

SPARSE MULTIREOLUTION REGRESSION FOR UNCERTAINTY PROPAGATION

Daniele Schiavazzi,^{1,} Alireza Doostan,² & Gianluca Iaccarino³*

¹*Mechanical and Aerospace Engineering Department, University of California, San Diego, California 92093, USA*

²*Aerospace Engineering Sciences Department, University of Colorado, Boulder, Colorado 80309-0429, USA*

³*Department of Mechanical Engineering, Stanford University, Stanford, California 94305, USA*

Original Manuscript Submitted: 10/09/2013; Final Draft Received: 01/27/2014

The present work proposes a novel nonintrusive, i.e., sampling-based, framework for approximating stochastic solutions of interest admitting sparse multiresolution expansions. The coefficients of such expansions are computed via greedy approximation techniques that require a number of solution realizations smaller than the cardinality of the multiresolution basis. The effect of various random sampling strategies is investigated. The proposed methodology is verified on a number of benchmark problems involving nonsmooth stochastic responses, and is applied to quantifying the efficiency of a passive vibration control system operating under uncertainty.

KEY WORDS: *uncertainty quantification, multiresolution approximation, compressive sampling, adaptive importance sampling, tree-based orthogonal matching pursuit, uncertain tuned mass damper*

1. INTRODUCTION

The possibilities offered by computational tools have dramatically affected the way modern design of engineering systems is approached and have significantly improved our understanding of how physical systems behave in response to changes in their underlying parameters. With the constant increase of available computer resources, the effect of the uncertainties associated with these parameters (aleatory uncertainties) as well as in the definition of the physical/mathematical models (epistemic uncertainty) can also be quantified to improve the *robustness* of our predictions.

In this context, a typical *forward* Uncertainty Quantification (UQ) process consists of various steps, e.g., formulation of the model/problem, representation of uncertainty sources, propagation of uncertainties to response Quantities of Interest (QoI), and possibly ranking of uncertainty sources. Depending on the application of interest, these tasks may be challenging and require careful investigation and efficient numerical treatment. In the present study, we focus exclusively on the uncertainty propagation step, for which we adopt a probabilistic framework. We, therefore, assume that the formulation of the physical model is completely defined in space and time, and the probability distribution functions of all input random variables or processes are specified. We are then interested in proposing an efficient methodology to compute the stochastic description of a response QoI.

Monte Carlo (MC) sampling has long been used for uncertainty propagation due to its straightforward implementation and robustness in computing statistics of general functionals. While various sampling strategies, e.g., stratified or importance sampling, have been proposed over the years to improve its convergence rate, MC sampling still maintains its appeal for general problems with high-dimensional random inputs.

*Correspond to Daniele Schiavazzi, E-mail: dschiavazzi@ucsd.edu

For problems with *smooth* responses and with moderate number of random inputs, approaches based on expansions in series of basis functions may lead to improved convergence. A description in terms of polynomial basis orthogonal with respect to the probability measure of random inputs—referred to as Polynomial Chaos (PC)—has been proposed in [1] as a generalization of the classical Wiener-Hermite chaos polynomials [2–4]. For sufficiently smooth stochastic responses, these expansion techniques may achieve as high as exponential convergence rate for increasing polynomial order [5–7]. Several numerical strategies, including stochastic Galerkin [1, 4, 5] and stochastic collocation—based either on tensorial product of one-dimensional quadrature rules or sparse grids [8–10]—have been developed to generate PC expansions of solutions to partial differential equations (PDEs) with random inputs.

Of particular interest in the present study is the situation where the QoI exhibits discontinuities or sharp gradients with respect to random inputs, i.e., nonsmooth QoI. Such cases are commonly observed in problems where the QoIs experience sharp variations in the physical space or time due to, for instance, the presence of shocks, bifurcations, or instabilities [11–13]. In these cases, approaches based on standard PC expansions are known to suffer from slow or no convergence [11].

