# Non-adapted sparse approximation of PDEs with stochastic inputs

**By A. Doostan†, H. Owhadi‡, A. Lashgari‡ AND G. Iaccarino**

## 1. Motivation and objectives

Realistic analysis and design of complex multiscale and multidisciplinary systems require not only a fine understanding and modeling of the underlying physics and their interactions but also recognition of intrinsic *uncertainties* and their influences on the quantities of interest. Uncertainty Quantification (UQ) is an emerging discipline that tries to address the latter issue: It aims at meaningful *characterization* of uncertainties in the physical models from the available measurements and efficient *propagation* of these uncertainties for a quantitative validation of model predictions.

Despite recent growing interest in UQ of complex systems, it remains as a grand challenge to efficiently propagate uncertainties through systems characterized by a large number of uncertain sources where the so-called curse of dimensionality is yet an unsolved problem. Additionally, development of *non-intrusive* uncertainty propagation techniques is of essence as the analysis of complex multidisciplinary systems often requires the use of sophisticated coupled deterministic solvers in which one cannot readily intrude to set up the necessary propagation infrastructure.

Sampling methods such as Monte Carlo simulation and its several variants have long been utilized as the primary scheme for uncertainty propagation. However, these methods are generally inefficient for large-scale systems due to their slow rate of convergence. There has been an increasing recent interest in developing more efficient alternative numerical methods. Most notably *stochastic Galerkin* schemes using polynomial chaos (PC) bases (Ghanem & Spanos 2003; Deb *et al.* 2001; Xiu & Karniadakis 2002; Babuška *et al.* 2004) have been applied successfully to a variety of engineering problems and are extremely useful when the number of uncertain parameters is not large. In their original form, the stochastic Galerkin schemes are *intrusive* as one has to modify the deterministic solvers for their implementation. *Stochastic collocation* schemes (Tatang 1995; Mathelin & Hussaini 2003; Xiu & Hesthaven 2005; Babuška *et al.* 2007; Nobile *et al.* 2008) belong to a different class of methods that rely upon (isotropic) sparse grid integration/interpolation in the stochastic space of the problem to reduce the curse of dimensionality associated with the conventional tensor-product integration/interpolation rules. As their construction is based primarily on the input parameter space, the computational cost of both stochastic Galerkin and stochastic collocation techniques increases rapidly when the number of input uncertainties is large.

More recently, efforts have been made to construct *solution-adaptive* uncertainty propagation techniques that exploit any structures in the solution to decrease the computational cost. Among these are the multiscale model reduction of Doostan *et al.* (2007) and sparse decomposition of Todor & Schwab (2007), Bieri & Schwab (2009), Bieri (2009), and Bieri *et al.* (2009) for the stochastic Galerkin technique, anisotropic and adaptive

† Affiliates University of Colorado, Boulder, CO, USA
‡ Affiliates California Institute of Technology, Pasadena, CA, USA.

sparse grids of Nobile *et al.* (2007) and Ma & Zabaras (2009) for the stochastic collocation scheme, and low-rank solution approximations of Nouy (2007, 2008) and Doostan & Iaccarino (2008).

In the present study, we discuss cases where the quantity of interest is *sparse* at the stochastic level, i.e., it can be accurately represented with only *a few* terms when linearly expanded into a stochastic, e.g., the polynomial chaos, basis. Interestingly, sparsity is salient in the analysis of high-dimensional problems where the number of energetic basis functions (those with large coefficients) is small relative to the cardinality of the full basis. For instance, it has been shown (Todor & Schwab 2007; Bieri & Schwab 2009) that, under some mild conditions, solutions to linear elliptic stochastic PDEs with high-dimensional random coefficients admit sparse representations with respect to the polynomial chaos basis. Consequently, an approach based on a zero-dimensional algebraic stochastic problem has been proposed (Bieri & Schwab 2009) to detect the sparsity pattern which then guides the stochastic Galerkin analysis of the original problem. Moreover a "quasi"-best $N$-term approximation for a class of elliptic stochastic PDEs has been proposed (Bieri *et al.* 2009).

In this work, using techniques based on *concentration of measure theory* and *compressive sensing* we derive a method for approximation of sparse solution of stochastic PDEs. The proposed method is

- Non-intrusive: it is based on the direct, i.e., non adapted, sampling of solutions. This sampling can be done by using any legacy code for the deterministic problem as a black box.
- Provably convergent: we obtain probabilistic bounds on the approximation error proving the convergence of the method in probability.
- Well suited for truly high dimensional problems (with slow decay in the spectrum).

*Compressive sampling* is an emerging direction in signal processing that aims at recovering *sparse* signals accurately (or even exactly) from a small number of their random projections (Chen *et al.* 2001; Candes *et al.* 2006; Donoho 2006; Candes & Romberg 2006; Candes & Tao 2006; Cohen *et al.* 2009; Bruckstein *et al.* 2009). A sparse signal is simply one with only few significant coefficients when linearly expanded into a basis $\{\psi_\alpha\}$.

For sufficiently sparse signals, the number of samples required is typically less than what is required by the Shannon-Nyquist sampling principle. Generally speaking, a successful signal reconstruction by compressive sampling is conditioned upon:

- *Sufficient* sparsity of the signal; and
- *Incoherent* (random) projections of signal.

A square-measurable stochastic function $u(\omega)$, defined on a suitable probability space $(\Omega, \mathcal{F}, \mathcal{P})$ can be expanded into a mean-squared convergent series of chaos polynomial bases, i.e., $u(\omega) \approx \sum_\alpha c_\alpha \psi_\alpha(\omega)$, with some cardinality $P$. The stochastic function $u(\omega)$ is then sparse in PC basis $\{\psi_\alpha\}$, if only a small fraction of coefficients $c_\alpha$ are significant. In this case, under certain conditions, the sparse PC coefficients $\boldsymbol{c}$ can be computed accurately and robustly using only $N << P$ random samples of $u(\omega)$ via compressive samplig. Given the $N$ random samples of $u(\omega)$, compressive sampling aims at finding the sparsest (or nearly sparsest) coefficients $\boldsymbol{c}$ from an optimization problem of the form

$$\min_{\boldsymbol{c}} \|\boldsymbol{c}\|_s \quad \text{subject to} \quad \|\boldsymbol{\Psi c} - \boldsymbol{u}\|_r \leq \delta_{p,N} \tag{1.1}$$

where $\|\boldsymbol{c}\|_s$ is a measure of the sparsity of $\boldsymbol{c}$ and $\|\boldsymbol{\Psi c} - \boldsymbol{u}\|_r$ is a measure of the accuracy

of the truncated PC basis in approximation of $u(\omega)$ samples. The $N$-vector $\boldsymbol{u}$ contains the independent random samples of $u(\omega)$ and the rows of the $N \times P$ matrix $\boldsymbol{\Psi}$ consist of the corresponding samples of the PC basis $\{\psi_\alpha\}$.

