ORNL REPORT

ORNL/TM-2012/186
Unlimited Release
Printed September 2012

An adaptive wavelet stochastic collocation method for irregular solutions of stochastic partial differential equations

M. Gunzburger, C. G. Webster and G. Zhang

Prepared by
Oak Ridge National Laboratory
One Bethel Valley Road, Oak Ridge, Tennessee 37831

The Oak Ridge National Laboratory is operated by UT-Battelle, LLC,
for the United States Department of Energy under Contract DE-AC05-00OR22725.

OAK RIDGE NATIONAL LABORATORY
MANAGED BY UT-BATTELLE FOR THE DEPARTMENT OF ENERGY

ORNL/TM-2012/186

Computer Science and Mathematics Division

AN ADAPTIVE WAVELET STOCHASTIC COLLOCATION METHOD FOR IRREGULAR SOLUTIONS OF STOCHASTIC PARTIAL DIFFERENTIAL EQUATIONS

M. Gunzburger [*] C. G. Webster [†] G. Zhang [‡]

Date Published: September 2012

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831-6283
managed by
UT-BATTELLE, LLC
for the
U.S. DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

[*]Department of Scientific Computing, 400 Dirac Science Library, Florida State University, Tallahassee, FL 32306–4120 (mgunzburger@fsu.edu)

[†]Computer Science and Mathematics Division, Oak Ridge National Laboratory, One Bethel Valley Road, P.O. Box 2008, MS-6164, Oak Ridge, TN 37831-6164 (webstercg@ornl.gov)

[‡]Computer Science and Mathematics Division, Oak Ridge National Laboratory, One Bethel Valley Road, P.O. Box 2008, MS-6367, Oak Ridge, TN 37831-6164 (zhangg@ornl.gov).

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

A novel multi-dimensional multi-resolution adaptive wavelet stochastic collocation method (AWSCM) for solving partial differential equations with random input data is proposed. The uncertainty in the input data is assumed to depend on a finite number of random variables. In case the dimension of this stochastic domain becomes moderately large, we show that utilizing a hierarchical sparse-grid AWSCM (sg-AWSCM) not only combats the curse of dimensionality but, in contrast to the standard sg-SCMs built from global Lagrange-type interpolating polynomials, maintains fast convergence without requiring sufficiently regular stochastic solutions. Instead, our non-intrusive approach extends the sparse-grid adaptive linear stochastic collocation method (sg-ALSCM) by employing a compactly supported wavelet approximation, with the desirable multi-scale stability of the hierarchical coefficients guaranteed as a result of the wavelet basis having the Riesz property. This property provides an additional lower bound estimate for the wavelet coefficients that are used to guide the adaptive grid refinement, resulting in the sg-AWSCM requiring an optimal (up to a constant) number of deterministic simulations for both smooth and irregular stochastic solutions. Second-generation wavelets constructed from a lifting scheme allow us to preserve the framework of the multi-resolution analysis, compact support, as well as the necessary interpolatory and Riesz property of the hierarchical basis. Several numerical examples are given to demonstrate the optimal convergence of our numerical scheme and to show the increased efficiency when compared to the sg-ALSCM method.

ACKNOWLEDGEMENTS

The first author was supported by the US Air Force Office of Scientific Research under grant number FA9550-11-1-0149. The second author was supported by the US Air Force Office of Scientific Research under grant #1854-V521-12. The second author was also sponsored by the Director's Strategic Hire Funds through the Laboratory Directed Research and Development (LDRD) Program of Oak Ridge National Laboratory (ORNL). The third author was supported by the US Air Force Office of Scientific Research under grant number FA9550-11-1-0149. The third author was also supported by the Advanced Simulation Computing Research (ASCR), Department of Energy, through the Householder Fellowship at ORNL. The ORNL is operated by UT-Battelle, LLC, for the United States Department of Energy under Contract DE-AC05-00OR22725.

1 INTRODUCTION

Many applications in engineering and science are affected by uncertainty in input data, including model coefficients, forcing terms, boundary condition data, media properties, source and interaction terms, as well as geometry. For example, highly heterogeneous materials may have properties that vary over small length scales so that these properties have to be often determined, e.g., by interpolating or extrapolating measurements obtained at a few locations. These types of uncertainties are known as *epistemic* because they are related to incomplete knowledge. In other situations, referred to as *aleatoric*, uncertainty is due to intrinsic variability in the system, e.g., fluctuations in turbulent flow fields. In practice, it is necessary to quantify both types of uncertainties, a process which is naturally referred to as *uncertainty quantification* (UQ).

The presence of random input uncertainties can be incorporated into a system of partial differential equations (PDEs) by formulating the governing equations as PDEs with random inputs. In practice, such PDEs may depend on a set of distinct random parameters with the uncertainties represented by a given joint probability distribution. In other situations, the input data varies randomly from one point of the physical domain to another and/or from one time instant to another; in these cases, uncertainties in the inputs are instead described in terms of *random fields* that can be expressed as an expansion containing an infinite number of random variables. For example, for correlated random fields, one has Karhunen-Loève (KL) expansions [30,31], Fourier-Karhunen-Loève expansions [29], or expansions in terms of global orthogonal polynomials [21, 50, 52]. However, in a large number of applications, it is reasonable to limit the analysis to just a finite number of random variables, either because the problem input itself can be described in that way (e.g., the random parameter case) or because the input random field can be approximated by truncating an infinite expansion [19] (e.g., the correlated random field case).

Currently, there are several numerical methods available for solving PDEs with random input data. Monte Carlo methods (MCMs) (see, e.g., [17]) are the classical and most popular approaches for approximating expected values and other statistical moments of quantities of interest that depend on the solution of PDEs with random inputs. MCMs are very flexible and trivial to implement and parallelize using existing deterministic PDE solvers, but they feature very slow convergence because they do not exploit any regularity the solution may possess with respect to the input stochastic parameters. On the other hand, the convergence rates of MCMs have mild dependence on the number of random variables so that for problems involving a large number of random parameters, MCMs remain the method of choice.

Several numerical approaches have been proposed that often feature much faster convergence rates. These include stochastic Galerkin methods (SGMs) [1, 3, 21, 35] and stochastic collocation methods (SCMs) [2, 34, 38, 39, 51]. The two approaches transform the original stochastic problem into a deterministic one with a large number of parameters and differ in the choice of the polynomial bases used and the resulting approximating spaces. To achieve increased rates of convergence relative to MCMs, both approaches are typically based on global polynomial approximations that take advantage of smooth behavior of the solution in the multi-dimensional parameter space. SGMs are based on orthogonal polynomials which leads to a coupling of the probabilistic and space/time degrees of freedom; for this reason, SGMs are referred to as being *intrusive*. On the

other hand, SCMs are based on interpolatory polynomials so that, when implemented, they result in ensemble-based *non-intrusive* approaches for which the probabilistic and space/time degrees of freedom are uncoupled.

We emphasize that the better convergence behavior of SGMs and SCMs relative to MCMs requires high regularity with respect to the random variables. However, often in scientific and engineering problems there are are irregular dependences, e.g., steep gradients, sharp transitions, bifurcations, or finite jump discontinuities, of a quantity of interest (QoI) with respect to the random variables. In such cases, global polynomial-based approximations such as SGMs and SCMs seriously deteriorate to the point that they converge very slowly or may even fail to converge. Indeed, for such applications, the use SGMs and SCMs often result in no improvements over MCMs. As a result, one turns to local approximation methods. To be effective, such approximations have to be implemented using refinement strategies that focus on regions of irregular behavior; otherwise, there would be an explosion in the required computational effort as the number of random variables increases, a phenomenon commonly referred to as *the curse of dimensionality*.

Not surprisingly, there have been many proposed methods that attempt to control the curse, i.e., to put off its inevitable fatal effect to higher dimensions. Several techniques involve domain decomposition approaches using h-type finite elements basis functions, similar to those constructed in the physical spatial domain. A *multi-element* approach utilized in [18] decomposes each parameter dimension into sub-domains and then uses tensor products to reconstruct the entire parameter space. This method has successfully been applied to moderate dimension problems, but the tensor-product decomposition inevitably re-introduces the curse of dimensionality. Similarly, [26, 27] presents a tensor product-based multi-resolution approximation based on a Galerkin projection onto a Wiener-Haar basis. This approach provides significant improvements over global orthogonal approximation approaches. However, in terms of robustness, dimension scaling is not possible due to the resulting dense coupled system and the lack of any rigorous criteria for triggering refinement.

It is recognized that any refinement strategy employed must be guided by an accurate estimation of both local and global errors. In [32, 33], an adaptive sparse-grid stochastic collocation strategy is applied that uses piecewise multi-linear hierarchical basis functions developed in [20, 22, 25]. This approach utilizes the hierarchical surplus as an error indicator to automatically detect the regions of importance (e.g., discontinuities) in stochastic parameter space and to adaptively refine the collocation points in this region. The adaptation process is continued until a prescribed global error tolerance is achieved. This goal, however, might be reached by using more points than necessary due to the instability of the multi-scale basis used; see §3.3 for a complete description of the properties of such multi-scale sparse grid approximations using hierarchical subspace splitting and see §4 for the additional properties required to construct an optimal multi-dimensional multi-resolution approximation.

In §4, we propose an adaptive wavelet stochastic collocation method that possesses the additional properties. The intent of our approach is to combat the curse of dimensionality while maintaining the increased convergence rates of standard SCM approaches by utilizing optimal compactly supported wavelet basis functions. The construction principles of such functions are highly developed; see, e.g., [6, 9, 12, 13, 23] and the references therein. Such bases are in ubiq-

uitous use in signal processing and other applications. They have also been *rigorously* shown to result in optimal approximations of PDEs (see, e.g., [7, 8, 10, 14, 15, 36, 42, 43]) and of PDE constrained optimal control problems (see, e.g., [11, 23]), when compared with traditional finite element approximations. In this paper, due to their general applicability to arbitrary domains, we consider second-generation wavelets constructed from a lifting scheme [46–48]. Moreover, in addition to maintaining compact support and the interpolatory properties of nodal bases, the beauty of the second-generation wavelets is that they also form a Riesz basis, a property that guarantees the stability of the hierarchical basis and allows one to construct an optimal multi-resolution approximation.

The outline of the paper is as follows. In §2, we introduce the mathematical description of a general stochastic initial-boundary problem and the main notation used throughout the paper. In §3, we briefly recall the stochastic collocation method and adaptive strategies using both global as well linear hierarchical polynomials. In §4, we propose our novel adaptive wavelet stochastic collocation method and the properties of the second-generation wavelets we employ. In §5, several numerical examples are given to demonstrate the effectiveness and efficiency of our method compared with classic approaches.

2 PROBLEM SETTING

We follow the notation in [2, 38, 39] and begin by letting D denote a bounded domain in $\mathbb{R}^d$, $d = 1, 2, 3$, and $(\Omega, \mathcal{F}, P)$ denote a complete probability space. Here, Ω denotes the set of outcomes, $\mathcal{F} \subset 2^\Omega$ the σ-algebra of events, and $P : \mathcal{F} \to [0, 1]$ a probability measure. We are interested the following stochastic initial-boundary value problem: find $u \colon \Omega \times \overline{D} \times [0, T] \to \mathbb{R}^m$ such that P-almost everywhere in Ω

$$\mathcal{L}(a)(u) = f \quad \text{in } D \times [0, T] \tag{2.1}$$

subject to the boundary and initial conditions

$$\begin{aligned} \mathcal{B}(b)(u) &= g \quad \text{on } \partial D \times [0, T] \\ u &= u_0 \quad \text{on } D \times \{t = 0\}. \end{aligned} \tag{2.2}$$

Here, $\mathcal{L}$ denotes a differential operator (linear or non-linear) depending on the coefficient(s) $a(\omega, x, t)$ with $(\omega, x, t) \in \Omega \times D \times [0, T]$; $\mathcal{B}$ denotes a boundary operator depending on the coefficient(s) $b(\omega, x, t)$ with $(\omega, x, t) \in \Omega \times \partial D \times [0, T]$. Similarly, the right-hand sides $f(\omega, x, t)$, $g(\omega, x, t)$, and $u_0(\omega, x)$ can be assumed to be random fields as well. Note that, in general, a, b, f, g, and u_0 belong to different probability spaces, but for economy of notation, we simply denote the stochastic dependences of these random data as if all belong to the same probability space. We denote by $W(D)$ a Banach space and assume the underlying stochastic input data are chosen so that the corresponding stochastic system (2.1)–(2.2) is well-posed so that it has an unique solution $u(\omega, x, t) \in L^2_P(\Omega) \otimes L^2(W(D); 0, T)$, where the space

$$\begin{aligned} &L^2_P(\Omega) \otimes L^2(W(D); 0, T) \\ &\quad := \left\{ u \ : \ \Omega \times \overline{D} \times [0, T] \to \mathbb{R}^m \ \middle| \ \int_0^T \int_\Omega \|u\|^2_{W(D)} \, dP(\omega) dt < +\infty \right\} \end{aligned} \tag{2.3}$$

consists of Banach-space valued functions that have finite second moments. Finally, we note that in this setting the solution u can either be a scalar or vector-valued function depending on the system of interest.