To tackle this shortcoming, various methodologies have been proposed. An expansion in terms of a Wiener-Haar basis is proposed in [11], resulting in faster convergence rates (error in expectation vs. number of solution realizations) for problems with discontinuities. In [12], a multiresolution approximation is formulated based on the Alpert multiwavelet basis [14–17] generated from a family of Legendre scaling functions (also referred to as *father* wavelets); adaptive strategies based on multiwavelet coefficients are also discussed in order to refine the approximation. Multi-element PC techniques, developed in [18, 19], are shown to be effective in approximating stochastic solutions with discontinuities and sharp gradients by adaptively partitioning the stochastic space into nonoverlapping elements. Rational polynomial expansions are investigated in [20]. In [21] an adaptive sparse grid collocation strategy is developed using piecewise multilinear hierarchical basis functions. Methods employing simplicial meshes have also been extended from deterministic to stochastic problems. In [22], a simplex stochastic collocation method is proposed, which is suitable for stochastic spaces of arbitrary shapes—not necessarily obtained as a product of one-dimensional intervals. Methods based on a finite volume discretization in the stochastic space have also been proposed in [23, 24].

The present study proposes a framework for sampling-based approximation of stochastic solutions exhibiting discontinuities or sharp gradients. We take advantage of the *sparsity* observed in multiwavelet expansions of piecewise-smooth responses in order to reduce the number of solution realizations needed for an accurate reconstruction via regression. By sparsity, we refer to the cases in which only a small fraction of multiwavelet basis functions are needed to represent the solution within the required accuracy. Motivated by the work in [25–30], the expansion coefficients are computed using tools extended from the compressive sampling framework [31–33], wherein assumptions on the sparsity of the expansion and random sampling of solution are translated into algorithms and convergence guarantees. To enhance the approximation accuracy, an adaptive importance sampling strategy is also introduced that relies on approximate multiwavelet coefficients computed at each iteration. The combination of these three elements, i.e., multiresolution approximations, sparse regression, and importance sampling, applied in the context of uncertainty propagation, makes the present approach innovative and effective for a broad range of problems.

We highlight some similarities to existing approaches. In particular, a similar expansion was employed in [12] in the context of intrusive uncertainty propagation. The construction of the multiwavelet basis proposed therein differs slightly from that originally discussed by Alpert [14–17], as fewer conditions are required in terms of vanishing statistical moments for the basis functions. Concepts like sparse recovery and importance sampling, closely related to the nonintrusive construction of this work, have not been discussed in [12]. Furthermore, it is worth highlighting the similarity between the concept of hierarchical surplus proposed in [21] and that of the multiwavelet basis used in the present work. In both approaches the information related to *details* is used to drive the adaptivity either in terms of samples and/or approximation basis.

The remaining of this paper is organized as follows. We introduce the problem of interest in Section 2. In Section 3, we briefly review the concepts of multiresolution analysis and multiwavelet approximation. The fundamentals of sparse approximation via compressive sampling are discussed in Section 4 with a special attention to multiwavelet expansions. The proposed adaptive importance sampling strategy for nonintrusive multiwavelet regression is discussed in Section 5. Numerical examples are provided in Section 6 showing the performance of the proposed framework on problems exhibiting discontinuities or sharp gradients with respect to random inputs.

2. PROBLEM SETUP

Consider a complete probability space $(\Omega, \mathcal{F}, \mathcal{P})$ in which Ω is the set of elementary events, \mathcal{F} is the σ algebra of possible events, and \mathcal{P} denotes a probability measure on \mathcal{F} . We assume that the input uncertainty is characterized by the independent random vector $\mathbf{y} = (y_1, \dots, y_d)$, $d \in \mathbb{N}$, where each random variable $y_i : \Omega \rightarrow [0, 1]$ is uniformly distributed over $[0, 1]$. Let $\Gamma \subset \mathbb{R}^D$, $D \in \mathbb{N}$, be the spatial domain with boundary $\partial\Gamma$ and $t \in [0, T]$ represent the temporal variable. We consider approximating the solution $\mathbf{u}(\mathbf{x}, t, \mathbf{y}) : \Gamma \times [0, T] \times \Omega \rightarrow \mathbb{R}^h$, $h \in \mathbb{N}$, to the problem