We will elaborate on the formulation of the compressive sampling problem (1.1) and the required conditions under which it leads to an accurate and stable approximation of the sparse solution of linear elliptic stochastic PDEs. In Section 2.2, we briefly overview the *spectral stochastic* discretization of stochastic PDEs using PC basis. We then summarize the analysis of Todor & Schwab (2007) and Bieri & Schwab (2009) on the sparsity of the solution of a class of linear elliptic PDEs with random inputs in Section 2.3. Section 2.4 introduces our proposed scheme based on the compressive sampling to approximate the sparse solution. Finally, in Section 3, a numerical experiment is performed on a 1-$D$ elliptic stochastic PDE with random diffusion coefficient to demonstrate the accuracy of the proposed procedure.

## 2. Numerical method

### 2.1. *Problem set up*

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a complete probability space where $\Omega$ is the set of elementary events, $\mathcal{F}$ is a $\sigma-$field, and $\mathcal{P}$ is a probability measure. We consider the following elliptic stochastic PDE defined on a bounded domain $\mathcal{D} \subset \mathbb{R}^D$ with boundary $\partial\mathcal{D}$,

$$\begin{aligned} -\nabla \cdot (a(\boldsymbol{x}, \omega)\nabla u(\boldsymbol{x}, \omega)) &= f(\boldsymbol{x}) \quad \boldsymbol{x} \in \mathcal{D}, \\ u(\boldsymbol{x}, \omega) &= 0 \quad \boldsymbol{x} \in \partial\mathcal{D}, \end{aligned} \tag{2.1}$$

$\mathcal{P} - a.s.\ \omega \in \Omega$. The diffusion coefficient $a(\boldsymbol{x}, \omega)$ is a stochastic function defined on $(\Omega, \mathcal{F}, \mathcal{P})$ and is the source of uncertainty in (2.1). We assume that $a(\boldsymbol{x}, \omega)$ is specified by a truncated Karhunen-Loève-like expansion

$$a(\boldsymbol{x}, \omega) = \bar{a}(\boldsymbol{x}) + \sum_{i=1}^{d} \sqrt{\lambda_i}\phi_i(\boldsymbol{x})y_i(\omega), \tag{2.2}$$

where $(\lambda_i, \phi_i)$, $i = 1, \cdots, d$, are the eigenpairs of the covariance function $C_a(\boldsymbol{x}_1, \boldsymbol{x}_2) \in L_2(\mathcal{D} \times \mathcal{D})$ of $a(\boldsymbol{x}, \omega)$ and $\bar{a}(\boldsymbol{x})$ is the mean of $a(\boldsymbol{x}, \omega)$. We further assume that $a(\boldsymbol{x}, \omega)$ satisfies the following conditions:

**A-I.** For all $\boldsymbol{x} \in \mathcal{D}$, there exist constants $a_{\min}$ and $a_{\max}$ such that

$$0 < a_{\min} \leq a(\boldsymbol{x}, \omega) \leq a_{\max} < \infty \quad \mathcal{P} - a.s.\ \omega \in \Omega \tag{2.3}$$

**A-II.** The covariance function $C_a(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is *piecewise analytic* on $\mathcal{D} \times \mathcal{D}$ (Bieri & Schwab 2009) implying that there exist real constants $c_1$ and $c_2$ such that for $i = 1, \cdots, d$,

$$0 \leq \lambda_i \leq c_1 e^{-c_2 i^{\kappa}} \tag{2.4}$$

and

$$\forall \boldsymbol{\alpha} \in \mathbb{N}^d : \qquad \sqrt{\lambda_i}\|\partial^{\boldsymbol{\alpha}}\phi_i\|_{L^\infty(\mathcal{D})} \leq c_1 e^{-c_2 i^{\kappa}}, \tag{2.5}$$

where $\kappa := 1/D$ and $\boldsymbol{\alpha} \in \mathbb{N}^d$ is a fixed multi-index. Notice that the decay rates in Eqs. (2.4) and (2.5) will be *algebraic* if $C_a(\boldsymbol{x}_1, \boldsymbol{x}_2)$ has $C^s(\mathcal{D} \times \mathcal{D})$ regularity for some $s > 0$.

**A-III.** The random variables $\{y_k(\omega)\}_{k=1}^d$ are independent and uniformly distributed on $\Gamma_k := [v_k, w_k]$, $k = 1, \cdots, d$, with probability distribution function $\rho_k(y_k) = 1/(w_k - v_k)$, with $w_k > v_k$, defined over $\Gamma_k$. The joint probability distribution function of the random vector $\boldsymbol{y} := \{y_k(\omega)\}_{k=1}^d$ is then given by $\rho(\boldsymbol{y}) := \prod_{k=1}^d \rho_k(y_k)$.

Given the assumption $(c)$ above, the solution $u(\boldsymbol{x}, \omega)$ of Eq. (2.1) also admits a finite dimensional representation, i.e.,

$$u(\boldsymbol{x}, \boldsymbol{y}) := u(\boldsymbol{x}, y_1(\omega), \cdots, y_d(\omega)) : \mathcal{D} \times \Gamma \to \mathbb{R}, \tag{2.6}$$

where $\Gamma := \prod_{k=1}^d \Gamma_k$. In the sequel, we briefly outline the *spectral stochastic* discretization of $u(\boldsymbol{x}, \boldsymbol{y})$.