An example problem posed in this setting is given as follows.

Example 2.1 [Linear parabolic PDE with random inputs] *Consider the initial-boundary value problem* [53]*: find a random field* $u : \Omega \times \overline{D} \times [0, T] \to \mathbb{R}$ *such that* P*-almost surely*

$$\begin{aligned} \partial_t u(\omega, x, t) - \nabla \cdot [a(\omega, x) \nabla u(\omega, x, t)] &= f(\omega, x, t) \quad \textit{in} \quad \Omega \times D \times [0, T] \\ u(\omega, x, t) &= 0 \quad \textit{on} \quad \Omega \times \partial D \times [0, T] \\ u(\omega, x, 0) &= u_0(\omega, x) \quad \textit{on} \quad \Omega \times D, \end{aligned} \tag{2.4}$$

where ∇ *denotes the gradient operator with respect to the spatial variable* $x \in D$*. To guarantee the well-posedness of the solution of* (2.4) *in* $L^2_P(\Omega) \otimes L^2(H^1(D); 0, T)$*, one assumes that almost surely the coefficient* $a(x, \omega)$ *is positive and uniformly bounded, i.e.,*

$$P\big(\omega \in \Omega : a_{\min} \le a(\omega, x) \le a_{\max} \, \forall \, x \in \overline{D}\big) = 1 \ \textit{ with } \ a_{\min}, a_{\max} \in (0, \infty) \tag{2.5}$$

and that the right-hand side satisfies

$$\int_0^T \int_D \mathbb{E}[f^2] dxdt := \int_0^T \int_D \int_\Omega f^2(\omega, x, t)\, dP(\omega) dxdt < +\infty \qquad P\text{-a.e. in } \Omega.$$

2.1 FINITE DIMENSIONAL NOISE

In many applications, the source of randomness can be approximated using just a finite number of uncorrelated, or even independent, random variables. As such, similar to [2, 38, 39], we make the following assumptions regarding the stochastic input data, i.e., the random coefficients a and b in $\mathcal{L}$ and $\mathcal{B}$ and the right-hand sides f, g, and u_0 in (2.1)–(2.2).

A_1) *The stochastic input coefficient a satisfies (2.5) and the other stochastic input data are bounded from above and below with probability 1, e.g., for the right-hand side $f(\omega, x, t)$, there exists $f_{\min} > -\infty$ and $f_{\max} < \infty$ such that*

$$P\big(\omega \in \Omega \ : \ f_{\min} \le f(\omega, x, t) \le f_{\max} \quad \forall x \in \overline{D},\ \forall t \in [0, T]\big) = 1 \tag{2.6}$$

and similarly for all remaining inputs.

A_2) *The stochastic input data have the form*

$$d_0(\omega, x, t) + \sum_{n=1}^{N} y_n(\omega) d_n(x, t), \tag{2.7}$$

where $N \in \mathbb{N}_+$ and $\mathbf{y} = [y_1(\omega), \ldots, y_N(\omega)] : \Omega \to \mathbb{R}^N$ is a real-valued vector of independent random variables.

In many applications, the stochastic input data may have a simple piecewise random representation whereas, in other applications, the coefficients a and b in (2.1) and the right-hand sides f, g, and u_0 in (2.2) may have spatial variation that can be modeled as a correlated random field, making them amenable to description by a Karhunen-Loève (KL) expansion [30, 31]. In practice, one has to truncate such expansions so that they are of the form (2.7), with the number N of terms kept depending on the degree of correlation and the desired accuracy of the simulation. Examples of both types of noise, each satisfying assumptions A_1 and A_2, are given next.

Example 2.2 [Piecewise constant random fields] *We assume the spatial domain D is the union of non-overlapping subdomains D_j, $j = 1, \ldots, J$, and the time interval $(0, T)$ is the union of disjoint subintervals (T_{k-1}, T_k), $k = 1, \ldots, K$. Then, we consider stochastic input data that is piecewise constant and random on each space-time subdomain $D_j \times (T_{k-1}, T_k)$, i.e.,*

$$d_0(x, t) = \sigma_0 \qquad \text{and} \qquad d_n(x, t) = \sigma_n 1_{D_j \times (T_{k-1}, T_k)}(x, t), \quad n = j + (k-1)K,$$

where σ_n, $n = 0, \ldots, N$, denote constants, $1_{D_j \times (T_{k-1}, T_k)}$ denotes the indicator function of the set $D_j \times (T_{k-1}, T_k) \subset D \times [0, T]$, and the random variables $y_n(\omega)$ are bounded and independent.

Once the bounded sample space Ω is defined, the constants $a_{\min}$, $a_{\max}$, $f_{\min}$, $f_{\max}$, etc. in the constraints (2.5)–(2.6) are easily deduced. Note that (2.5) requires restrictions on the constants σ_n and the bounds on the random variables $y_n(\omega)$ corresponding to the coefficient a; in practice, such restrictions would be deduced from the physics of the problem.

Example 2.3 [Karhunen-Loève expansion] *Any second-order correlated random field $c(\omega, x, t)$ with continuous covariance function $Cov[c](\widetilde{x}_1, \widetilde{x}_2)$, where $\widetilde{x}_1 = (x_1, t_1)$ and $\widetilde{x}_2 = (x_2, t_2)$ are space-time coordinates, can be represented as an infinite sum of random variables by means of, e.g., a KL expansion. For $\widetilde{x} = (x, t)$, we define the operator $F : L^2(D) \times L^2(0,T) \to L^2(D) \times L^2(0,T)$ by*

$$Fv(\widetilde{x}) := \int_0^T \int_D Cov[c](\widetilde{x}_1, \widetilde{x}) v(\widetilde{x}_1)\, d\widetilde{x}_1 \quad \forall\, v \in L^2(D) \times L^2(0,T). \tag{2.8}$$

Because of the symmetry and positivity properties of covariance functions, the operator F has real, non-negative eigenvalues $\{\lambda_n\}_{n=1}^{\infty}$ that may be arranged in non-increasing order and corresponding real orthonormal eigenfunctions $\{c_n(\widetilde{x})\}_{n=1}^{\infty}$. For simplicity of the exposition, we assume that the eigenvalues are positive. Furthermore, if we define mutually uncorrelated real random variables by

$$y_n(\omega) := \frac{1}{\sqrt{\lambda_n}} \int_0^T \int_D \left(c(\omega, \widetilde{x}) - \mathbb{E}[c](\widetilde{x})\right) c_n(\widetilde{x})\, d\widetilde{x}, \qquad n = 1, 2, \ldots$$

with zero mean and variance $Var[y_n] = \sqrt{\lambda_n}$, then $c(\omega, x, t)$ can be represented by the truncated N-term KL expansion satisfying assumption A_2 with $d_0(x,t) = \mathbb{E}[c](x,t)$ and $d_n(x,t) = c_n(x,t)$ for $n = 1, \ldots, N$. Finally, note that if the process is Gaussian, then the random variables $\{y_n\}_{n=1}^{\infty}$ are standard identically independent distributed random variables.

In what follows, we denote by $\Gamma_n \equiv y_n(\Omega) \subset \mathbb{R}$ the image of the random variable y_n, then set $\Gamma := \prod_{n=1}^{N} \Gamma_n$, where $N \in \mathbb{N}_+$, and assume that the components of the real-valued random vector $\mathbf{y} = [y_1(\omega), \ldots, y_N(\omega)] : \Omega \to \mathbb{R}^N$ have a joint probability density function (PDF)

$$\rho(\mathbf{y}) : \Gamma \to \mathbb{R}_+ \quad \text{with} \quad \rho(\mathbf{y}) \in L^{\infty}(\Gamma)$$

such that $\rho(\mathbf{y}) = \prod_{n=1}^{N} \rho_n(y_n)$. We note that, given assumption A_1, the image set Γ is a bounded hypercube in $\mathbb{R}^N$. Moreover, from assumption A_2, we have that the solution u to (2.1)–(2.2) depends on a realization $\omega \in \Omega$ through an instantiation of the random vector $\mathbf{y} \in \Gamma$. Therefore, the probability space $(\Omega, \mathcal{F}, P)$ is equivalent to $(\Gamma, \mathcal{B}(\Gamma), \rho(\mathbf{y})d\mathbf{y})$, where $\mathcal{B}(\Gamma)$ is the Borel σ-algebra on Γ and $\rho(\mathbf{y})d\mathbf{y}$ is the finite measure of the random vector $\mathbf{y}$.

We are now in position to restate the random input data in terms of $\mathbf{y}$ as follows:

$$\begin{aligned} a(\omega, x, t) &= a(\mathbf{y}, x, t), \quad f(\omega, x, t) = f(\mathbf{y}, x, t) \quad \text{for} \quad (\mathbf{y}, x, t) \in \Gamma \times D \times [0, T] \\ b(\omega, x, t) &= b(\mathbf{y}, x, t), \quad g(\omega, x, t) = g(\mathbf{y}, x, t) \quad \text{for} \quad (\mathbf{y}, x, t) \in \Gamma \times \partial D \times [0, T] \\ u_0(\omega, x) &= u_0(\mathbf{y}, x) \quad \text{for} \quad (\mathbf{y}, x) \in \Gamma \times D. \end{aligned} \tag{2.9}$$

As a result, the problem (2.1)–(2.2) can be restated as follows: find a random function $u(\mathbf{y}, x, t) : \Gamma \times \overline{D} \times [0, T] \to \mathbb{R}^m$ such that ρ-almost everywhere $\mathbf{y} \in \Gamma$, we have that

$$\mathcal{L}(a(\mathbf{y}, x, t))(u) = f(\mathbf{y}, x, t) \quad \text{in } D \times [0, T] \tag{2.10}$$

subject to the boundary and initial conditions

$$\begin{aligned} \mathcal{B}(b(\mathbf{y}, x, t))(u) &= g(\mathbf{y}, x, t) \quad \text{on } \partial D \times [0, T] \\ u &= u_0(\mathbf{y}, x) \quad \text{on } D \times \{t = 0\}. \end{aligned} \tag{2.11}$$

Then, by the Doob-Dynkin lemma [40], the unique solution u of (2.1) and (2.2) can also be characterized by the same random vector $\mathbf{y}$, i.e.,

$$u(\omega, x, t) = u(y_1(\omega), \ldots, y_N(\omega), x, t) \in L^2_\rho(\Gamma) \otimes L^2(W(D); 0, T), \tag{2.12}$$

where $L^2_\rho(\Gamma)$ is the space of square integrable functions of Γ with respect to the measure $\rho(\mathbf{y})d\mathbf{y}$. As indicated in (2.12), the solution $u(\mathbf{y}, x, t)$ belongs to the function space $L^2_\rho(\Gamma) \otimes L^2(W(D); 0, T)$ that is defined by

$$\begin{aligned} &L^2_\rho(\Gamma) \otimes L^2(W(D); 0, T) \\ &\quad := \left\{ u \,:\, \Gamma \times \overline{D} \times [0, T] \to \mathbb{R}^m \;\middle|\; \int_\Gamma \int_0^T \|u\|^2_{W(D)} \rho(\mathbf{y}) dt d\mathbf{y} < \infty \right\}. \end{aligned} \tag{2.13}$$

We once again note that, in general, each appearance of $\mathbf{y}$ in (2.10)–(2.11) can be a different vector of random variables each belonging to a different probability space and that in the end, the solution u depends on all the different $\mathbf{y}$'s which collectively belong to the product space of the individual probability spaces. However, again to economize notation, we do not explicitly differentiate between the different vectors of random variables.

Thus far we have turned the possibly infinite-dimensional stochastic initial-boundary value problem given by (2.1)–(2.2) into a finite-dimensional parametric problem (2.10)–(2.11). Without loss of generality, we will assume the support of the random variables y_n is $\Gamma_n = [0, 1]$ for $n = 1, \ldots, N$ and therefore the bounded stochastic (or parameter) space is the N-dimensional unit hypercube $\Gamma = [0, 1]^N$. At this point, we can apply any stochastic approximation technique, e.g., spectral-Galerkin, locally adaptive, etc. However, the focus of our work involves *non-intrusive* approximations (such as Monte Carlo sampling or stochastic collocation methods) in probability space for which, for any realization $\mathbf{y}(\omega_k)$ of the random parameters, solutions can be constructed using standard deterministic approximation techniques in space-time, e.g., finite difference methods, finite element methods, finite volume methods, etc. for spatial discretization and backward Euler or Crank-Nicolson schemes for temporal discretization [37, 53].

3 ADAPTIVE STOCHASTIC COLLOCATION METHODS

To provide context and background for the new method we present in §4, in this section we discuss, in general terms, adaptive stochastic collocation methods (SCMs). These approximations are computed via Lagrange interpolation of the random parameter dependence of solutions of (2.10)-(2.11), described in §3.1, constructed from either globally or locally supported basis functions, described in §3.2 and §3.3 respectively. In §3.3, we discuss in somewhat more detail the special case of hierarchical piecewise polynomial basis functions, leading to the hierarchical, locally adaptive, piecewise linear approximations. The latter is the closest precursor to our new method and, naturally, we use it for comparison purposes.

