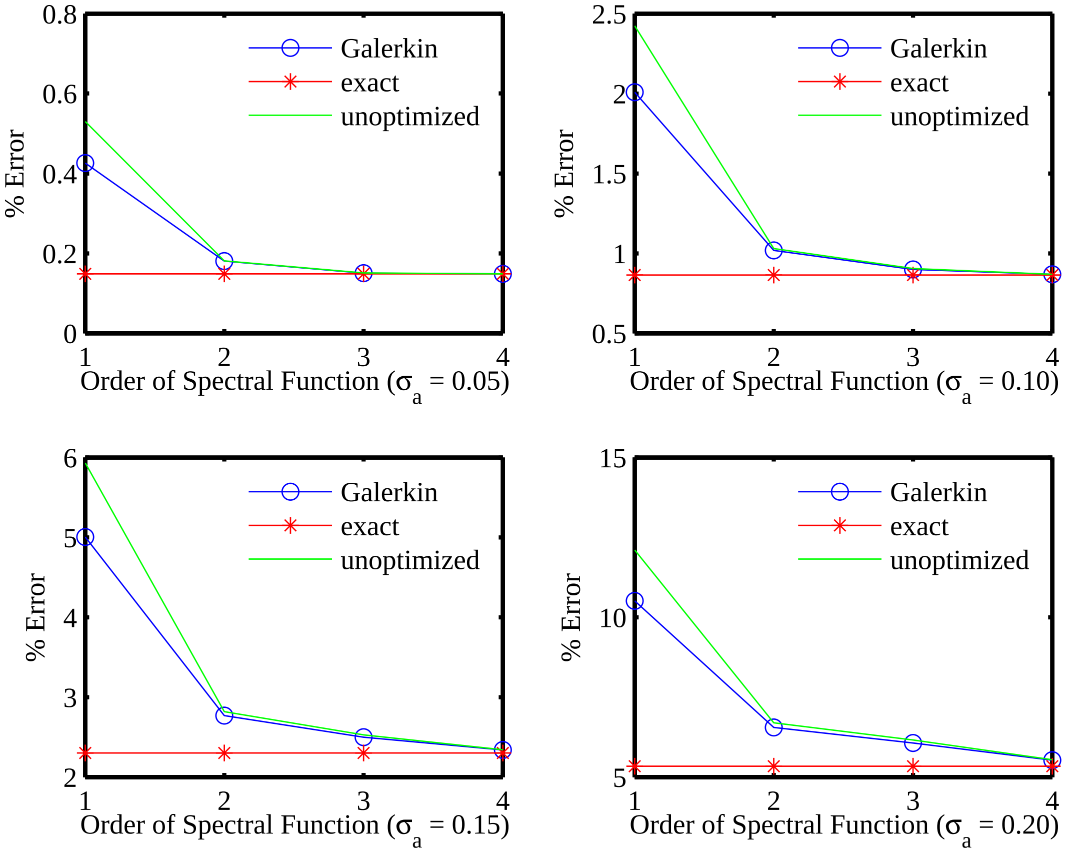


Figure 5. Plot of the percentage error of the standard deviation of the stochastic response as a function of the order of the spectral functions (SF) used to derive the basis functions for different degrees of variability of the random field $\sigma_a = [0.05, 0.10, 0.15, 0.20]$. Correlation length is $L/10$.

of σ_a, and that as the order of the spectral functions are increased the curves converge towards the (almost constant) value of the exact scheme of orthogonalization. This conclusively establishes the accuracy of the exact scheme over the Bubnov-Galerkin. The 'unoptimized' curve indicates the case where the residual has not been optimized with any coefficients and is found to be deviate more from the accurate solution for lower order of the spectral functions and high variability of the random parameter.

IX. Conclusions

We consider discretized stochastic elliptic partial differential equations. In the classical spectral stochastic finite element approach, the solution is projected into an infinite dimensional orthonormal basis functions and the associated constant vectors are obtained using the Galerkin type of error minimization approach. Here an alternative approach is investigated. The solution is projected into a finite dimensional complete orthonormal vector basis and the associated coefficient functions are obtained. The coefficient functions, called as the spectral functions, are expressed in terms of the spectral properties of the deterministic matrices appearing in the discretized governing equation. It is proved that, if infinite order of terms in the functions are retained, then the resulting series converges to the exact solution in probability 1. This is a stronger convergence compared to the classical polynomial chaos which converges in the mean-square sense in the Hilbert space.