### 2.2. *Spectral stochastic discretization*

Let $\mathbb{N}_0^d := \{(\alpha_1, \cdots, \alpha_d) : \alpha_j \in \mathbb{N} \cup \{0\}\}$ be the set of multi-indices of size $d$ defined on non-negative integers. In the context of the *spectral stochastic* methods, the solution $u(\boldsymbol{x}, \boldsymbol{y})$ of Eq. (2.1) is represented by an infinite series of the form

$$u(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}_0^d} c_{\boldsymbol{\alpha}}(\boldsymbol{x}) \psi_{\boldsymbol{\alpha}}(\boldsymbol{y}). \tag{2.7}$$

The basis functions $\{\psi_{\boldsymbol{\alpha}}(\boldsymbol{y})\}$ are multidimensional Legendre polynomials, referred to as Legendre polynomial chaos (PC), that are orthogonal with respect to the joint probability measure $\rho(\boldsymbol{y})$ of the random vector $\boldsymbol{y}$. Each basis function $\psi_{\boldsymbol{\alpha}}(\boldsymbol{y})$ is a tensor product of univariate Legendre polynomials $\psi_{\alpha_i}(y_i)$ of degree $\alpha_i$,

$$\psi_{\boldsymbol{\alpha}}(\boldsymbol{y}) = \psi_{\alpha_1}(y_1)\psi_{\alpha_2}(y_2)\cdots\psi_{\alpha_d}(y_d) \qquad \boldsymbol{\alpha} \in \mathbb{N}_0^d. \tag{2.8}$$

We here assume that the univariate Legendre polynomials $\psi_{\alpha_i}(y_i)$ are also normalized thus satisfying the orthonormality condition

$$\int_{\Gamma_i} \psi_{\alpha_i}(y_i)\psi_{\alpha_i'}(y_i)\rho_i(y_i)dy_i = \delta_{\alpha_i\alpha_i'}, \quad i = 1, \cdots, d, \tag{2.9}$$

where $\delta_{\alpha_i\alpha_i'}$ is the Kronecker delta.

The generalized Fourier coefficients $c_{\boldsymbol{\alpha}}(\boldsymbol{x})$ in Eq. (2.7), referred to as the PC coefficients, are computed by projection of $u(\boldsymbol{x}, \boldsymbol{y})$ onto each basis function $\psi_{\boldsymbol{\alpha}}(\boldsymbol{y})$,

$$c_{\boldsymbol{\alpha}}(\boldsymbol{x}) := \mathbb{E}\left[u(\boldsymbol{x}, \boldsymbol{y})\psi_{\boldsymbol{\alpha}}(\boldsymbol{y})\right] = \int_\Gamma u(\boldsymbol{x}, \boldsymbol{y})\psi_{\boldsymbol{\alpha}}(\boldsymbol{y})\rho(\boldsymbol{y})d\boldsymbol{y}. \tag{2.10}$$

In practice, the expansion in Eq. (2.7) is finite, that is only a finite number of spectral modes is needed to approximate $u(\boldsymbol{x}, \boldsymbol{y})$ given a target accuracy. Traditionally, a finite order truncation of the basis $\{\psi_{\boldsymbol{\alpha}}(\boldsymbol{y})\}$ is adopted for approximation, i.e.,

$$u_{p,d}(\boldsymbol{x}, \boldsymbol{y}) := \sum_{\boldsymbol{\alpha} \in \Lambda_{p,d}} c_{\boldsymbol{\alpha}}(\boldsymbol{x})\psi_{\boldsymbol{\alpha}}(\boldsymbol{y}), \tag{2.11}$$

where the set of multi-indices $\Lambda_{p,d}$ is

$$\Lambda_{p,d} := \left\{ \boldsymbol{\alpha} \in \mathbb{N}_0^d : \|\boldsymbol{\alpha}\|_1 \leq p,\ \|\boldsymbol{\alpha}\|_0 \leq d \right\} \tag{2.12}$$

and has the cardinality

$$P := |\Lambda_{p,d}| = \frac{(p+d)!}{p!d!}. \tag{2.13}$$

Here $\|\boldsymbol{\alpha}\|_1 = \sum_{i=1}^d \boldsymbol{\alpha}_i$ and $\|\boldsymbol{\alpha}\|_0 = \#\{i : \alpha_i > 0\}$ are the total order (degree) and dimensionality of the basis function $\psi_{\boldsymbol{\alpha}}(\boldsymbol{y})$. The approximation is then refined by increasing $p$ to achieve a given target accuracy.

### 2.3. *Sparsity of the solution*

Alternative to the $p$th-order truncated PC expansion of (2.15), one may ideally seek a proper index set $\Lambda_{p,d}^{\epsilon} \subseteq \Lambda_{p,d}$, with sufficiently large $p$, such that

$$\Lambda_{p,d}^{\epsilon} := \arg\min \left\{ |\tilde{\Lambda}_{p,d}| : \tilde{\Lambda}_{p,d} \subseteq \Lambda_{p,d}, \|u - \tilde{u}_{p,d}\|_{L^\infty(\Gamma)\otimes H_0^1(\mathcal{D})} \leq \epsilon \right\}, \tag{2.14}$$

in which

$$\tilde{u}_{p,d}(\boldsymbol{x},\boldsymbol{y}) := \sum_{\boldsymbol{\alpha}\in\tilde{\Lambda}_{p,d}\subseteq\Lambda_{p,d}} c_{\boldsymbol{\alpha}}(\boldsymbol{x})\psi_{\boldsymbol{\alpha}}(\boldsymbol{y}). \tag{2.15}$$

This will lead to the so-called *sparse approximation* of $u(\boldsymbol{x},\boldsymbol{y})$ if

$$|\Lambda_{p,d}^{\epsilon}| << |\Lambda_{p,d}| = P. \tag{2.16}$$

Such a reduction in the number of basis functions in (2.16) is possible as, given an accuracy $\epsilon$, the *effective* dimensionality $\nu$ of $u(\boldsymbol{x},\boldsymbol{y})$ in $\Gamma$ is potentially smaller than $d$, the size of random vector $\boldsymbol{y}$. More precisely, under assumptions **A-I**, **A-II**, and **A-III** stated in Section 2.1, (Todor & Schwab 2007; Bieri & Schwab 2009) show that the discretization of $u(\boldsymbol{x},\boldsymbol{y})$ using a sparse index set $\Lambda_{p,\nu}$,

$$\Lambda_{p,\nu} := \left\{ \alpha \in \mathbb{N}_0^d : \|\alpha\|_1 \leq p,\ \|\alpha\|_0 \leq \nu \leq d \right\} \tag{2.17}$$

preserves the exponential decay of the approximation error in $L^\infty(\Gamma)\otimes H_0^1(\mathcal{D})$. For completeness, we cite this from Bieri & Schwab (2009) in the following lemma.