We note that the use of polynomials having the property that the interpolation matrix is diagonal, i.e. the "delta property" (see Remark 3.1), leads to approximations that some authors refer to as *stochastic collocation methods* (SCMs). Others use that terminology to refer to any method for which the parameter and spatial/temporal degrees of freedom uncouple; with this view, which is the one we adopt, all methods discussed below and in this section and in §4 would be referred to as being SCMs.

3.1 LAGRANGE INTERPOLATION IN THE PROBABILISTIC DOMAIN

The goal is to construct a numerical approximation of the solution of (2.10)–(2.11) in a finite-dimensional subspace $\mathcal{P}(\Gamma) \otimes L^2(W_h(D); 0, T)$. Here, $W_h(D) \subset W(D)$ is a standard finite element space of dimension $\dim(W_h) = M_h$, used for spatial discretization and $\mathcal{P}(\Gamma) \subset L^2_\rho(\Gamma)$ is a finite-dimensional space of dimension $\dim(\mathcal{P}(\Gamma)) = M$, used for approximation in parameter space. Of course, a temporal discretization, usually via a finite difference method, is implied as well. Interpolatory approximations in parameter space start with the selection of a set of distinct points $\{\mathbf{y}_k\}_{k=1}^M \in \Gamma$, in parameter space and a set of basis functions[1] $\{\psi_k(\mathbf{y})\}_{k=1}^M \in \mathcal{P}(\Gamma)$. Then, we seek an approximation $u_{M_h,M} \in \mathcal{P}(\Gamma) \otimes L^2(W(D); 0, T)$ of the solution u of the problem (2.10)-(2.11) of the form

$$u_{M_h,M}(\mathbf{y}, x, t) = \sum_{k=1}^{M} c_k(x, t)\psi_k(\mathbf{y}). \tag{3.1}$$

The Lagrange interpolant is defined by first obtaining M realizations $u_{M_h}(\mathbf{y}_k, x, t)$ of the finite element approximation[2] of the solution $u(\mathbf{y}_k, x, t)$ of the problem (2.10)-(2.11), i.e., one solves for the finite element approximation for each of the interpolation points in the set $\{\mathbf{y}_k\}_{k=1}^M$. Then, the coefficient functions $\{c_k(x, t)\}_{k=1}^M$ are determined by imposing the interpolation conditions

$$\sum_{\ell=1}^{M} c_\ell(x, t)\psi_\ell(\mathbf{y_k}) = u_{M_h}(\mathbf{y}_k, x, t) \qquad \text{for } k = 1, \ldots, M. \tag{3.2}$$

[1] In general, the number of points and number of basis functions do not have to be the same, e.g., for Hermite interpolation. However, because here we only consider Lagrange interpolation, we let M denote both the number of points and the dimension of the basis.

[2] Extensions to other methods, e.g., finite difference, finite volume, spectral or h-p, etc. are straightforward.

Thus, the coefficient functions $\{c_\ell(x,t)\}_{\ell=1}^M$ are each a linear combination of the data functions $\{u_{M_h}(\mathbf{y}_k,x,t)\}_{k=1}^M$; the specific linear combinations are determined in the usual manner from the entries of the inverse of the $M \times M$ interpolation matrix $\mathbf{K}$ having entries $K_{k\ell} = \psi_\ell(\mathbf{y}_k)$, $k,\ell = 1,\ldots,M$. The sparsity of $\mathbf{K}$ heavily depends on the choice of basis; that choice could result in matrices that range from fully dense to diagonal.

A main attraction of interpolatory approximations of parameter dependences is that it effects a complete decoupling of the spatial/temporal and probabilistic degrees of freedom. Clearly, once the interpolation points $\{\mathbf{y}_k\}_{k=1}^M$ are chosen, one can solve M deterministic problems (i.e., the spatial/temporal discretization of (2.10)-(2.11)), one for each parameter point $\mathbf{y}_k$, with total disregard to what basis $\{\psi_k(\mathbf{y})\}_{k=1}^M$ one choose to use. Then, the coefficients $\{c_k(x,t)\}_{k=1}^M$ defining the approximation (3.1) of the solution of (2.10)-(2.11) are found from the interpolation matrix $\mathbf{K}$ as discussed above; its only in this last step that the choice of basis enters into the picture. Note that this decoupling property makes the implementation of Lagrange interpolatory approximations of parameter dependences as trivial as it is for Monte Carlo sampling. However, if that dependence is smooth, because of the higher accuracy of, e.g., polynomial approximations in the space $\mathcal{P}(\Gamma)$, interpolatory approximations require substantially fewer sampling points to achieve a desired tolerance.

Remark 3.1 (The "delta property") *Given a set of interpolation points, to complete the setup of a Lagrange interpolation problem, one has to then choose a basis. The simplest and most popular choice are the Lagrange polynomials, i.e., polynomials that have the "delta property" $\psi_\ell(\mathbf{y}_k) = \delta_{k\ell}$, where $\delta_{k\ell}$ denotes the Kroeneker delta. In this case, the interpolating conditions (3.2) reduce to $c_k(x,t) = u_h(\mathbf{y}_k,x,t)$ for $k = 1,\ldots,M$, i.e., the interpolation matrix $\mathbf{K}$ is simply the $M \times M$ identity matrix. In this sense, the use of Lagrange polynomial bases can be viewed as resulting in pure sampling methods, much the same as Monte Carlo methods, but instead of randomly sampling in the parameter space Γ, the sample points are deterministically structured as, e.g., tensor product or sparse grid points.*

3.2 ADAPTIVE GLOBAL SPARSE-GRID LAGRANGE INTERPOLATION

When the solution is analytic with respect to the noise parameterization [2,38,39], the most widely used approach to for constructing approximations of the form (3.1) involves the construction of a global Lagrange interpolant, by replacing the polynomial space $\mathcal{P}(\Gamma)$ by $\mathcal{P}_p(\Gamma)$, defined as the span of product polynomials, i.e.,

$$\mathcal{P}_p(\Gamma) = \text{span}\left\{ \prod_{n=1}^N y_n^{p_n} \quad \text{with } \mathbf{p} = (p_1,\ldots,p_N) \in \mathcal{J}(p) \right\},$$

where the index set $\mathcal{J}(p)$ determines the type of polynomial space used. Thus, the dimension M of $\mathcal{P}_p(\Gamma)$ is the cardinalty of the index set $\mathcal{J}(p)$. Two obvious choices are tensor product spaces of one-dimenional polynomials of degree p for which $\mathcal{J}(p) = \{\mathbf{p} \in \mathbb{N}^N : \max_{1\le n\le N} p_n \le p\}$ and total degree p spaces for which $\mathcal{J}(p) = \{\mathbf{p} \in \mathbb{N}^N : \sum_{n=1}^N p_n \le p\}$.

Both these choices are problematical even for problems having moderately large parameter dimension N. The first choice results in $M = (p+1)^N$ interpolation points, a number which grows explosively with increasing N; this is perhaps the most egregious instance of the curse of dimensionality. For the second, we have $M = (N+p)!/(N!p!)$ interpolation points, i.e., much slower growth than for the tensor product case, so that the inevitable fatal effects of the curse are postponed to higher dimensions. However, this choice requires a judicious choice of the location of the interpolation points because arbitrary choices can result in large Lebesgue constants which can lead to serious deterioration in accuracy. Unfortunately, good choices of total degree interpolation points in N-dimensional cubes are not known, even for moderate values of N.

A third choice for the interpolation abscissas are sparse-grid points, constructed from the roots of either the nested Chebyschev (Clenshaw-Curtis) or the Guassian polynomials. [2, 38, 39]. Typically, in these approaches the index set is defined using the Smolyak method [5, 45] where

$$\mathcal{J}(p) = \left\{ \mathbf{p} \in \mathbb{N}^N : \sum_{n=1}^{N} \alpha_n f(p_n) \leq f(p) \right\} \text{ with } f(p) = \begin{cases} 0, & p = 0 \\ 1, & p = 1 \\ \lceil \log_2(p) \rceil, & p \geq 2 \end{cases}.$$

Other polynomial spaces have been described and considered in, e.g., [4, 49].

For any choice of interpolation points, a reduction in the number of interpolation points can be effected by using *dimension-adaptive* global polynomials. For example, for the tensor product and total degree cases, one can use the index sets $\mathcal{J}(p) = \{\mathbf{p} \in \mathbb{N}^N : \max_{1 \leq n \leq N} \alpha_n p_n \leq \alpha_{min} p\}$ and $\mathcal{J}(p) = \{\mathbf{p} \in \mathbb{N}^N : \sum_{n=1}^{N} \alpha_n p_n \leq \alpha_{min} p\}$, respectively, where the weights $\alpha_n > 0$, $n = 1, \ldots, N$, can be computed either a priori or a posteriori; see [38].

3.3 ADAPTIVE HIERARCHICAL SPARSE-GRID LAGRANGE INTERPOLATION

None of the approaches discussed above are effective in approximating solutions $u(\mathbf{y}, x, t)$ of (2.10)-(2.11) that have irregular dependence with respect to the random parameters. This is perhaps even truer for those methods, commonly refereed to as polynomial chaos methods, that use global orthogonal polynomials in the parameter space. What is required for the effective approximation of solutions having irregular dependence with respect to the random parameters is an approximating space that allows for, through a judicious choice of basis, a multi-level, multi-scale decomposition. Such an approach can be constructed using piecewise polynomial approximations in the parameter space with multi-level, multi-scale hierarchical bases. A step in this direction was the development of an adaptive piecewise linear hierarchical sparse-grid approximation [5, 22] and their utilization for solving problems with random inputs [32, 33]. In this section we discuss hierarchical sparse-grid Lagrange interpolation approaches and also specialize to the approach [5, 22, 32, 33]. In §4, we extend this technique by developing a multi-dimensional multi-resolution interpolating wavelet-based approximation.

Instead of using global polynomial interpolating spaces that attempt to achieve greater accuracy by increasing the degree p of the polynomial space, piecewise polynomial interpolation spaces

attempt to achieve greater accuracy with a fixed polynomial degree by refining the grid that is the underpinning of the definition of the space. Problems having solutions with irregular behavior cannot take advantage of increases in the degree of the polynomials used; however, through local grid refinement in regions where the solution exhibits irregular behavior, piecewise polynomial spaces have the potential to be effective for such problems. However, realizing that potential for problems with even moderate parameter dimension N is not a straightforward matter.

Piecewise polynomial spaces for Lagrange interpolation are most often implemented in a standard "finite element" manner using locally supported nodal basis functions. One advantage of this approach is that the basis functions have the "delta property" (see Remark 3.1). However, such choices do not result in a multi-scale basis so that defining reliable error indicators for adaptive refinement is a difficult matter and, in fact, obtaining approximations that are optimal with respect to the number of degrees of freedom used to achieve a desired accuracy is not possible. We also focus the discussion on sparse-grid hierarchical polynomial interpolation because multi-dimensional approximations based on tensor product grids are not viable for high-dimensional parameter spaces, even for polynomial degree $p = 1$, because of the large number of degrees of freedom involved.

That is, for each parameter dimension $n = 1, \ldots, N$, we define $V_n := L^2_\rho(\Gamma_n)$. Then, the desired approximation is based on a sequence of subspaces $\{V_{i_n}\}_{i_n=0}^{\infty}$ of V of increasing dimension M_{i_n} which is dense in V_n, i.e., $\cup_{i_n=0}^{\infty} V_{i_n} = V_n$. The sequence of spaces is also required to be nested, i.e.,

$$V_0 \subset V_1 \subset V_2 \subset \cdots \subset V_{i_n} \subset V_{i_n+1} \subset \cdots \subset V_n. \tag{3.3}$$

A set of subspaces satisfying these requirements are defined as the span of a nodal piecewise polynomial basis of order p, i.e.,

$$V_{i_n} = \text{span}\{\phi^{i_n}_{j_n}(y_n) \mid 0 \le j_n \le 2^{i_n}\}, \tag{3.4}$$

where i_n denotes the scaling level of all the basis functions $\phi^{i_n}_{j_n}$ with compact support, i.e., $\text{supp}(\phi^{i_n}_{j_n}) = O(2^{-i_n})$, and $\phi^{i_n}_{j_n}(y_n)$ is a polynomial of degree p. For example, suppose M_{i_n} distinct points are selected in the interval $\overline{\Gamma}_n$ such that the maximum distance between any two neighboring points is of order $O(2^{-i_n})$. Then, the simplest choice for the set $\{\phi^{i_n}_{j_n}\}_{j_n=1}^{M_{i_n}}$ are the linear "hat" functions corresponding to the selected points in $\overline{\Gamma}_n$; in this case, we indeed have that the support of each $\phi^{i_n}_{j_n}(y_n)$ is of order $O(2^{-i_n})$.