The explicit closed-form relationship between the spectral functions and polynomial chaos functions has been derived. It is shown that the spectral functions effectively represent a sum of infinite number of polynomial chaos functions associated with linearly dependent vectors. This significantly reduces the number of unknown constants to be calculated. Since the spectral functions have to be truncated for numerical calculations, an error minimization approach has to be developed. This is done by orthogonalization of the residual vector to the subspace spanned by the stochastic basis functions. The exact scheme of orthogonalization for each random sample has been compared with the Bubnov-Galerkin error minimization scheme, where the latter is known to be the zero-order approximation of the former. The results demonstrate that the exact scheme is more accurate compared to the Bubnov-Galerkin approach. We also establish the relationship between the orthogonalization approach and the direct least square minimization of the residual vector, and it shows that the latter establishes an $\mathbf{A}$-orthogonal relationship between the residue and the basis vectors.

Following this, the Galerkin error minimization approach has been adopted and developed in context of the present stochastic system. It is shown that the number of unknown constants are obtained by solving a system of linear equations which has the same dimension as the original discretized equation. A numerical approach based on the first and second-order spectral functions has been developed. Based on these, a hybrid analytical-simulation approach is proposed to obtain the statistical properties of the solution. The method is applied to a stochastic beam for illustration. The statistics of the deflection were obtained with 24 and 67 random variables and degrees of freedom $n = 100$. A second-order polynomial chaos approach for these problems would require the solution of algebraic equations of dimension 32,400 and 234,500 respectively. In comparison, the proposed Galerkin approach using the Monte Carlo simulation requires the solution of algebraic equations of dimension n only. Promising accuracy compared to the direct Monte Carlo simulation, especially with the second-order spectral function has been observed.

Acknowledgments

SA gratefully acknowledges the support The Royal Society of London through the Wolfson Research Merit Award. AK acknowledges the support Swansea University through the award of Zienkiewicz Fellowship.

References

[1] Nouy, A., "Recent Developments in Spectral Stochastic Methods for the Numerical Solution of Stochastic Partial Differential Equations," *Archives of Computational Methods in Engineering*, Vol. 16, 2009, pp. 251–285.

[2] Charmpis, D. C., Schueeller, G. I., and Pellissetti, M. F., "The need for linking micromechanics of materials with stochastic finite elements: A challenge for materials science," *Computational Materials Science*, Vol. 41, No. 1, 2007, pp. 27–37.

[3] Stefanou, G., "The stochastic finite element method: Past, present and future," *Computer Methods in Applied Mechanics and Engineering*, Vol. 198, No. 9-12, 2009, pp. 1031 – 1051.

[4] Vanmarcke, E. H., *Random fields*, MIT press, Cambridge Mass., 1983.

[5] Ghanem, R. and Spanos, P. D., *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, New York, USA, 1991.

[6] Matthies, H. G., Brenner, C. E., Bucher, C. G., and Soares, C. G., "Uncertainties in probabilistic numerical analysis of structures and solids - Stochastic finite elements," *Structural Safety*, Vol. 19, No. 3, 1997, pp. 283–336.

[7] Papoulis, A. and Pillai, S. U., *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, Boston, USA, 4th ed., 2002.

[8] Matthies, H. G. and Keese, A., "Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations," *Computer Methods in Applied Mechanics and Engineering*, Vol. 194, No. 12-16, 2005, pp. 1295–1331.

[9] Babuska, I., Tempone, R., and Zouraris, G. E., "Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation," *Computer Methods in Applied Mechanics and Engineering*, Vol. 194, No. 12-16, 2005, pp. 1251–1294.

[10] Horn, R. A. and Johnson, C. R., *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.

[11] Kleiber, M. and Hien, T. D., *The Stochastic Finite Element Method*, John Wiley, Chichester, 1992.

[12] Liu, W. K., Belytschko, T., and Mani, A., "Random field finite-elements," *International Journal for Numerical Methods in Engineering*, Vol. 23, No. 10, 1986, pp. 1831–1845.