**Lemma 1.** Proposition 3.10 of Bieri & Schwab (2009) *Given assumptions* **A-I**, **A-II**, *and* **A-III** *in Section 2.1, there exist constants* $c_1, c_2, c_3, c_4 > 0$*, depending only on* $a(\boldsymbol{x},\omega)$ *and* $f(\boldsymbol{x})$ *but independent of* $d, p, \nu$ *such that*

$$\|u - u_{p,\nu}\|_{L^\infty(\Gamma)\otimes H_0^1(\mathcal{D})} \leq c_1 \left( e^{-c_2\nu^{1+\kappa}} + e^{c_3\nu(\ln d + \ln p) - c_4 p} \right), \tag{2.18}$$

*for any* $d, p, \nu \in \mathbb{N}$ *with* $\nu \leq d$ *and* $\kappa = 1/D$.

Particularly, for $d \geq c_d|\ln\epsilon|^{1/\kappa}$, choosing

$$p(\epsilon) = \lceil c_p d^{\kappa} \rceil \quad \text{and} \quad \nu(\epsilon) = \lceil c_\nu d^{\kappa/(\kappa+1)} \rceil, \tag{2.19}$$

leads to

$$\|u - u_{p,\nu}\|_{L^\infty(\Gamma)\otimes H_0^1(\mathcal{D})} \leq \epsilon, \tag{2.20}$$

where

$$|\Lambda_{p,\nu}| \lesssim \epsilon^{-1/s}, \tag{2.21}$$

for some arbitrary large $s > 0$ and constants $c_d, c_p$, and $c_\nu$ independent of $d, p$, and $\nu$ (Bieri & Schwab 2009).

In practice, the sparse set $\Lambda_{p,d}^\epsilon$ in Eq. (2.14) (equivalently $\Lambda_{p,\nu}$ in Eq. (2.21)) is not known a priori. In Bieri & Schwab (2009) a *heuristic* approach based on an algebraic purely stochastic problem is proposed to adaptively identify $\Lambda_{p,d}^\epsilon$. Having done this, the coefficients of the the spectral modes are computed via the intrusive stochastic Gakerkin scheme (Ghanem & Spanos 2003; Xiu & Karniadakis 2002). Alternatively, in the present work, we propose a *non-intrusive* and *non-adaptive* algorithm within the framework of *compressing sampling* for the sparse spectral stochastic approximation of $u(\boldsymbol{x}, \boldsymbol{y})$.

### 2.4. *Sparse approximation using compressive sampling*

*Compressive sampling* is an emerging theory in signal and image processing that hinges around the idea that a *sparse* signal can be accurately or exactly reconstructed using only a *small* number (much smaller than the cardinality of the signal) of its random projections. A sparse signal is one that can be represented using a small number of terms when expanded into a basis/frame. In the context of sparse solution of problem (2.1), compressive sampling can be interpreted as follows. If the solution $u(\boldsymbol{x}, \boldsymbol{y})$ is *sparse* at the stochastic level, i.e., it has a PC expansion with only small number of significant coefficients, it can be accurately recovered using $N << P$ samples $\{u(\boldsymbol{x}, \boldsymbol{y}_i)\}_{i=1}^N$, where $P$ is the cardinality of the PC basis $\{\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{y})\}$. We attempt now to elaborate on the above statement and address how such a reconstruction is achieved and under what conditions it is successful.

Let $\{u(\boldsymbol{y}_i)\}_{i=1}^N$ be *i.i.d.* samples of $u(\boldsymbol{x}, \boldsymbol{y})$ for a fixed point $\boldsymbol{x}$ in $\mathcal{D}$. For the time being, let us assume that the $p$th-order PC basis $\{\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\boldsymbol{y})\}$ is a complete basis to expand $u(\boldsymbol{y})$; we relax this assumption as we proceed. Due to the analyticity of $u(\boldsymbol{y})$ with respect to $\boldsymbol{y}$, (Babuška *et al.* 2007), we can write

$$u(\boldsymbol{y}_i) = \sum_{\boldsymbol{\alpha}\in\Lambda_{p,d}} c_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{y}_i) \quad i = 1, \cdots, N, \tag{2.22}$$

or equivalently,

$$\boldsymbol{\Psi}\boldsymbol{c} = \boldsymbol{u}. \tag{2.23}$$

Where $\boldsymbol{c} \in \mathbb{R}^P$ is the vector of PC coefficients to be determined, $\boldsymbol{u} \in \mathbb{R}^N$ is the vector of samples of $u(\boldsymbol{y})$, and each column of the *measurement* matrix $\boldsymbol{\Psi} \in \mathbb{R}^{N\times P}$ contains samples of the PC basis, i.e.,

$$\boldsymbol{\Psi}[i, j] = \psi_{\boldsymbol{\alpha}_j}(\boldsymbol{y}_i), \quad i = 1, \cdots, N, \quad j = 1, \cdots, P. \tag{2.24}$$

We are interested in the case the number $N$ of solution samples is much smaller than the unknown PC coefficients $P$, i.e., $N << P$. Without any additional constraints on $\boldsymbol{c}$, the underdetermined linear system (2.23) is illposed and, in general, has infinitely many solutions. From the analysis of Section 2.3, we know that the vector of coefficients $\boldsymbol{c}$ is

sparse: only a few of the coefficients are significant. Without knowing such a sparsity pattern, we henceforth seek a solution $\boldsymbol{c}$ that has minimum number of nonzeros. This can be readily formulated in the optimization problem

$$(P_0): \qquad \min_{\boldsymbol{c}} \|\boldsymbol{c}\|_0 \quad \text{subject to} \quad \boldsymbol{\Psi c} = \boldsymbol{u}, \tag{2.25}$$

where the semi-norm $\|\boldsymbol{c}\|_0 := \#\{\boldsymbol{\alpha} : c_{\boldsymbol{\alpha}} \neq 0\}$ is the number of nonzero components of $\boldsymbol{c}$. In general, the global minimum solution of $(P_0)$ is not unique and $NP$-hard to compute: The cost of a global search is exponential in $P$. Further development in compressive sampling resulted in a convex relaxation of problem $(P_0)$ by minimization of the $\ell_1$ norm of the solution $\boldsymbol{c}$ instead, i.e.,

$$(P_1): \qquad \min_{\boldsymbol{c}} \|\boldsymbol{c}\|_1 \quad \text{subject to} \quad \boldsymbol{\Psi c} = \boldsymbol{u}, \tag{2.26}$$

where $\|\boldsymbol{c}\|_1 = \sum_{\boldsymbol{\alpha}} |c_{\boldsymbol{\alpha}}|$ is the $\ell_1$ norm of $\boldsymbol{c}$. Since the $\ell_1$ norm functional $\|\boldsymbol{c}\|_1$ is convex, the optimization problem $(P_1)$ admits a unique solution which is also the unique solution to problem $(P_0)$ for sufficiently sparse $\boldsymbol{c}$ and with some constraints on $\boldsymbol{\Psi}$ (Bruckstein *et al.* 2009). Among several available solvers, modern interior-point methods can efficiently solve $(P_1)$ even for large-scale systems where $d$ and $p$ (and thus $P$) are large.