$$\begin{aligned} \mathcal{L}(\mathbf{x}, t, \mathbf{y}; \mathbf{u}) &= \mathbf{f}(\mathbf{x}, t, \mathbf{y}) \quad \text{on } \Gamma, \\ \mathbf{u}(\mathbf{x}, t, \mathbf{y}) &= \mathbf{u}_b(\mathbf{x}, t, \mathbf{y}) \quad \text{on } \partial\Gamma, \\ \mathbf{u}(\mathbf{x}, t, \mathbf{y}) &= \mathbf{u}_0(\mathbf{x}, \mathbf{y}) \quad \text{at } t = 0, \end{aligned} \quad (1)$$

which holds \mathcal{P} -a.s. in Ω . Here, we assume the well-posedness (in \mathcal{P} -a.s. sense) of (1) with respect to the choices of the forcing, boundary, and initial functions \mathbf{f} , \mathbf{u}_b , and \mathbf{u}_0 , respectively. We seek to approximate $\mathbf{u}(\mathbf{x}, t, \mathbf{y})$ at a fixed location $\mathbf{x}^* \in \Gamma$ in space and time instance $t^* \in [0, T]$ by using realizations $\{\mathbf{u}(\mathbf{x}^*, t^*, \mathbf{y}^{(i)}) : i = 1, \dots, M\}$ of $\mathbf{u}(\mathbf{x}^*, t^*, \mathbf{y})$ corresponding to M random samples $\{\mathbf{y}^{(i)} : i = 1, \dots, M\}$ of \mathbf{y} . To simplify the notation and presentation, we henceforth drop the space and time variables \mathbf{x}^* , t^* , and describe our approach for a scalar, multivariate solution $u(\mathbf{y})$, i.e., with $h = 1$.

It is worth highlighting that the framework introduced here is applicable to cases in which the solution of interest is defined over multiple spatial locations or time instances, or is vector-valued, i.e., $h > 1$. In these cases, a direct—but not necessarily the most efficient—approach is to independently repeat the proposed constructions for each and every spatial location, time instance, or component of the solution of interest.

3. MULTIREOLUTION AND MULTIWAVELETS

3.1 Multiresolution Analysis

A *multiresolution* approximation of $\mathbf{L}^2([0, 1])$ is expressed by means of a nested sequence of closed subspaces $\mathbf{V}_0 \subset \mathbf{V}_1 \subset \dots \subset \mathbf{V}_j \subset \dots \subset \mathbf{L}^2([0, 1])$, where each subspace \mathbf{V}_j corresponding to *resolution* j is given by $\mathbf{V}_j = \text{span}\{\phi_{j,k}(y) : k = 0, \dots, 2^j - 1\}$. Here, the functions

$$\phi_{j,k}(y) = 2^{j/2} \phi(2^j y - k)$$

are dilations and translations of a *scaling* function $\phi(y) : [0, 1] \rightarrow \mathbb{R}$, and $\phi(y)$ is such that the closure of the union of \mathbf{V}_j , $\overline{\bigcup_{k=0}^{\infty} \mathbf{V}_k}$, is dense in $\mathbf{L}^2([0, 1])$. Let the *wavelet* subspace \mathbf{W}_j denote the orthogonal complement of \mathbf{V}_j in \mathbf{V}_{j+1} , that is $\mathbf{V}_{j+1} = \mathbf{V}_j \oplus \mathbf{W}_j$ and $\mathbf{V}_j \perp \mathbf{W}_j$. It can be shown that $\mathbf{W}_j = \text{span}\{\varphi_{j,k}(y) : k = 0, \dots, 2^j - 1\}$ where $\varphi_{j,k}(y)$ is generated from dilation and translation of a *mother* wavelet function $\varphi(y) : [0, 1] \rightarrow \mathbb{R}$,