Similarly, for an N-dimensional problem, we define $\mathcal{V}^N := L^2_\rho(\Gamma)$. Then, a sequence of subspaces $\{\mathcal{V}^N_l\}_{l=0}^{\infty}$ of $\mathcal{V}^N$ can be constructed using a sparse-grid framework, i.e.,

$$\mathcal{V}^N_l = \bigcup_{|\mathbf{i}|\le l} \bigotimes_{n=1}^{N} V_{i_n} = \bigcup_{|\mathbf{i}|\le l} \text{span}\left\{ \prod_{n=1}^{N} \phi^{i_n}_{j_n}(y_n) \;\middle|\; 0 \le j_n \le 2^{i_n} \right\}, \tag{3.5}$$

where $\mathbf{i} = (i_1, \ldots, i_N) \in \mathbb{N}^N_+$ is a multi-index and $|\mathbf{i}| \equiv i_1 + \cdots + i_N \le l$ defines the resolution of the sparse-grid approximation in $\mathcal{V}^N_L$. Note that full tensor-product resolution is defined by simply replacing the index set by $\max_{n=1\ldots,N} i_n \le l$.

Instead of using locally supported nodal bases, we construct a hierarchical approximation at level L using a truncation $\mathcal{V}^N_L$ of the infinite expansion $\mathcal{V}^N$. We begin with a basis for $\mathcal{V}^N_0$ and

then, due to the nested property of $\{\mathcal{V}_l^N\}_{l=0}^{\infty}$, we express the finer subspaces of $\mathcal{V}_l^N$ as a direct sum $\mathcal{V}_l^N = \mathcal{V}_{l-1}^N \oplus \mathcal{W}_l^N$, where $\mathcal{W}_l^N = \mathcal{V}_l^N \,/ \oplus_{m=0}^{l-1} \mathcal{V}_m^N$. Therefore, we have that

$$\mathcal{V}_L^N = \mathcal{V}_0^N \oplus \mathcal{W}_1^N \oplus \cdots \oplus \mathcal{W}_L^N. \tag{3.6}$$

Then, the *hierarchical sparse-grid approximation* of $u_{M_h,M}(\mathbf{y}, x, t) \in \mathcal{V}_L^N \otimes L^2(W(D); 0, T)$ in (3.1) is defined by

$$u_{M_h,M}(\mathbf{y}, x, t) \equiv \mathcal{I}_L^N(u)(\mathbf{y}, x, t) = \sum_{l=0}^{L} \sum_{|\mathbf{i}|=l} \sum_{\mathbf{j} \in B_\mathbf{i}} c_\mathbf{j}^\mathbf{i}(x, t) \psi_\mathbf{j}^\mathbf{i}(\mathbf{y}), \tag{3.7}$$

where $\mathcal{I}_L^N : \mathcal{V}^N \to \mathcal{V}_L^N$ denotes the approximation operator, $\psi_\mathbf{j}^\mathbf{i} = \prod_{n=1}^N \phi_{j_n}^{i_n}$ denotes a multi-dimensional hierarchical polynomial, and $B_\mathbf{i}$ a multi-index set defined by

$$B_\mathbf{i} \equiv \left\{ \mathbf{j} \in \mathbb{N}_+^N \;\middle|\; \begin{array}{ll} 0 \le j_n \le 2^{i_n},\ j_n \text{ odd}\,, & 1 \le n \le N,\ \text{if } i_n > 0 \\ j_n = 0, 1, & 1 \le n \le N,\ \text{if } i_n = 0 \end{array} \right\}. \tag{3.8}$$

The approximation space $\mathcal{P}_p(\Gamma) = \mathcal{V}_L^N$ and the particular basis chosen are required to possess the following properties.

P_1) **Nested hierarchical subspaces:** $\mathcal{V}_0^N \subset \mathcal{V}_1^N \subset \cdots \subset V_\infty^N$.

P_2) **Small compact support:** $\text{supp}\left(\prod_{n=1}^N \phi_{j_n}^{i_n}\right) = O\left(2^{-\sum_{n=1}^N i_n}\right)$.

P_3) **Interpolatory basis:** $\{\phi_j^i\}$ in (3.4) is an interpolating basis for V_i, e.g., the "hat" functions, so that the approximation operator $\mathcal{I}_L^N$ in (3.7) is a multi-dimensional interpolation operator.

P_4) **Decay of the coefficients for smooth functions in $L_\rho^2(\Gamma)$:** there exits a constant C, independent of the level L, such that for every $u(\mathbf{y}, \cdot, \cdot) \in L_\rho^2(\Gamma)$ the following holds:

$$\sum_{l=0}^{L} \sum_{|\mathbf{i}|=l} \sum_{\mathbf{j} \in B_\mathbf{i}} \left|c_\mathbf{j}^\mathbf{i}\right|^2 2^{2l} \le CL \|u\|_{L_\rho^2(\Gamma)}^2. \tag{3.9}$$

Denote by $\mathcal{H}_\mathbf{i} = \{\mathbf{y}_\mathbf{j}^\mathbf{i} \mid \mathbf{j} \in B_\mathbf{i}\}$ the set of points corresponding to the basis functions $\psi_\mathbf{j}^\mathbf{i}(\mathbf{y})$ with $\mathbf{j} \in B_\mathbf{i}$; then, the set of points corresponding to the subspace $\mathcal{W}_l^N$ is given by $\cup_{|\mathbf{i}|=l} \mathcal{H}_\mathbf{i}$ and the set of points used by $\mathcal{I}_L^N(u)$ is defined by

$$\mathcal{H}_L^N = \bigcup_{|\mathbf{i}| \le L} \mathcal{H}_\mathbf{i} \tag{3.10}$$

which is the sparse grid corresponding to $\mathcal{V}_L^N$. Note that due to property P_1, the sparse grid $\mathcal{H}_L^N$ is also nested, i.e., $\mathcal{H}_{L-1}^N \subset \mathcal{H}_L^N$. In Figure 1, we plot the structure of a level $L = 2$ sparse grid in

$N = 2$ dimensions, without considering boundary points. The left nine sub-grids $\mathcal{H}_{\mathbf{i}}$ correspond to the nine multi-index sets $B_{\mathbf{i}}$, where

$$\mathbf{i} \in \{(0,0),(0,1),(0,2),(1,0),(1,1),(1,2),(2,0),(2,1),(2,2)\}.$$

The level $L = 2$ sparse grid $\mathcal{H}_2^2$ shown on the right (top) includes only six of the nine sub-grids in black according to the criterion $|\mathbf{i}| \leq 2$. Moreover, due to the nested property of the hierarchical basis, $\mathcal{H}_2^2$ has only 17 points, as opposed to the 49 points of the full tensor-product grid.

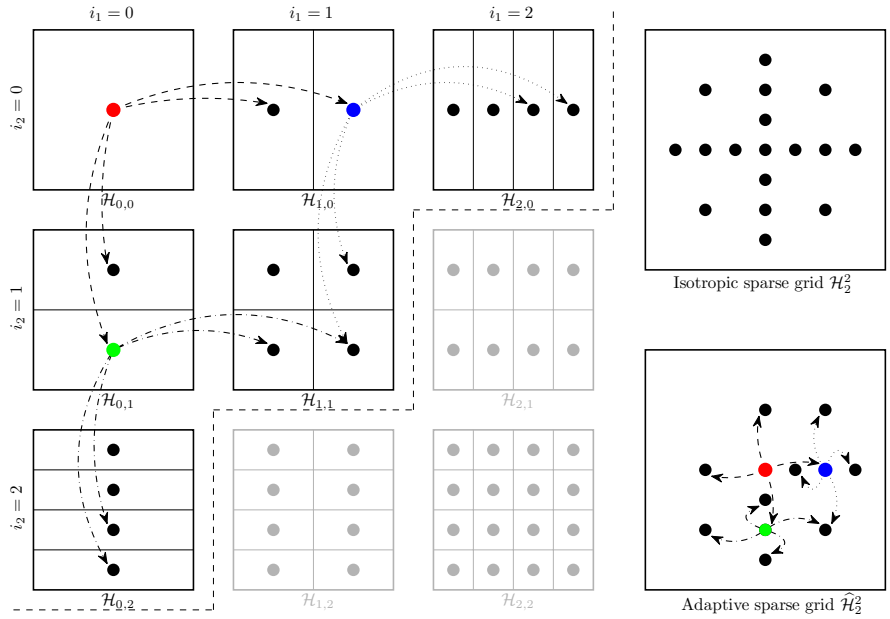

Figure 1: Nine tensor-product sub-grids (left) for level $L = 0, 1, 2$ of which only the 6 sub-grids with $i_1 + i_2 \leq 2$ are chosen to appear in the level $L = 2$ isotropic sparse grid $\mathcal{H}_2^2$ (right-top) containing 17 points. With adaptivity, each point that corresponds to a large surplus, e.g., the points in red, blue, or green, lead to 2 children points added in each direction resulting in the adaptive sparse grid $\widehat{\mathcal{H}}_2^2$ (right-bottom) containing 12 points.

Next, we explain how to compute the coefficients $c_{\mathbf{j}}^{\mathbf{i}}(x,t)$. In general, this requires the solution of a linear system whose right-hand-side depends only on the value of the finite element approximation of the solution u at each collocation point. Moreover, the structure of the coefficient matrix depends on the type of hierarchical polynomials used in (3.7). However, for some choices of the basis, these coefficients can be computed explicitly.

Example 3.2 (Linear hierarchical piecewise polynomials) *We can take the hierarchical one-dimensional functions to be the standard piecewise linear finite element basis, i.e., the basis function ϕ_j^i in (3.4) are obtained by the dilation and translation of the function*

$$\phi(y) = \begin{cases} 1 - |y| & \text{if } y \in [-1,1] \\ 0 & \text{otherwise.} \end{cases} \tag{3.11}$$

This basis possesses properties $P_1 - P_4$. Examples of higher-degree bases are given in [5]. Then, $\mathcal{I}_L^N(u)$ in (3.7) can be rewritten as

$$\mathcal{I}_L^N(u)(\mathbf{y}, x, t) = \mathcal{I}_{L-1}^N(u)(\mathbf{y}, x, t) + \Delta\mathcal{I}_L^N(u)(\mathbf{y}, x, t), \tag{3.12}$$

where $\mathcal{I}_{L-1}^N(u)$ is the sparse-grid approximation in $\mathcal{V}_{L-1}^N$ and $\Delta\mathcal{I}_L^N(u)$ is the hierarchical difference interpolant corresponding to $\mathcal{W}_L^N$. Due to property P_1, the set of grid points used by $\Delta\mathcal{I}_L^N(u)$ can be denoted by $\Delta\mathcal{H}_L^N = \mathcal{H}_L^N \backslash \mathcal{H}_{L-1}^N$. Then, due to the interpolatory property P_3 and the choice of the basis function (3.11), by substituting $\mathbf{y}_\mathbf{j}^\mathbf{i} \in \Delta\mathcal{H}_L^N$ in (3.12), we obtain that

$$\begin{aligned} c_\mathbf{j}^\mathbf{i}(x,t) &= \mathcal{I}_L^N(u)(\mathbf{y}_\mathbf{j}^\mathbf{i}, x, t) - \mathcal{I}_{L-1}^N(u)(\mathbf{y}_\mathbf{j}^\mathbf{i}, x, t) \\ &= u(\mathbf{y}_\mathbf{j}^\mathbf{i}, x, t) - \mathcal{I}_{L-1}^N(u)(\mathbf{y}_\mathbf{j}^\mathbf{i}, x, t) \end{aligned} \tag{3.13}$$

as the hierarchical surplus. This is simply the difference between the solution u at a point $\mathbf{y}_\mathbf{j}^\mathbf{i}$ on the current level of interpolation and the interpolated value of the previous level [25] at that point. Therefore, using the recursive formula (3.12), we can compute all the coefficients $c_\mathbf{j}^\mathbf{i}$ in (3.7) by calculating the coefficients of $\Delta\mathcal{I}_L^N(u)$ for $l = 1, \ldots, L$.

According to the analysis in [25], for smooth functions described by property P_4, the hierarchical surpluses tend to zero as the interpolation level goes to infinity. On the other hand, for irregular functions having, e.g., steep slopes or jump discontinuities, the magnitude of the surplus is an indicator of the interpolation error and, as such, can be used to control the error adaptively. That is, for the sparse grid $\mathcal{H}_L^N$, abscissas involved in each direction can be considered as a tree-like data structure as shown in Figure 1.

For example, on the left, we show that the red point in $\mathcal{H}_{0,0}$ has 2 children points at level $L = 1$ in each of the horizontal and vertical directions; the 4 children are indicated by the arrows emanating from the red point. Each of its 4 children also have 4 children of their own at level $L = 2$, and so on for subsequent levels. Suppose the magnitude of the coefficients (the surplus) associated with the blue and green children are larger than a prescribed tolerance, but those for the two black children of the red point are smaller than the tolerance. In this case, refinement is effected only from the blue and green children; no refinement is done of the black children. This is indicated by having, at level $L = 2$, four arrows emanate from the blue and green points, but none from the black points. We arrive at the adaptive space grid $\widehat{\mathcal{H}}_{2,2}$ that has 12 total collocation points, shown on the right (bottom) of Figure 1. The analagous (non-adaptive) isotropic sparse grid, which has 17 collocation points, is also shown on the right (top).

In general, a grid point in a N-dimensional space has $2N$ children which are also the neighbor points of the parent node. However, note that the children of a parent point correspond to hierarchical basis functions on the next interpolation level, so that we can build the interpolant $\mathcal{I}_L^N(u)$ in (3.7) from level $L-1$ to L by adding those points on level L whose parent has a surplus greater than our prescribed tolerance. In this way, we can refine the sparse grid locally and end up with an adaptive sparse grid which is a sub-grid of the corresponding isotropic sparse grid.