In general, a $p$th-order PC basis is not complete for the exact representation of $u(\boldsymbol{y})$ and, therefore, we have to account for the truncation error. This can be accommodated, for instance in $(P_1)$, by allowing a nonzero residual in Eq. (2.23). Therefore, as in Section 3.2.1 of Bruckstein *et al.* (2009), the proposed algorithm in this paper is an error-tolerant version of $(P_1)$, with error tolerance $\delta_{p,N}$:

$$(P_{1,\delta}): \qquad \min_{\boldsymbol{c}} \|\boldsymbol{c}\|_1 \quad \text{subject to} \quad \|\boldsymbol{\Psi c} - \boldsymbol{u}\|_2 \leq \delta_{p,N} \tag{2.27}$$

Theorem 2, then shows that (modulo a control on the coherence matrix given by Theorem 1) if problem (2.27) admits a sparse solution then it is at a distance proportional to $\delta_{p,N}$ for the true solution $u$ in mean squared norm. Hence one must choose $\delta_{p,N}$ as small as possible but so that with high probability problem a sparse solution satisfies the inequality constraint of (2.27).

Let $\boldsymbol{c}_u$ be the projection coefficients of $u$ onto the basis $\Psi$. Define

$$\sigma_{p,N} := \|\boldsymbol{\Psi c}_u - \boldsymbol{u}\|_2 \tag{2.28}$$

Observe that $\sigma_{p,N}$ controls the $\ell_2$ norm of the residual and depends on the order $p$ of the PC expansion as well as the number of samples $N$. If $\delta_{P,N} > \zeta\sigma_{p,N}$ with $\zeta > 1$ then, with high probability (for large $N$), problem (2.27) admits a sparse solution. Indeed, notice that by the strong law of large numbers

$$\frac{1}{N}\sigma_{p,N}^2 \xrightarrow{a.s.} \sigma_p^2 \quad \text{as} \quad N \longrightarrow \infty, \tag{2.29}$$

where $\sigma_p^2 := \mathbb{E}[(u - u_{p,d})^2]$ is the mean-squares error of the $p$th-order PC expansion of $u(\boldsymbol{y})$ whose upper bound asymptotically decays exponentially fast as $p$ increases according to Lemma 1. In practice, the exact value of $\sigma_{p,N}$ (and $\sigma_p$) is not known a priori and has to be estimated through statistical tools such as cross-validation.

The success of problem $(P_{1,\sigma})$ in accurately approximating the sparse PC coefficients $\boldsymbol{c}$ is not readily evident from its construction as we require $N << P$ samples of the solution. We next briefly introduce the general factors involved in the success of compressive

sampling. We furthermore extend the sufficient requirements of Donoho *et al.* (2006) on problem $(P_{1,\sigma})$ for a stable and accurate approximation of sparse PC coefficients $\boldsymbol{c}$ using $N << P$ samples.

2.4.1. *Ingredients of a successful recovery*

The ability of problem $(P_{1,\sigma})$ to accurately approximate the sparse PC coefficients $\boldsymbol{c}$, and hence $u(\boldsymbol{y})$, depends on two main factors: *a*) sparsity of $\boldsymbol{c}$ and *b*) *mutual coherence* of measurement matrix $\boldsymbol{\Psi}$. In Section 2.2, we rationalized the sparsity property of $\boldsymbol{c}$ based on the analysis of Bieri & Schwab (2009). In the present section, we focus on quantifying sufficient requirements on the sparsity of $\boldsymbol{c}$ together with the mutual coherence of $\boldsymbol{\Psi}$, for a success recovery of solution $u(\boldsymbol{y})$ via $(P_{1,\sigma})$. Let us first start by mutual defining the coherence of $\boldsymbol{\Psi}$.

**Definition 1 (Mutual Coherence)** (Donoho *et al.* 2006). *The mutual coherence $\mu(\boldsymbol{\Psi})$ of a matrix $\boldsymbol{\Psi} \in \mathbb{R}^{N\times P}$ is the maximum of absolute normalized inner products of its columns. Let $\boldsymbol{\psi}_j$ and $\boldsymbol{\psi}_k$ be two distinct columns of $\boldsymbol{\Psi}$. Then*

$$\mu(\boldsymbol{\Psi}) := \max_{1\le j,k\le P,\ j\neq k} \frac{|\boldsymbol{\psi}_j^T \boldsymbol{\psi}_k|}{\|\boldsymbol{\psi}_j\|_2 \|\boldsymbol{\psi}_k\|_2}. \tag{2.30}$$

In other words, the mutual coherence is a measure of how close to orthogonal a matrix is. Clearly for a general matrix

$$0 \le \mu(\boldsymbol{\Psi}) \le 1, \tag{2.31}$$

where the lower bound is achieved, for instance, by unitary matrices. However, for the case of $N < P$ the mutual coherence is strictly positive. Given that matrices with smaller mutual coherence are better able to recover a sparse solution, we examine the mutual coherence of the random measurement matrix $\boldsymbol{\Psi}$ in Eq. (2.23). We first observe that, by the orthogonality of the Legendre PC basis, the mutual coherence $\mu(\boldsymbol{\Psi})$ converges to zero almost surely for asymptotically large random sample sizes $N$. However, it is essential for our purpose to investigate if a desirably small $\mu(\boldsymbol{\Psi})$ can be achieved by a sample size $N << P$ and to quantify how large $\mu(\boldsymbol{\Psi})$ can get for a *finite* $N$. These are stated in the following theorem.