[13] Yamazaki, F., Shinozuka, M., and Dasgupta, G., "Neumann expansion for stochastic finite element analysis," *Journal of Engineering Mechanics-ASCE*, Vol. 114, No. 8, 1988, pp. 1335–1354.

[14] Adhikari, S. and Manohar, C. S., "Dynamic analysis of framed structures with statistical uncertainties," *International Journal for Numerical Methods in Engineering*, Vol. 44, No. 8, 1999, pp. 1157–1178.

[15] Grigoriu, M., "Galerkin solution for linear stochastic algebraic equations," *Journal of Engineering Mechanics-Asce*, Vol. 132, No. 12, 2006, pp. 1277–1289.

[16] Falsone, G. and Impollonia, N., "A new approach for the stochastic analysis of finite element modelled structures with uncertain parameters," *Computer Methods in Applied Mechanics and Engineering*, Vol. 191, No. 44, 2002, pp. 5067–5085.

[17] Li, C. F., Feng, Y. T., and Owen, D. R. J., "Explicit solution to the stochastic system of linear algebraic equations $(\alpha_1 A_1 + \alpha_2 A_2 + \cdots + \alpha_m A_m)x = b$," *Computer Methods in Applied Mechanics and Engineering*, Vol. 195, No. 44-47, 2006, pp. 6560–6576.

[18] Feng, Y. T., "Adaptive preconditioning of linear stochastic algebraic systems of equations," *Communications in Numerical Methods in Engineering*, Vol. 23, No. 11, 2007, pp. 1023–1034.

[19] Foo, J. and Karniadakis, G. E., "Multi-element probabilistic collocation method in high dimensions," *Journal of Computational Physics*, Vol. 229, No. 5, 2010, pp. 1536 – 1557.

[20] Ma, X. and Zabaras, N., "An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations," *Journal of Computational Physics*, Vol. 228, No. 8, 2009, pp. 3084–3113.

[21] Nair, P. B. and Keane, A. J., "Stochastic reduced basis methods," *AIAA Journal*, Vol. 40, No. 8, 2002, pp. 1653–1664.

[22] Xiu, D. B. and Karniadakis, G. E., "The Wiener-Askey polynomial chaos for stochastic differential equations," *Siam Journal on Scientific Computing*, Vol. 24, No. 2, 2002, pp. 619–644.

[23] Wan, X. L. and Karniadakis, G. E., "Beyond wiener-askey expansions: Handling arbitrary pdfs," *Journal of Scientific Computing*, Vol. 27, No. (-3, 2006, pp. 455–464.

[24] Sarkar, A., Benabbou, N., and Ghanem, R., "Domain decomposition of stochastic PDEs: Theoretical formulations," *International Journal for Numerical Methods in Engineering*, Vol. 77, No. 5, 2009, pp. 689–701.

[25] Blatman, G. and Sudret, B., "An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis," *Probabilistic Engineering Mechanics*, Vol. 25, No. 2, 2010, pp. 183 – 197.

[26] Nouy, A., "A generalized spectral decomposition technique to solve a class of linear stochastic partial differential equations," *Computer Methods in Applied Mechanics and Engineering*, Vol. 196, No. 45-48, 2007, pp. 4521–4537.

[27] Nouy, A., "Generalized spectral decomposition method for solving stochastic finite element equations: Invariant subspace problem and dedicated algorithms," *Computer Methods in Applied Mechanics and Engineering*, Vol. 197, No. 51-52, 2008, pp. 4718–4736.

[28] Luenberger, D. G., *Optimization by Vector Space Methods*, John Wiley & Sons, NY, USA, 1969.

[29] Wiener, N., "The homogeneous chaos," *American Journal of Mathematics*, Vol. 60, No. 4, 1938, pp. 897–936.

[30] Nair, P. B., "On the theoretical foundations of stochastic reduced basis methods," *AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and Exhibit, 42nd, Seattle, WA*, Vol. AIAA-2001-1677, 2001.

[31] Xiu, D. B. and Karniadakis, G. E., "Modeling uncertainty in flow simulations via generalized polynomial chaos," *Journal of Computational Physics*, Vol. 187, No. 1, 2003, pp. 137–167.

[32] Zienkiewicz, O. C. and Taylor, R. L., *The Finite Element Method*, McGraw-Hill, London, 4th ed., 1991.