A *sparse grid adaptive linear stochastic collocation method* (sg-ALSCM) that utilizes a locally supported linear hierarchical basis, given by (3.11), to approximate random functions in the multi-dimensional hypercube $\Gamma \subset \mathbb{R}^N$ were considered in [16, 24, 32, 33]. As mentioned in Example 3.2,

the expansion coefficients $c_{\mathbf{j}}^{\mathbf{i}}(x,t)$ in (3.7) are simply the hierarchical surpluses and adaptive refinement is guided by the magnitude $|c_{\mathbf{j}}^{\mathbf{i}}|$ of those coefficients. However, this approach has a major drawback: one cannot estimate the error from *below* with constants independent of the number of hierarchical levels involved. Thus, the linear hierarchical basis does not form a stable multi-scale splitting of the approximation space [41] and the absolute value of a hierarchical coefficient is just a local error indicator and not a true error estimator. As a result, one obtains sufficiently refined sparse approximations for which the error is behaving as predicted, but in doing so, one may have used many more grid points than needed to achieve a prescribed error tolerance for the adaptive procedure. This scheme has no guarantee of efficiency so that some previous claims of optimality with respect to complexity for this approach are heuristic, not provable and, in general, not valid.

Our approach is generally similar, but uses multi-resolution *wavelet* approximations that possess all the properties ($P_1 - P_4$) of the linear basis functions, but also possess an additional property that guarantee optimality. We will introduce this essential criteria in Section 4 and more importantly, also explain the advantages of our novel adaptive wavelet stochastic collocation method (AWSCM).

4 ADAPTIVE WAVELET STOCHASTIC METHOD

As discussed several times in the paper, UQ for complex stochastic systems that require the approximation and/or resolution of statistical QoIs involving, e.g., steep gradients, sharp transitions, bifurcations, or finite jump discontinuities, in possibly high-dimensional probabilistic domains, require sophisticated multi-dimensional multi-resolution adaptive algorithms. To be effective, however, refinement strategies must be guided by accurate estimates of errors (both local and global) while not expending unnecessary computational effort approximating the QoI with respect to any random dimension. In the sg-ALSCM described in §3.3, given the hierarchical sparse-grid approximation (3.7) that satisfies properties $P_1 - P_4$, optimal approximations of such irregular problems cannot be guaranteed. Here, by optimal we mean achieving a prescribed error tolerance with a minimal number of grid points. This can result in an explosion in computational effort for high-dimensional problems. Towards alleviating this effect, we require the following additional property of the basis functions $\psi_{\mathbf{j}}^{\mathbf{i}}$ (3.7), namely

P_5) **Riesz property:** the basis $\{\psi_{\mathbf{j}}^{\mathbf{i}}\}$ in (3.7)$\}$ is a Riesz basis so that there exists a constant $C_R > 0$, independent of the level L, such that for all $\mathcal{I}^L(u)$ given by (3.7) the following holds:

$$C_R^{-1} \sum_{l=0}^{L} \sum_{|\mathbf{i}|=l} \sum_{\mathbf{j} \in B_{\mathbf{i}}} \left| c_{\mathbf{j}}^{\mathbf{i}} \right|^2 \leq \left\| \mathcal{I}_N^L(u) \right\|_{\mathcal{V}_N}^2 \leq C_R \sum_{l=0}^{L} \sum_{|\mathbf{i}|=l} \sum_{\mathbf{j} \in B_{\mathbf{i}}} \left| c_{\mathbf{j}}^{\mathbf{i}} \right|^2, \tag{4.1}$$

where the set of multi-indices $B_{\mathbf{i}}$ is defined as in (3.8) and $\mathcal{I}_N^L(u) \equiv \mathcal{I}_N^L(u)(\mathbf{y}, \cdot, \cdot)$.

Unfortunately, finite element bases such as the linear hierarchical polynomials used in the sg-ALSCM of [32, 33] are not Riesz bases, so norms of such approximations can only be bounded from above (but not from below) by sequence norms of the corresponding coefficients. In other words, they are not L^2_ρ-stable, as implied by property P_5. The same can be said for the high-order hierarchical polynomial basis in [5], the Lagrangian interpolation polynomials used in [38, 39], as well as the orthogonal polynomials of [26, 28, 52].

On the other hand, standard Riesz bases, e.g., Fourier and orthogonal polynomials, consist of functions that are globally supported. In the numerical PDE setting, this has the disadvantage of leading to dense stiffness matrices and, in the UQ setting, to intrusive methods. However, *certain classes of hierarchical wavelet and pre-wavelet bases are not only Riesz bases, but consist of compactly supported basis functions.* Thus, we have the best of both worlds: the compact support property of standard finite element bases and the Riesz basis property of spectral bases. Moreover, an interpolating wavelet basis can be utilized for the approximation given by (3.7), satisfies all the properties $P_1 - P_5$, and forms a stable multi-resolution analysis of the stochastic space L^2_ρ as defined in [13]. Hence, for the interpolating wavelet basis, we obtain the two-sided estimates given by P_5. Therefore, the magnitude of the wavelet coefficients $|c_{\mathbf{j}}^{\mathbf{i}}|$ in (3.7) can serve as *true local error estimators* and the the lower bounds provided by the Riesz basis property gives us *a rigorous indicator of the efficiency of adaptive schemes.* This means a prescribed error tolerance is reached at the cost of just the optimal number of points (up to a constant) in a sparse grid

adaptation process. This results in a superior convergence rate when compared to methods using other hierarchical multi-scale basis functions. We choose one particular class of second-generation wavelets, namely lifted interpolating wavelets on the bounded interval, to achieve this goal. We next provide details about such wavelets.

4.1 SECOND-GENERATION WAVELETS AND THE LIFTING SCHEME

Second-generation wavelets are a generalization of biorthogonal wavelets that are more easily applied to functions defined on bounded domains. Second-generation wavelets form a Reisz basis for some function space, with the wavelets being local in both "spatial" domain (in our context, the parameter domain) and the frequency domain and often having many vanishing polynomial moments, but they do not possess the translation and dilation invariance of their biorthogonal cousins. The lifting scheme [46, 47] is a tool for constructing second-generation wavelets that are no longer dilates and translates of one single scaling function. In contrast to first-generation wavelets, which use the Fourier transform to build the wavelet basis, a construction using lifting is performed exclusively in the "spatial" domain so that wavelets can be custom designed for complex domains and irregularly sampled data.

The basic idea behind lifting is to start with simple multi-resolution analysis and gradually build a multi-resolution analysis with specific, a priori defined properties. The lifting scheme can be viewed as a process of taking an existing first-generation wavelet and modifying it by adding linear combinations of the scaling function at the coarse level. To explain the procedure in detail, we follow the notation in Section 3.3. The approximation space $V_i = \text{span}\{\phi_j^i \mid 0 \le j \le 2^i\}$ in (3.4) has a decomposition $V_i = V_{i-1} \oplus W_i$, where V_{i-1} and W_i are defined by

$$V_{i-1} = \text{span}\{\phi_j^{i-1} | 0 \le j \le 2^{i-1}\} \quad \text{and} \quad W_i = \text{span}\{\phi_j^i | 0 \le j \le 2^i,\ j \text{ odd }\}. \tag{4.2}$$

Here, W_i is viewed as the hierarchical subspace on level i, and $\phi_j^i \in W_i$ are the first-generation interpolating wavelets. Then, the corresponding second-generation wavelet $\widehat{\phi}_j^i$ is constructed by "lifting" ϕ_j^i as

$$\widehat{\phi}_j^i \equiv \phi_j^i + \sum_{\hat{j}=0}^{2^{i-1}} \alpha_{\hat{j},j}^{i-1} \phi_{\hat{j}}^{i-1}, \tag{4.3}$$

where the weights $\alpha_{\hat{j},i}^{i-1}$ in the linear combination are chosen in such a way that the new wavelet $\widehat{\phi}_j^i$ has more vanishing moments than ϕ_j^i and thus provides a stabilization effect. If we apply this approach to the piecewise linear hierarchical basis, i.e., to the "hat" functions, in such a way that the lifting wavelet basis has two vanishing moments, we end up with

$$\begin{aligned}
\widehat{\phi}_j^i &= \phi_j^i - \frac{1}{4}\phi_{\frac{j-1}{2}}^{i-1} - \frac{1}{4}\phi_{\frac{j+1}{2}}^{i-1} && \text{for } 1 < j < 2^i - 1,\ j \text{ odd} \\
\widehat{\phi}_j^i &= \phi_j^i - \frac{3}{4}\phi_{\frac{j-1}{2}}^{i-1} - \frac{1}{8}\phi_{\frac{j+1}{2}}^{i-1} && \text{for } j = 1 \\
\widehat{\phi}_j^i &= \phi_j^i - \frac{1}{8}\phi_{\frac{j-1}{2}}^{i-1} - \frac{3}{4}\phi_{\frac{j+1}{2}}^{i-1} && \text{for } j = 2^i - 1,
\end{aligned} \tag{4.4}$$

where the three equations define the central "mother" wavelet, the left-boundary wavelet, and the right-boundary wavelet, respectively. We illustrate the three lifting wavelets in Figure 2. For additional details, see [48].

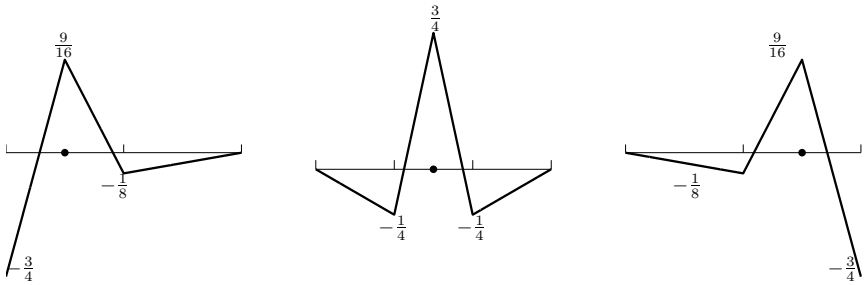

Figure 2: Left-boundary wavelet (left), central wavelet (middle), right-boundary wavelet (right).

Due to the fact that our second-generation wavelets are lifted from the first-generation wavelets, properties $P_1 - P_4$ are guaranteed. In addition, from the analysis provided in [46, 47], we know that they also constitute a Riesz basis so that property P_5 is satisfied. Therefore, introduction of the lifted wavelet basis into the hierarchical sparse-grid approximation framework results in a novel non-intrusive sparse grid adaptive wavelet stochastic collocation method (sg-AWSCM). This method allow us to encapsulate the advantages of the increased convergence rates of standard SCM and polynomial chaos approaches resulting from higher-order polynomial expansion (p-refinement) [38] with the robustness of optimal local decompositions (h-refinement) [12] for the approximation of irregular solutions and QoIs coming from PDEs with random inputs in high-dimensional stochastic domains.

Note that due to the interpolatory property of the wavelet basis, when computing the wavelet coefficients in (3.1), we only face an interpolation problem. That is, from the construction procedure of the lifted wavelets described above, we observe that neighboring wavelet basis function at the same level have overlapping support such that the resulting interpolation matrix for our sg-AWSCM is has greater bandwidth compared to that for sg-ALSCM. For the one-dimensional problem, the paper [48] proposed fast algorithms for computing wavelet coefficients. We are currently working on extending their algorithms to the multi-dimensional case, but in this paper, we just use mature numerical libraries, e.g., LINPACK. to solve the linear system for the interpolation coefficients.

5 NUMERICAL EXAMPLES

This section illustrates the convergence properties of the sparse grid adaptive wavelet collocation method for solving three problems. In all examples, we use the linear hierarchical second generation lifted wavelets described in Section 4.1. The first example is used to compare our sg-AWSCM with the sg-ALSCM for approximating irregular (deterministic) functions in $N = 2$ parameter dimensions. In the second example, we apply our new approach to solve: (a) the Burgers equation with random boundary condition data and (b) the time-dependent Riemann problem for the Burgers equation with random initial conditions. Finally, in the third example, we investigate the ability of the sg-AWSCM to detect the important random dimensions in a elliptic problem having a moderately high number of random parameter inputs. As opposed to previous dimension-adaptive approach of [39], our new sg-AWSCM does not require *a priori* nor *a posterori* estimates to guide adaptation. Instead, as described in Section 4, our multi-dimension multi-resolution adaptive approach uses only the sparse grid wavelet coefficient to guide refinement while maintaining near-optimal convergence. We will also use this problem to compare the convergence of our sg-AWSCM with other ensemble-based methods such as the isotropic sparse grid method and the sg-ALSCM and to compare all these approaches to the best N-term sparse grid approximation.