**Theorem 1.** *Let $\boldsymbol{\Psi} \in \mathbb{R}^{N\times P}$ be the measurement matrix corresponding to $N$ independent samples of the Legendre polynomial chaos basis of order $p$ in $d$ i.i.d. uniform random variables $\boldsymbol{y}$ as defined in Eq. (2.23). There exists a positive constant $C_\gamma$ depending only on $\gamma := d/p$ such that*

- *If*

$$0 \le r := 2(1+\zeta)\sqrt{(P^{4C_\gamma} \ln P)/N} \le 1/2, \tag{2.32}$$

*for some $\zeta > 0$, then*

$$\mathbb{P}\left[\mu(\boldsymbol{\Psi}) \ge \frac{r}{1-r}\right] \le 4P^{2-2(1+\zeta)^2}, \tag{2.33}$$

- *$C_\gamma < \frac{1}{4}$ for $\gamma \ge 2.1896$.*

**Remark 1.** Theorem 1 simply says that for cases where the number $d$ of random variables $\boldsymbol{y}$ representing the uncertainty in the problem is at least twice that of the

required order $p$ of PC expansion of the solution $u(\boldsymbol{y})$, it is sufficient to have $N \sim \mathcal{O}(4P^{4C_\gamma} \ln P) << P$ to keep $\mu(\boldsymbol{\Psi})$ bounded from above with a large probability. Notice that such a requirement on $\gamma$ is particularly suited to high-dimensional problems.

**Remark 2.** We observe that, given the definition of $r$ in Eq. (2.32), the upper bound of $\mu(\boldsymbol{\Psi})$ in (2.33) decays like $1/\sqrt{N}$ for asymptotically large $N$, which is consistent with the central limit theorem.

To summarize, given the discussions of Section 2.2 we know that the solution to problem (2.1) is sparse in the Legendre polynomial chaos basis. We also observed that the mutual coherence $\mu(\Psi)$ of measurement matrix $\boldsymbol{\Psi}$ in Eq. (2.23) can be arbitrarily bounded from above by increasing the number $N$ of independent solution samples. Following (Donoho *et al.* 2006; Bruckstein *et al.* 2009), we now state a sufficient condition on sparsity (or equivalently the mutual coherence) such that the problem $(P_{1,\sigma})$ leads to a *stable* solution that accurately approximates the exact sparse solution $u(\boldsymbol{y})$. By stability, we simply mean that the polynomial chaos approximation of $u(\boldsymbol{y})$ with coefficients obtained from $\ell_1$-minimization problem $(P_{1,\sigma})$ does not blow up in the presence of PC expansion truncation error $\sigma_{p,N}$.

**Theorem 2. (Stability of $(P_{1,\delta})$).** *Assume that a vector of PC coefficients $\boldsymbol{c}$ satisfy the sparsity condition $\|\boldsymbol{c}\|_0 < (1+1/\mu(\boldsymbol{\Psi}))/4$ and that $\|\boldsymbol{\Psi c} - \boldsymbol{u}\|_2 \leq \delta_{p,N}$ as stated in (2.27). Then the solution $u_{1,\delta}$ of $(P_{1,\delta})$ must obey*

$$\mathbb{E}\left[(u - u_{1,\delta})^2\right] \leq \frac{4\delta_{p,N}^2}{1 - \mu(\boldsymbol{\Psi})(4\|\boldsymbol{c}\|_0 - 1)} \tag{2.34}$$

*where $u_{1,\delta}(\boldsymbol{y})$ is the PC expansion of $u(\boldsymbol{y})$ with coefficient $\boldsymbol{c}_{1,\delta}$ computed from $\ell_1$-minimization problem $(P_{1,\delta})$.*

*Proof.* Due to orthonormality of $\{\psi_{\boldsymbol{\alpha}}(\boldsymbol{y})\}$ we have $\mathbb{E}\left[(u - u_{1,\delta})^2\right] = \|\boldsymbol{c} - \boldsymbol{c}_{1,\delta}\|_2^2$. But $\|\boldsymbol{c} - \boldsymbol{c}_{1,\delta}\|_2^2 \leq \frac{4\delta_{p,N}^2}{1-\mu(\boldsymbol{\Psi})(4\|\boldsymbol{c}\|_0-1)}$ from Donoho *et al.* (2006).

**Remark 4.** In fact, under the condition of Theorem 2, the error in reconstruction of exact PC coefficients $\boldsymbol{c}$ grows proportional to the error due to truncation of PC expansion of solution samples.

In the next section, we explore some aspects of the proposed scheme through its application to a 1-$D$ (in space) elliptic stochastic PDE with random diffusion coefficient.

## 3. Results

Consider the problem

$$\begin{aligned} -\frac{d}{dx}\left(a(x,\omega)\frac{du(x,\omega)}{dx}\right) &= 1 \quad x \in (0,1), \\ u(0,\omega) = u(1,\omega) &= 0, \end{aligned} \tag{3.1}$$

where the diffusion coefficient $a(x,\omega)$ is given by the expansion

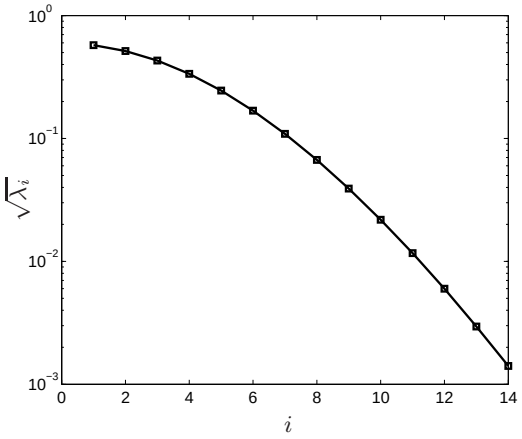



FIGURE 1. Decay of eigenvalues of the covariance kernel of $a(x, \omega)$

$$a(x, \omega) = \bar{a} + \sigma_a \sum_{i=1}^{d} \sqrt{\lambda_i} \phi_i(x) y_i(\omega). \tag{3.2}$$

Here, $\{\lambda_i\}_{i=1}^d$ and $\{\phi_i(x)\}_{i=1}^d$ are, respectively, $d = 14$ largest eigenvalues (shown in Fig. 1) and the corresponding eigenfunctions of the Gaussian covariance kernel

$$C_{aa}(x_1, x_2) = \exp\left[-25(x_1 - x_2)^2\right]. \tag{3.3}$$

The random variables $\{y_i(\omega)\}_{i=1}^d$ are assumed to be independent and uniformly distributed $U[-1, 1]$. We further assume that $\bar{a} = 0.1$ and $\sigma_a = 0.03$ to ensure that all realization of $a(x, \omega)$ is positive on $(0, 1)$.

We solve the problem (3.1) using Lain-Hypercube Monte Carlo simulation, the isotropic sparse grid stochastic collocation with Clenshaw-Curtis abscissas, and the $\ell_1$-minimization technique $(P_{1,\sigma})$. We compare errors in the mean and standard deviation of the solution at $x = 0.5$ using the above methods. A reference solution is computed using the level 8 stochastic collocation for which the approximation errors are negligible in our comparisons.

The 3rd-order, $p = 3$, Legendre PC expansion of $u(0.5, \boldsymbol{y})$ results in a mean-squares error of $\mathcal{O}(10^{-7})$. Among $P = 681$ terms in the full PC expansion, only $S = 21$ terms corresponding to 21 largest coefficients are sufficient to approximate the solution with a mean-squares error of $\mathcal{O}(10^{-6})$. The objective is then to verify the accuracy of the proposed $\ell_1$-minimization $(P_{1,\sigma})$ for approximating these coefficients. Hence, we compute the Legendre PC coefficients $\boldsymbol{c}$ using $N = \{29, 80, 120, 140, 180, 240, 300, 360, 421\}$ number of samples to demonstrate the convergence of the algorithm. The sampling is implemented in a nested fashion: Only additional samples are performed when a larger sample size is needed. The smallest and the largest sample sizes $N = 29$ and $N = 421$ correspond to the level $l = 1$ and level $l = 2$ of isotropic stochastic collocation respectively.

Figure 2 compares the exact PC coefficients with those obtained from the proposed scheme using three different sample sizes. The largest $S = 21$ coefficients distinguished by the dashed line in Fig. 2 correspond to PC basis functions of degree $p = 1$, $p = 3$, and $p = 3$.

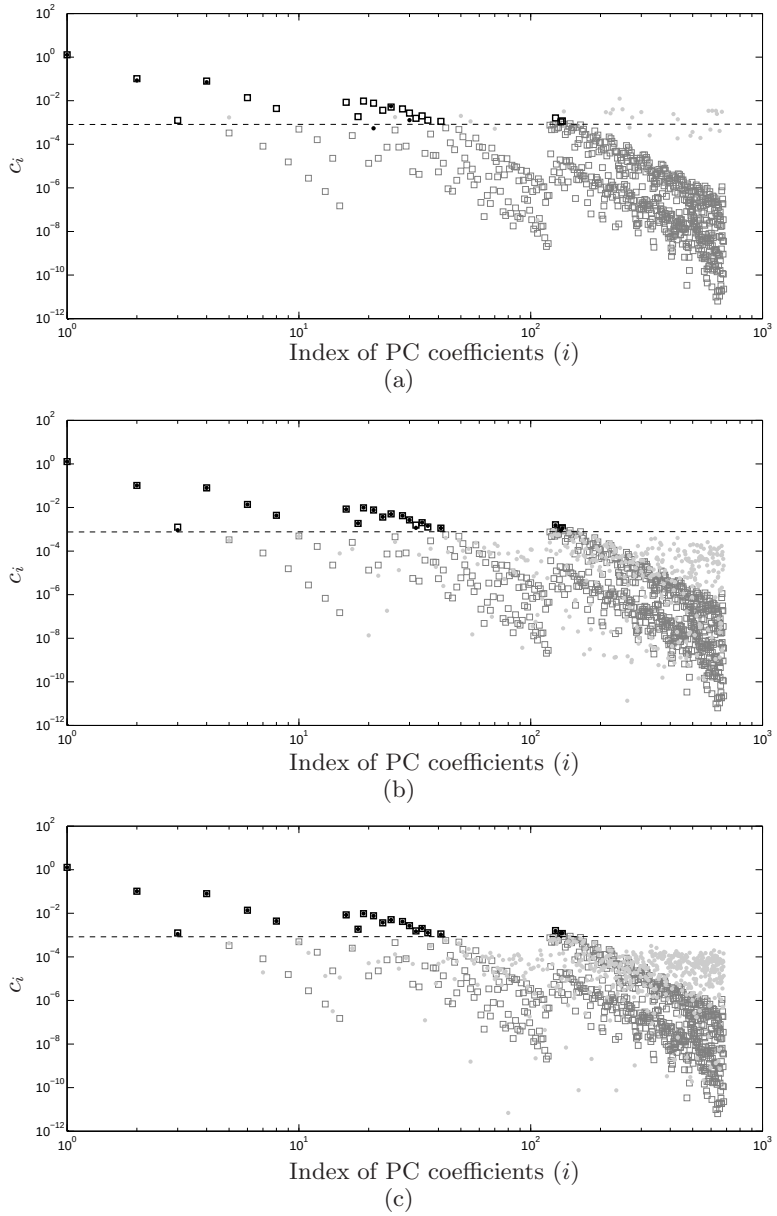
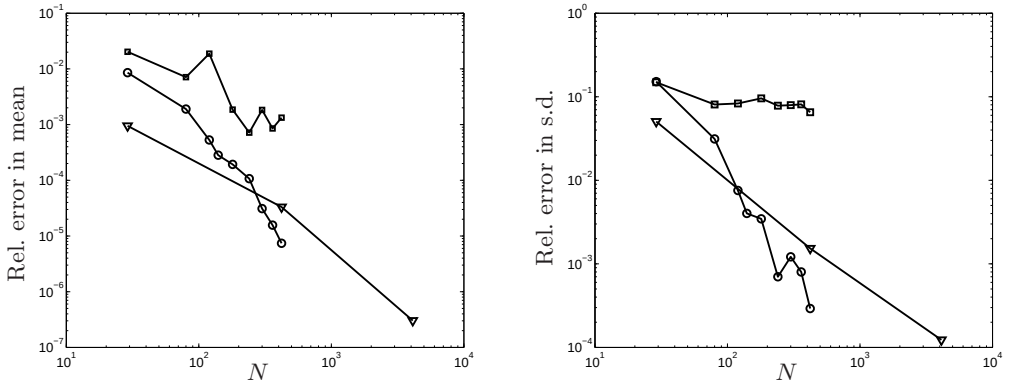



FIGURE 2. Approximation of polynomial chaos (PC) coefficients $c_i$ of $u(0.5, \boldsymbol{y})$ using $\ell_1$-minimization. (a) $N = 29$, (b) $N = 240$, (c) $N = 421$. ('Exact' coefficients computed from level 8 stochastic collocation $\square$ ; $\ell_1$-minimization $\bullet$).