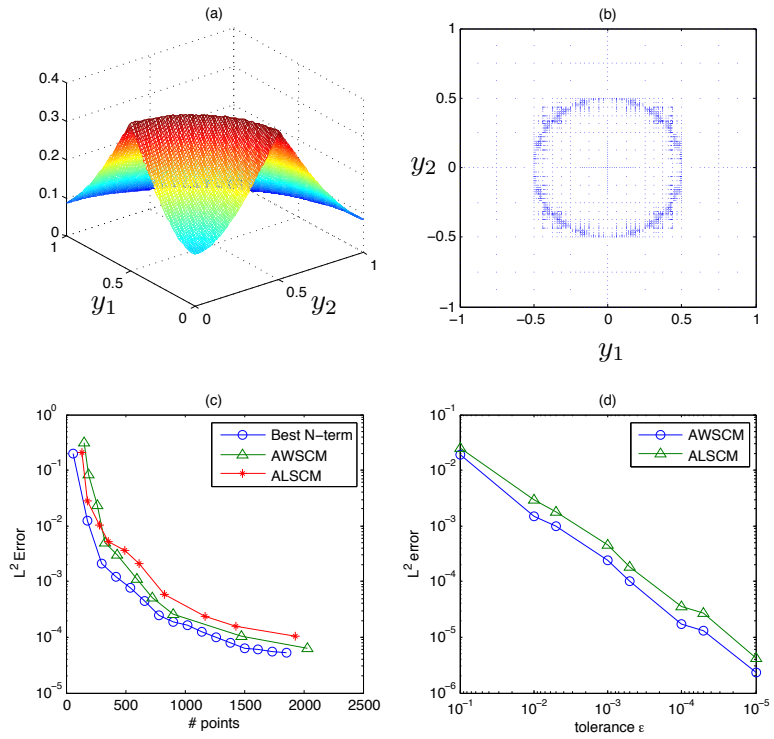

Figure 3: Results for $f_1(y_1, y_2)$ in Example 1: (a) the target function; (b) the points used by the sg-AWSCM for a tolerance $\varepsilon = 10^{-3}$; (c) error decay vs. number of points; (d) error decay vs. the tolerance ε.

5.1 APPROXIMATION OF IRREGULAR DETERMINISTIC FUNCTIONS

Consider the two bivariate functions $f_1(y_1, y_2)$ on $[-1, 1]^2$ and $f_2(y_1, y_2)$ on $[0, 1]^2$ defined by

$$f_1(y_1, y_2) = \begin{cases} \exp(-2(y_1^2 + y_2^2)) & \text{if } y_1^2 + y_2^2 \geq 0.5 \\ 2\exp(-\frac{1}{2}) - \exp(-2(y_1^2 + y_2^2)) & \text{if } y_1^2 + y_2^2 < 0.5 \end{cases} \tag{5.1}$$

$$f_2(y_1, y_2) = \frac{1}{|0.15 - y_1^2 - y_2^2| + 0.1} \quad \text{on} \quad [0, 1] \times [0, 1]. \tag{5.2}$$

It is easy to see that $f_1(y_1, y_2)$ and $f_2(y_1, y_2)$ represent two types of irregular behavior. The function $f_1(y_1, y_2)$ has a jump in its first-order derivatives $\partial f_1/\partial y_1$ and $\partial f_1/\partial_2$ across the circle $y_1^2+y_2^2 = 0.5$ whereas $f_2(y_1, y_2)$ has a steep gradient across the curve $y_1^2 + y_2^2 = 0.15$. To construct interpolants for both $f_1(y_1, y_2)$ and $f_2(y_1, y_2)$ using the sg-ALSCM and the sg-AWSCM, we first build a level $L = 3$ isotropic sparse grid as the initial grid, then add nodes adaptively guided by linear hierarchical surpluses or wavelet coefficients, respectively. The interpolation results for $f_1(y_1, y_2)$ are shown in Figure 3. Figure 3(a) displays the function f_1; only the first quadrant is shown due to the symmetries of the function. Figure 3(b) reveals the resulting adaptive sparse interpolation grid constructed from the lifted wavelets for a tolerance $\varepsilon = 10^{-3}$. In Figure 3(c), we show the optimality of our approximation by plotting the convergence rates for the sg-ALSCM and sg-AWSCM approximations as well as for the best N-term approximation that is obtained by extracting the N terms with the N biggest coefficients from an approximation on a non-adaptive, isotropic sparse grid. We observe that the convergence behavior of the sg-AWSCM more closely matches that of the best N-term approximation, compared to the sg-ALSCM, which results in a reduction in the number of function evaluations to achieve the desired accuracy $\varepsilon = 10^{-3}$. In Figure 3(d), we also plot the convergence behavior of both methods versus the tolerance ε. We see that for the same prescribed tolerance for the hierarchical surpluses, the sg-AWSCM can achieve higher accuracy than the sg-ALSCM. Similar conclusions can be made by examining Table 1, where we show the number of sparse grid points required by the various interpolants to achieve a desired accuracy. In all cases the sg-AWSCM outperforms the sg-ALSCM and more closely matches the best N-term approximation.

error α	sg-ALSCM	sg-AWSCM	best N-term
5.0E-03	366	330	265
1.0E-03	774	623	479
5.0E-04	920	737	640
1.0E-04	1927	1548	1261

Table 1: For $N = 2$ dimensions, we compare the number of function evaluations required by the isotropic sparse grid (ISG), the sg-ALSCM, and sg-AWSCM and the best N-term approximation to compute the interpolated approximation of $f_1(y_1, y_2)$ to an accuracy smaller than the prescribed error tolerance α, i.e., so that $\|\mathcal{I}_L^N(f_1)(y_1, y_2) - f_1(y_1, y_2)\| \leq \alpha$.

error α	sg-ALSCM	sg-AWSCM	Best N-term
5.0E-02	243	238	237
1.0E-02	445	414	359
5.0E-03	638	491	431
1.0E-03	1392	1062	902

Table 2: For $N = 2$ dimensions, we compare the number of function evaluations required by the isotropic sparse grid (ISG), the sg-ALSCM, and sg-AWSCM and the best N-term approximation to compute the interpolated approximation of $f_2(y_1, y_2)$ to an accuracy smaller than the prescribed error tolerance α, i.e., so that $\|\mathcal{I}_L^N(f_1)(y_1, y_2) - f_1(y_1, y_2)\| \leq \alpha$.

The same observations and conclusions can be reached for the function $f_2(y_1, y_2)$ by examining Figure 4 and Table 2. Additionally, in Figure 5, we show the condition number of the linear system used to construct the interpolation wavelet coefficients $f_1(y_1, y_2)$; we see that the interpolation matrix is well-conditioned. Therefore, as expected, due to the additional property P_5 and the well-conditioning of the interpolation matrix for the wavelet coefficients, when approximating functions with discontinuous derivatives, the sg-AWSCM substantially reduces the complexity of determining an accurate interpolant compared to the sg-ALSCM.

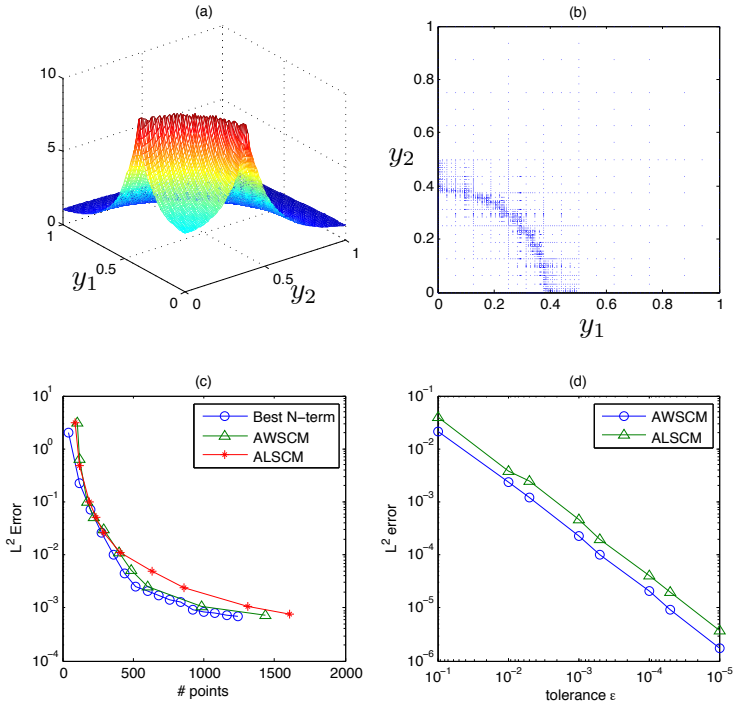

Figure 4: Results for $f_2(y_1, y_2)$ in Example 1: (a) the target function; (b) the points used by the sg-AWSCM for a tolerance $\varepsilon = 10^{-2}$; (c) error decay vs. number of points; (d) error decay vs. the tolerance ε.

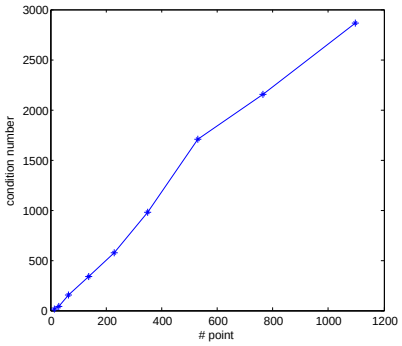

Figure 5: Condition number vs. the number of function evaluations for sg-AWSCM for Example 1.

5.2 BURGERS EQUATION WITH RANDOM INPUTS

We next apply our novel sg-AWSCM to construct optimal approximations of solutions of two Burgers equation problems. First, we consider the steady viscous Burgers equation with random boundary condition data:

$$\begin{cases} \dfrac{1}{2}\dfrac{\partial u^2}{\partial x} - \nu \dfrac{\partial^2 u}{\partial x^2} = 0 \quad \text{in} \quad [-1,1] \\ u(-1) = y(\omega), \qquad u(1) = 0, \end{cases} \tag{5.3}$$

where the left boundary condition data is a uniformly distributed random variable $y(\omega) \sim \mathcal{U}[0.95, 1.05]$, i.e., the parameter space $\Gamma = [0.95, 1.05]$ and the PDF $\rho(y) = 10$; the viscosity is set to $\nu = 0.1$.

The deterministic solver used for this problem is as a finite difference discretization of the conservative form of the equation followed by an application of Newton's method to solve the resulting nonlinear system. Figure 6 show the computed realizations of the solution $u(y, \cdot)$ for several values of the left boundary value $y(\omega)$. Observe that perturbing $y(\omega)$ from 1 to 1.005 effects a startlingly large perturbation to the solution $u(y, \cdot)$. Thus, we conclude that the solution $u(y, \cdot)$ is very sensitive to $y(\omega)$ near $y(\omega) = 1$. In particular, this holds for the point x_0 at which the solution u changes sign. Thus, if we choose the quantity of interest to be the point x_0, we again have an instance of irregular behavior. Therefore, we focus on quantifying the uncertainty of x_0 propagated from $y(\omega)$ following the uniform distribution $\mathcal{U}[0.95, 1.05]$. To build the multi-scale interpolant using the AWSCM, we start with a 4-level uniform grid on $[0.95, 1.05]$ and then add points adaptively, guided by the size of the wavelet coefficients. The tolerance ε is set to 10^{-3}. The relation between x_0 and $y(\omega)$ and the corresponding adaptive grid are shown in Figure 7. We can see that $x_0(y)$ has a steep slope around $y(\omega) = 1$ (which accounts for its high sensitivity near that value) and that the corresponding adaptive grid is refined around the point $y(\omega) = 1$. The convergence rate of $\mathbb{E}[x_0]$ is shown in Figure 8 and compared to that of the best N-term approximation obtained by extracting N terms with N largest coefficients from an approximation on an non-adaptive, uniform grid.

Next, we solve a time-dependent Riemann problem for a Burgers equation with random initial

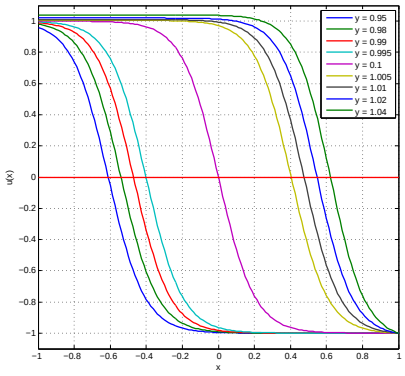

Figure 6: Sensitivity of x_0 to the left boundary value $y(\omega)$ in example (5.3).

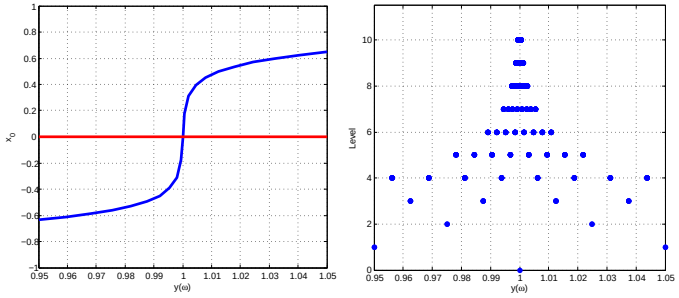

Figure 7: The relation between x_0 and $y(\omega)$ (left), adaptive grid with tolerance $\varepsilon = 10^{-3}$ (right) in example (5.3).

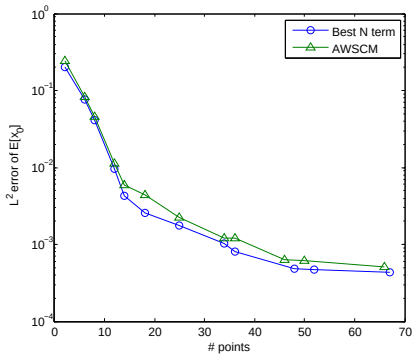

Figure 8: The convergence rate of the sg-AWSCM with tolerance $\varepsilon = 10^{-3}$ in example (5.3).

shock location [44]:

$$\begin{cases} \dfrac{\partial u}{\partial t} + \dfrac{\partial}{\partial x}\left(\dfrac{1}{2}u^2\right) = 0, \quad (x,t) \in [-1,1] \times (0,+\infty) \\ u_0(x,\omega) = \begin{cases} 1 & \text{if} \quad x < y(\omega) \\ 0 & \text{if} \quad x \geq y(\omega). \end{cases} \end{cases} \tag{5.4}$$

The initial shock location depends on an uniform random variable $y(\omega) \sim \mathcal{U}[-0.1, 0.1]$, i.e., we have the parameter space $\Gamma = [-0.1, 0.1]$ and the PDF $\rho(y) = 5$. A formula for the expectation $\mathbb{E}[u]$ and variance $\mathrm{Var}[u]$ of the exact solution u can be found in [44].

The deterministic solver used for this problem is a weighted essentially non-oscillatory (WENO) scheme. Here we consider the solution at time $t = 0.2$. We compute the approximate deterministic solution on a uniform spatial grid with 1025 points in spatial domain $[-1, 1]$. In Figure 9, we plot the expectation and variance of the approximate shock profile at $t = 0.2$, computed with the AWSCM; also plotted are the corresponding exact statistics. To test the adaptive wavelet procedure, we choose our quantities of interest to be the expectations of $u(y(\omega), x)$ at 3 locations, namely $\mathbb{E}[u](x = 0.036, t = 0.2)$, $\mathbb{E}[u](x = 0.127, t = 0.2)$, and $\mathbb{E}[u](x = 0.590, t = 0.2)$. We then build the grids using AWSCM. At each location, we start with a 2-level uniform grid on $[-0.1, 0.1]$ in the parameter space and then add points guided by the magnitudes of the wavelet coefficients. In Figure 9, we plot the adaptive grids for the three cases. We can see that the singular point of $u(y(\omega), x, t = 0.2)$ with respect to $y(\omega)$ depends on the value of x. At the time instant $t = 0.2$: if $x \in [0, 0.2]$ such as $x = 0.036$ or $x = 0.127$, then $u(y(\omega), x, t = 0.2)$ has a singular point but its location is determined by the value of x; on the other hand, there is no singular point in $u(y(\omega), x, t = 0.2)$ for $x \in [-1, 0) \cup (0.1, 1]$, including for $x = 0.590$, so that grid refinement in parameter spaces is not needed; the AWSCM method recognizes this so that the 2-level initial grid is not changed by the adaptation procedure.

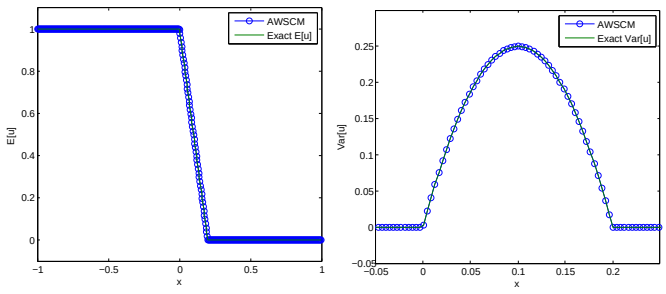

Figure 9: Expectation (left) and variance (middle) of the probabilistic shock profile in example (B_2).

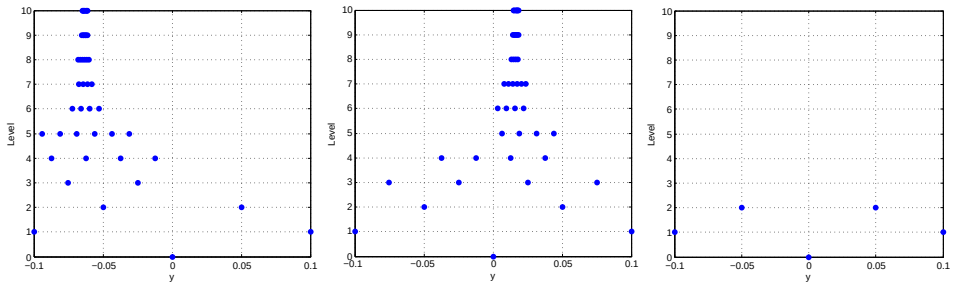

Figure 10: Adaptive grids for quantities of interest being $\mathbb{E}[u](x)$ at 3 spatial points: $x = 0.036$ (left), $x = 0.127$ (middle), $x = 0.590$ (right) in example (B_2).

5.3 ELLIPTIC PDE WITH RANDOM INPUTS

Similar to [38, 39], we consider an elliptic PDE in two spatial dimensions with random inputs. As shown in the previous examples, the AWSCM and the sg-AWSCM can accurately capture the irregular, even non-smooth regions in a low-dimensional stochastic parameter space. If the solution depends on a moderately large number of random variables with sufficient regularity (analytic in this case), the major challenge of the numerical approximation is anisotropic (dimension-adaptive) refinement. Therefore, in this example we demonstrate the ability of the sg-AWSCM method to detect important dimensions when the random variables do not "weigh equally" in the stochastic solution.

The specific problem we solve is given by

$$\begin{cases} -\nabla \cdot \big(a(\omega,x)\nabla u(\omega,x)\big) &= \cos(x_1)\sin(x_2) \quad \text{in } \Omega \times D \\ u(\omega,x) &= 0 \quad \text{on } \Omega \times \partial D, \end{cases} \tag{5.5}$$

where $D = [0,1]^2$, x_1 and x_2 denote the components of the spatial position vector x, and ∇ denotes the gradient operator with respect to x. The forcing term $f(\omega,x)$ is deterministic. The random diffusion coefficient $a(\omega,x)$ has a one-dimensional spatial dependence and is given by

$$\log(a(\omega,x)-0.5) = 1 + y_1(\omega)\left(\frac{\sqrt{\pi}C}{2}\right)^{1/2} + \sum_{n=2}^{N} \zeta_n \varphi_n(x_1) y_n(\omega), \tag{5.6}$$

where, for $n \geq 2$,

$$\zeta_n := (\sqrt{\pi}C)^{1/2} \exp\left(-\frac{\left(\lfloor \frac{n}{2} \rfloor \pi C\right)^2}{8}\right) \tag{5.7}$$

and

$$\varphi_n(x) := \begin{cases} \sin\left(\dfrac{\left(\lfloor \frac{n}{2} \rfloor \pi x_1\right)^2}{C_p}\right) & \text{for } n \text{ even} \\ \cos\left(\dfrac{\left(\lfloor \frac{n}{2} \rfloor \pi x_1\right)^2}{C_p}\right) & \text{for } n \text{ odd.} \end{cases} \tag{5.8}$$

In this example, the random variables $\{y_n(\omega)\}_{n=1}^{\infty}$ are independent, have zero mean and unit variance, i.e., $\mathbb{E}[y_n] = 0$ and $\mathbb{E}[y_n y_m] = \delta_{nm}$ for $n, m \in \mathbb{N}_+$, and are uniformly distributed in the interval $[-\sqrt{3}, \sqrt{3}]$. For $x_1 \in [0,1]$, let C_L be the physical correlation length of the stationary covariance function

$$Cov[\log(a-0.5)](x_1, x_1') = \exp\left(-\frac{(x_1 - x_1')^2}{C_L^2}\right). \tag{5.9}$$

Then, the parameter C_p in (5.8) is $C_p = \max(1, 2C_L)$ and the parameter C in (5.6) and (5.7) is $C = C_L/C_p$. Also, ζ_n and φ_n, for $n = 1, \ldots, N$ are the eigenvalues and eigenfunctions (given by (5.7) and (5.8) respectively) of the covariance operator defined by substituting (5.9) into (2.8). The eigenvalues ζ_n in (5.6) decay with increasing n with large values of the correlation length

C_L corresponding to fast decay. Thus, the parameter dimensions have decreasing influence on the solution as their index increases and, for large C_L, the influence decreases quickly. Therefore, an approximation requires less accurate resolution of the dimensions having small influence, compared with that for more influential dimensions so that, to achieve maximum efficiency (e.g., the fewest sample points in parameter space) for a given overall accuracy, one should use anisotropic, or dimension-adaptive set of sparse grid points.

For the numerical results, we set $C_L = \frac{1}{2}$ and retain 7 terms of the expansion (5.6) and treat the truncated version as the exact diffusion field. In this case, the eigenvalues are $\zeta_1 = 0.665$, $\zeta_2 = \zeta_3 = 0.692$, $\zeta_4 = \zeta_5 = 0.274$, $\zeta_6 = \zeta_7 = 0.059$. In order to investigate convergence rates, we compare the expected value $\mathbb{E}[u]$ approximated by our sg-AWSCM method with a tolerance $\varepsilon = 10^{-5}$ to the "exact" solution determined from simulations based on 10^6 Monte Carlo (MC) samples of the seven-dimensional parameter. Specifically, in Figure 11, for several values of the level L, we plot $\|\mathbb{E}[\mathcal{I}_L^N(u)] - \mathbb{E}_{MC}[u]\|$, i.e., the $L^2(D)$ norm of the "error" between the expected values obtained using the sg-AWSCM method and the densely sampled MC method. Also provided in that figure are the corresponding errors for the isotropic sparse grid, for the sg-ALSCM, and for the best N-term approximation defined by taking the N terms in the isotropic sparse grid solution with the largest coefficients. The errors are plotted against the number of points in parameter space each method requires to achieve the desired accuracy. As expected, due to the fast decay of the eigenvalues, the convergence, with respect to the number of points used, of both the sg-AWSCM and the sg-ALSCM is much faster than the approximation based on an isotropic sparse grid because fewer points are placed along the non-important dimensions associated with small eigenvalues. Furthermore, our new sg-AWSCM also reduces the overall complexity when compared to the sg-ALSCM approximation, and nearly matches that for the best N-term approximation. Further proof of this can be seen in Table 3 that shows a reduction in the computational complexity for computing the expected value using the sg-AWSCM, when compared to the isotropic sparse grid and sg-ALSCM, by approximately a factor of 20 and 3 respectively, to achieve a desired accuracy of 10^{-7}. In fact, for this higher-dimensional problem, the savings incurred by the sg-AWSCM compared to the sg-ALSCM as much more significant than for the previous low-dimensional examples. One can expect the relative savings to increase as one further increases the parameter dimension. This optimal performance is guaranteed by our sparse grid wavelet multi-resolution analysis.

error α	isotropic SG	sg-ALSCM	sg-AWSCM	best N-term
1.0E-05	73	34	30	25
5.0E-06	344	85	74	60
1.0E-06	2435	772	476	248
5.0E-07	7767	1271	1038	840
1.0E-07	85861	9604	3824	2812

Table 3: For $N = 7$ dimensions, we compare the number of function evaluations required by the isotropic sparse grid (ISG), the sg-ALSCM, the sg-AWSCM, and the best N-term approximation to compute the expected value of the solution to within a prescribed global error tolerance α, i.e., so that $\|\mathbb{E}[\mathcal{I}_L^N(u)] - \mathbb{E}_{MC}[u]\| \leq \alpha$.

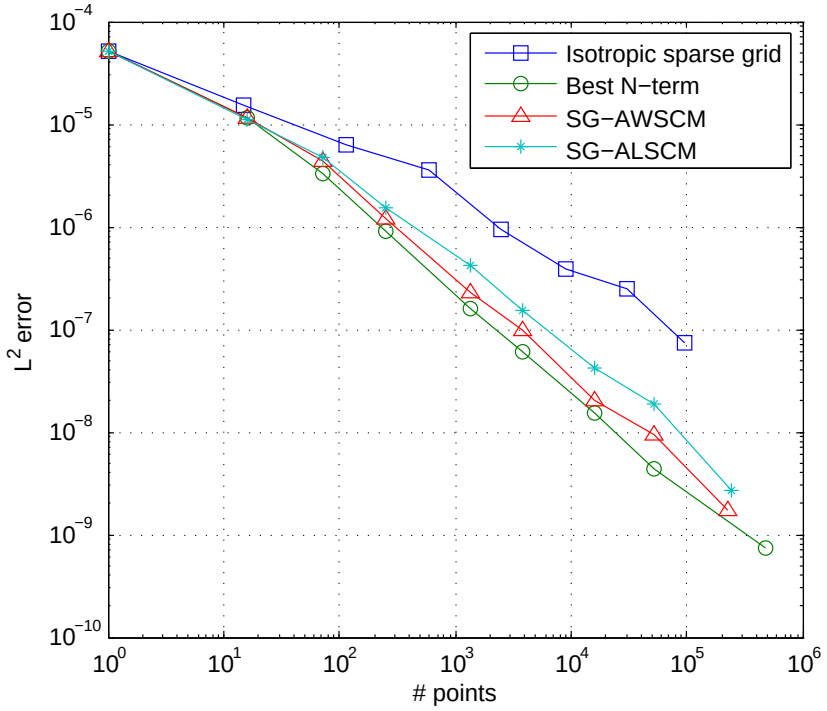

Figure 11: The convergence rate of isotropic sparse grid, the sg-ALSCM, and the sg-AWSCM approximations with tolerance $\varepsilon = 10^{-5}$.