The convergence of the mean and standard deviation of $u(0.5, \boldsymbol{y})$ with respect to different sample sizes $N$ is illustrated in Fig. 3. To make a meaningful comparison, for each $N$, the samples used to compute the solution statistics by the Latin-Hypercube Monte Carlo and the $\ell_1$-minimization are identical. In this sense, the $\ell_1$-minimization can be seen as only a post-processing step in the Monte Carlo simulation. Despite the stochastic collocation approach where the refinement of the approximation requires a certain num-




(a) (b)

FIGURE 3. Relative error in estimation of solution mean and standard deviation at $x = 0.5$ using Lain-Hypercube Monte Carlo simulation, isotropic sparse grid stochastic collocation with Clenshaw-Curtis abscissas, and the $\ell_1$-minimization. (Monte Carlo simulation —□— ; stochastic collocation —▽— ; $\ell_1$-minimization —○— ).

ber of extra samples, the $\ell_1$-minimization can be implemented using arbitrary numbers of additional samples which is an advantage, particularly, when only a limited number of samples can be afforded.

## 4. Conclusion and future directions

The present study proposes a *non-intrusive* approach based on the *compressive sampling* formalism for the approximation of the sparse solution of stochastic PDEs. When sufficiently sparse in the polynomial chaos (PC) basis, compressive sampling is able to recover the solution using a number of its Monte Carlo samples that is significantly smaller than the cardinality of the PC basis. The performance of the proposed technique was explored through its application to a linear elliptic PDE with random diffusion where the sparsity of the solution with respect to the PC basis is guaranteed. The proposed method is not restricted to elliptic PDEs as its underlying applicability assumptions are universal. We plan to explore the efficiency of the proposed method for the analysis of stochastic systems with multi-physics nature.

## Acknowledgments

This work was supported by the United States Department of Energy under the Predictive Science Academic Alliance Program (PSAAP) at Stanford University and California Institute of Technology.

## REFERENCES

BABUŠKA, I., NOBILE, F. & TEMPONE, R. 2007 A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.* **45**(3), 1005–1034.

Babuška, I., Tempone, R. & Zouraris, G. 2004 Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numer. Anal.* **42**(2), 800–825.

Bieri, M. 2009 A sparse composite collocation finite element method for elliptic sPDEs. *Research Report No. 2009-08*, Seminar für Angewandte Mathematik, SAM, Zürich, Switzerland.

Bieri, M., Andreev, R. & Schwab, C. 2009 Sparse tensor discretization of elliptic sPDEs. *Research Report No. 2009-07*, Seminar für Angewandte Mathematik, SAM, Zürich, Switzerland.

Bieri, M. & Schwab, C. 2009 Sparse high order FEM for elliptic sPDEs. *Computer Methods in Applied Mechanics and Engineering* **198**, 1149–1170.

Bruckstein, A.M., Donoho, D.L. & Elad, M. 2009 From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review.* **51**(1), 34–81.

Candes, E.J. & Romberg, J. 2006 Quantitative robust uncertainty principles and optimally sparse decompositions. *Foundations of Comput. Math.* **6**(2), 227–254.

Candes, E.J., Romberg, J. & Tao, T. 2006 Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory* **52**(2), 489–509.

Candes, E.J. & Tao, T. 2006 Near optimal signal recovery from random projections: Universal encoding strategies?. *IEEE Transactions on information theory.* **52**(12), 5406–5425.

Chen, S.S., Donoho, D.L. & Saunders, M.A. 2001 Atomic decomposition by basis pursuit. *SIAM Rev.* **43**(1), 129–159.

Cohen, A., Dahmen, W. & DeVore, R. 2009 Compressed sensing and best $k-$term approximation. *J. Amer. Math. Soc.* **22**, 211–231.

Deb, M., Babuška, I. & Oden, J. 2007 Solution of stochastic partial differential equations using Galerkin finite element techniques. *Computer Methods in Applied Mechanics and Engineering.* **190**(48), 6359–6372.

Donoho, D.L. 2006 Compressed Sensing. *IEEE Transactions on information theory.* **52**(4), 1289–1306.

Donoho, D.L., Elad, M. & Temlyakov, V. 2006 Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on information theory.* **52**, 6–18.

Doostan, A., Ghanem, R. & Red-Horse, J. 2007 Stochastic model reduction for chaos representations. *Computer Methods in Applied Mechanics and Engineering* **196**, 3951–3966.

Doostan, A. & Iaccarino, G. 2008 A least-squares approximation of partial differential equations with high-dimensional random inputs. *Journal of Computational Physics* **228**, 4332–4345.

Ghanem, R. & Spanos, P. 2003 *Stochastic Finite Elements: A Spectral Approach.* Mineola, New York: Dover.

Loeve, M. 1977 *Probability Theory,* $4^{th}$ *edition.* Springer-Verlag, NewYork.

Ma, X. & Zabaras, N. 2009 An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. *Journal of Computational Physics* **228**, 3084–3113.

Mathelin, L. & Hussaini, M. 2003 A stochastic collocation algorithm for uncertainty

analysis. *Technical Report NAS 1.26:212153; NASA/CR-2003-212153*, Langley, Virginia: NASA Langley Research Center.

Nobile, F., Tempone, R. & Webster, C. 2007 An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.* **46**(5), 2411–2442.

Nobile, F., Tempone, R. & Webster, C. 2008 A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.* **46**(5), 2309–2345.

Nouy, A. 2007 A generalized spectral decomposition technique to solve a class of linear stochastic partial differential equations. *Computer Methods in Applied Mechanics and Engineering.* **196** 4521–4537.

Nouy, A. 2008 Generalized spectral decomposition method for solving stochastic finite element equations: Invariant subspace problem and dedicated algorithms. *Computer Methods in Applied Mechanics and Engineering.* **197** 4718–4736.

Tatang, M.A. 1995 Direct incorporation of uncertainty in chemical and environmental engineering systems. *PhD thesis*, Massachusetts Institute of Technology.

Todor, R.A. & Schwab, C. 2007 Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients. *IMA Journal of Numerical Analysis.* **27**(2) 232–261.

Xiu, D. & Hesthaven, J. 2005 High-order collocation methods for differential equations with random inputs. *SIAM J. Sci. Comput.* **27**(3), 1118–1139.

Xiu, D. & Karniadakis, G. 2002 The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.* **24**(2), 619–644.