6 CONCLUSIONS

This work proposed a novel sparse-grid adaptive wavelet stochastic collocation method for both smooth and irregular solutions of partial differential equations with random input data. This method can be viewed as a major improvement to previous works; isotropic and anisotropic global Lagrange-type stochastic collocation based on tensor product approximations [2] or sparse grid approaches [39, 39, 51], as well as the hierarchical sparse-grid locally adaptive linear stochastic collocation method [32, 33].

The new technique consists of any standard deterministic approximation in the physical space (e.g. Galerkin finite element) and an adaptive collocation in the probability domain at sparse-grid points in the random parameter space, along with a hierarchical multi-dimensional multi-resolution linear wavelet basis. This compactly supported Riesz basis guarantees the stability of the multi-scale coefficients and leads to optimal hierarchical sparse grid approximations. That is, we are able to guide adaptive refinement by the magnitude of the wavelet coefficient which results in a minimal number of grid points to achieve a prescribed tolerance. This alleviates the curse of dimensionality by reducing the computational complexity for problems having high stochastic dimension. Moreover, as a consequence of the interpolation property, guaranteed by the proposed lifting scheme, our approach remains completely non-intrusive and naturally allow for the solution of uncoupled deterministic problems that are trivially parallelizable, as for the Monte Carlo method.

The numerical examples included in this work provide computational verification of the optimality of our novel algorithm. The numerical results compare our new approach with several classical and heavily utilized techniques for solving stochastic problems whose solutions are both highly regular and even non-smooth with respect to the random variables. The results show that, in particular, for moderately large-dimensional problems, the sparse grid adaptive wavelet stochastic collocation approach seems to be very efficient and superior to all methods it is compared to.

Future directions of this research will include the a full convergence analysis of our new approach that will incorpoorate an examination of the complexity of our algorithm with respect to the number of collocation points on the sparse grid, as the dimension of the problem increases. However, as the computational results suggest, we also want to use the theoretical results to fully explain the optimality of this techniques when compared to previous methods. Finally, we want to avoid solving an interpolating matrix equation for the wavelet coefficients and intend to develop fast algorithms for calculating the wavelet coefficients in sparse tensor product spaces.

REFERENCES

[1] I. M. BABUŠKA, R. TEMPONE, AND G. E. ZOURARIS, *Galerkin finite element approximations of stochastic elliptic partial differential equations*, SIAM J. Numer. Anal., 42 (2004), pp. 800–825 (electronic). 2

[2] I. BABUŠKA, F. NOBILE, AND R. TEMPONE, *A Stochastic Collocation Method for Elliptic Partial Differential Equations with Random Input Data*, SIAM Journal on Numerical Analysis, 45 (2007), pp. 1005–1034. 2, 5, 6, 10, 11, 29

[3] I. M. BABUŠKA, R. TEMPONE, AND G. E. ZOURARIS, *Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 1251–1294. 2

[4] J. BECK, F. NOBILE, L. TAMELLINI, AND R. TEMPONE, *Stochastic spectral galerkin and collocation methods for PDEs with random coefficients: a numerical comparison*, Lecture Notes in Computational Science and Engineering, 76 (2011), pp. 43–62. 11

[5] H.-J. BUNGARTZ AND M. GRIEBEL, *Sparse grids*, Acta Numerica, 13 (2004), pp. 1–123. 11, 15, 17

[6] C. CHUI AND J. WANG, *A general framework of compactly supported splines and wavelets*, Journal of Approximation Theory, 71 (1992), pp. 263 – 304. 3

[7] A. COHEN, W. DAHMEN, AND R. DEVORE, *Adaptive wavelet methods for elliptic operator equations – convergence rates*, Math. Comp., 70 (2001), pp. 27–75. 4

[8] A. COHEN, W. DAHMEN, AND R. DEVORE, *Adaptive wavelet methods for elliptic operator equations II – beyond the elliptic case*, Found. Comput. Math., 2 (2002), pp. 203–245. 4

[9] A. COHEN, I. DAUBECHIES, AND J. FEAUVEAU, *Biorthogonal Bases of Compactly Supported Wavelets*, Communications on Pure and Applied Mathematics, 45 (1992), pp. 485–560. 3

[10] S. DAHLKE, W. DAHMEN, AND K. URBAN, *Adaptive wavelet methods for saddle point problems – optimal convergence rates*, SIAM J. Numer. Anal., 40 (2002), pp. 1230–1262. 4

[11] W. DAHMEN AND A. KUNOTH, *Adaptive wavelet methods for linear-quadratic elliptic control problems: Convergence rates*, SIAM J. Contr. Optim., 43 (2002), pp. 1640–1675. 4

[12] I. DAUBECHIES, *Orthonormal bases of compactly supported wavelets*, Communications on Pure and Applied Mathematics, 41 (1988), pp. 909–996. 3, 19

[13] I. DAUBECHIES, *Wavelets - Algorithms and Applications*, Science, 262 (1993), pp. 1589–1591. 3, 17

[14] D. DIAZ, M. GUNZBURGER, AND A. KUNOTH, *An adaptive wavelet viscosity method for hyperbolic conservation laws*, Numer. Math., 24 (2008), pp. 1388–1404. 4

[15] T. J. DIJKEMA, C. SCHWAB, AND R. STEVENSON, *An adaptive wavelet method for solving high-dimensional elliptic PDEs*, Constr Approx, 30 (2009), pp. 423–455. 4

[16] H. ELMAN AND C. MILLER, *Stochastic collocation with kernel density estimation*, tech. rep., Department of Computer Science, University of Maryland, 2011. 15

[17] G. FISHMAN, *Monte Carlo*, Springer Series in Operations Research, Springer-Verlag, New York, 1996. Concepts, algorithms, and applications. 2

[18] J. FOO, X. WAN, AND G. KARNIADAKIS, *The multi-element probabilistic collocation method (ME-PCM): Error analysis and applications*, Journal of Computational Physics, 227 (2008), pp. 9572–9595. 3

[19] P. FRAUENFELDER, C. SCHWAB, AND R. A. TODOR, *Finite elements for elliptic problems with stochastic coefficients*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 205–228. 2

[20] T. GERSTNER AND M. GRIEBEL, *Dimension-adaptive tensor-product quadrature*, Computing, 71 (2003), pp. 65–87. 3

[21] R. G. GHANEM AND P. D. SPANOS, *Stochastic finite elements: a spectral approach*, Springer-Verlag, New York, 1991. 2

[22] M. GRIEBEL, *Adaptive sparse grid multilevel methods for elliptic PDEs based on finite differences*, Computing, 61 (1998), pp. 151–179. 3, 11

[23] M. GUNZBURGER AND A. KUNOTH, *Space-time adaptive wavelet methods for optimal control problems constrained by parabolic evolution equations*, Siam J Control Optim, 49 (2011), pp. 1150–1170. 3, 4

[24] J. D. JAKEMAN, R. ARCHIBALD, AND D. XIU, *Characterization of discontinuities in high-dimensional stochastic problems on adaptive sparse grids*, Journal of Computational Physics, 230 (2011), pp. 3977–3997. 15

[25] A. KLIMKE AND B. WOHLMUTH, *Algorithm 847: Spinterp: piecewise multilinear hierarchical sparse grid interpolation in matlab*, ACM Transactions on Mathematical Software (TOMS), 31 (2005), pp. 561–579. 3, 15

[26] O. P. LE MAÎTRE, O. M. KNIO, H. N. NAJM, AND R. G. GHANEM, *Uncertainty propagation using Wiener-Haar expansions*, J. Comput. Phys., 197 (2004), pp. 28–57. 3, 17

[27] O. P. LE MAÎTRE, H. N. NAJM, R. G. GHANEM, AND O. M. KNIO, *Multi-resolution analysis of Wiener-type uncertainty propagation schemes*, J. Comput. Phys., 197 (2004), pp. 502–531. 3

[28] O. P. LE MATRE AND O. M. KNIO, *Spectral Methods for Uncertainty Quantification*, Springer, 2010. 17

[29] C. F. LI, Y. T. FENG, D. R. J. OWEN, D. F. LI, AND I. M. DAVIS, *A Fourier-Karhunen-Loève discretization scheme for stationary random material properties in SFEM*, Int. J. Numer. Meth. Engng, www.interscience.wiley.com (2007). 2

[30] M. LOÈVE, *Probability theory. I*, Springer-Verlag, New York, fourth ed., 1977. Graduate Texts in Mathematics, Vol. 45. 2, 6

[31] M. LOÈVE, *Probability theory. II*, Springer-Verlag, New York, fourth ed., 1978. Graduate Texts in Mathematics, Vol. 46. 2, 6

[32] X. MA AND N. ZABARAS, *An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations*, Journal of Computational Physics, 228 (2009), pp. 3084–3113. 3, 11, 15, 17, 29

[33] X. MA AND N. ZABARAS, *An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations*, Journal of Computational Physics, 229 (2010), pp. 3884–3915. 3, 11, 15, 17, 29

[34] L. MATHELIN, M. Y. HUSSAINI, AND T. A. ZANG, *Stochastic approaches to uncertainty quantification in CFD simulations*, Numer. Algorithms, 38 (2005), pp. 209–236. 2

[35] H. G. MATTHIES AND A. KEESE, *Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 1295–1331. 2

[36] P. NITSCHE, *Sparse Approximation of Singularity Functions*, Constructive Approximation, 21 (2005), pp. 63–81. 4

[37] F. NOBILE AND R. TEMPONE, *Analysis and implementation issues for the numerical approximation of parabolic equations with random coefficients*, Int. J. Numer. Meth. Engng, 80 (2009), pp. 979–1006. 8

[38] F. NOBILE, R. TEMPONE, AND C. G. WEBSTER, *An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM J. Numer. Anal., 46 (2008), pp. 2411–2442. 2, 5, 6, 10, 11, 17, 19, 26

[39] F. NOBILE, R. TEMPONE, AND C. G. WEBSTER, *A sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM J. Numer. Anal., 46 (2008), pp. 2309–2345. 2, 5, 6, 10, 11, 17, 20, 26, 29

[40] B. ØKSENDAL, *Stochastic differential equations*, Universitext, Springer-Verlag, Berlin, sixth ed., 2003. An introduction with applications. 8

[41] P. OSWALD, *Hierarchical conforming finite-element methods for the biharmonic equation*, Siam J Numer Anal, 29 (1992), pp. 1610–1625. 16

[42] C. SCHWAB AND R. STEVENSON, *Adaptive Wavelet Algorithms for Elliptic PDE's On Product Domains*, Mathematics of Computation, 77 (2008), pp. 71–92. 4

[43] C. SCHWAB AND R. STEVENSON, *Fast evaluation of nonlinear functionals of tensor product wavelet expansions*, Numerische Mathematik, 119 (2011), pp. 765–786. 4

[44] C. SCHWAB AND S. TOKAREVA, *High order approximation of probabilistic shock profiles in hyperbolic conservation laws with uncertain initial data*, tech. rep., SAM Research Report No. 2011-53, 2011. 24, 25

[45] S. SMOLYAK, *Quadrature and interpolation formulas for tensor products of certain classes of functions*, Dokl. Akad. Nauk SSSR, 4 (1963), pp. 240–243. 11

[46] W. SWELDENS, *The lifting scheme: A custom-design construction of biorthogonal wavelets*, Applied and Computational Harmonic Analysis, 3 (1996), pp. 186–200. 4, 18, 19

[47] W. SWELDENS, *The Lifting Scheme: A Construction of Second Generation Wavelets*, SIAM Journal on Mathematical Analysis, 29 (1998), pp. 511–546. 4, 18, 19

[48] W. SWELDENS AND P. SCHRODER, *Building your own wavelets at home*, Computer, 90 (2000), pp. 72–107. 4, 19

[49] R. A. TODOR AND C. SCHWAB, *Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients*, IMA Journal of Numerical Analysis, 27 (2006), pp. 232–261. 11

[50] N. WIENER, *The homogeneous chaos*, Amer. J. Math., 60 (1938), pp. 897–936. 2

[51] D. XIU AND J. HESTHAVEN, *High-order collocation methods for differential equations with random inputs*, Siam J Sci Comput, 27 (2005), pp. 1118–1139. 2, 29

[52] D. XIU AND G. E. KARNIADAKIS, *The Wiener-Askey polynomial chaos for stochastic differential equations*, SIAM J. Sci. Comput., 24 (2002), pp. 619–644 (electronic). 2, 17

[53] G. ZHANG AND M. GUNZBURGER, *Error analysis of a stochastic collocation method for parabolic partial differential equations with random input data*, SIAM Journal on Numerical Analysis, 50 (2012), pp. 1922–1940. 5, 8


OAK RIDGE NATIONAL LABORATORY
MANAGED BY UT-BATTELLE FOR THE DEPARTMENT OF ENERGY