$$\varphi_{j,k}(y) = 2^{j/2} \varphi(2^j y - k).$$

By the construction of wavelet subspaces \mathbf{W}_j , it is straightforward to see that $\mathbf{V}_j = \mathbf{V}_0 \oplus \left(\bigoplus_{k=0}^{j-1} \mathbf{W}_k \right)$, and consequently $\mathbf{V}_0 \oplus \left(\bigoplus_{k=0}^{\infty} \mathbf{W}_k \right) = \mathbf{L}^2([0, 1])$. Therefore, any function $u(y) \in \mathbf{L}^2([0, 1])$ admits an orthogonal decomposition of the form

$$u(y) = \tilde{\alpha}_{0,0} \phi_{0,0}(y) + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \alpha_{j,k} \varphi_{j,k}(y), \quad (2)$$

where $\tilde{\alpha}_{0,0} = \langle u, \phi_{0,0} \rangle_{\mathbf{L}^2([0,1])}$, $\alpha_{j,k} = \langle u, \varphi_{j,k} \rangle_{\mathbf{L}^2([0,1])}$, and the inner product $\langle u, v \rangle_{\mathbf{L}^2([0,1])} = \int_0^1 u(y) v(y) dy$. For the interest of notation, we rewrite (2) in the form

$$u(y) = \sum_{i=1}^{\infty} \alpha_i \psi_i(y), \quad (3)$$

in which we establish a one-to-one correspondence between elements of the basis sets $\{\phi_{0,0}, \phi_{j,k} : k = 0, \dots, 2^j - 1, j = 0, \dots, \infty\}$ in (2) and $\{\psi_i : i = 1, \dots, \infty\}$ in (3).

3.2 Multiwavelet Basis in One Dimension

In the present study, we adopt the multiresolution basis of Alpert [14–17], where multiple scaling functions $\{\phi_i(y) : i = 0, \dots, m-1\}$ are used to construct the polynomial space \mathbf{V}_0^m . Specifically, we choose $\phi_i(y)$ as the Legendre polynomial of degree i defined on $[0, 1]$. An orthonormal piecewise polynomial basis $\{\varphi_i(y) \in \mathbf{L}^2([0, 1]) : i = 0, \dots, m-1\}$ of \mathbf{W}_0^m is also established such that

$$\int_0^1 \varphi_i(y) y^l dy = 0, \quad i, l = 0, \dots, m-1.$$

Stated differently, $\varphi_i(y)$ has m vanishing moments. The *multiwavelet* basis functions $\varphi_{j,i,k}(y)$, hence the multiwavelet subspaces \mathbf{W}_j^m , are generated by dilations and translations of $\{\varphi_i(y) : i = 0, \dots, m-1\}$, that is

$$\varphi_{j,i,k}(y) = 2^{j/2} \varphi_i(2^j y - k), \quad i = 0, \dots, m-1, k = 0, \dots, 2^j - 1. \quad (4)$$

With certain additional constraints on $\varphi_i(y)$ described in [15], the resulting basis is unique (up to the sign) and provides a generalization of the Legendre and Haar basis. In particular, the Legendre polynomials may be obtained by limiting (4) to the resolution $j = 0$, while Haar wavelets are obtained by setting $m = 1$. Example of functions in $\varphi_i(y)$ for $m = 3$ and $m = 4$ are illustrated in Fig. 1. We refer the interested reader to [15, Section 1.1] for an in-depth derivation of the Alpert multiwavelet basis.

3.3 Multiwavelet Basis in Multiple Dimensions

For $d > 1$, we consider the vector $\mathbf{m} = (m_1, \dots, m_d) \in \mathbb{N}^d$ and introduce $\mathbf{V}_0^{\mathbf{m}} = \mathbf{V}_0^{m_1} \otimes \dots \otimes \mathbf{V}_0^{m_d}$ as the product space spanned by the tensorizations of univariate Legendre polynomials defined on $[0, 1]$. We also introduce a multi-index $\mathbf{i} \in \mathcal{I} = \{(i_1, \dots, i_d) : 0 \leq i_l < m_l, l = 1, \dots, d\}$, and define a single d -dimensional scaling function

$$\phi_{\mathbf{i}}(\mathbf{y}) = \phi_{i_1}(y_1) \dots \phi_{i_d}(y_d).$$

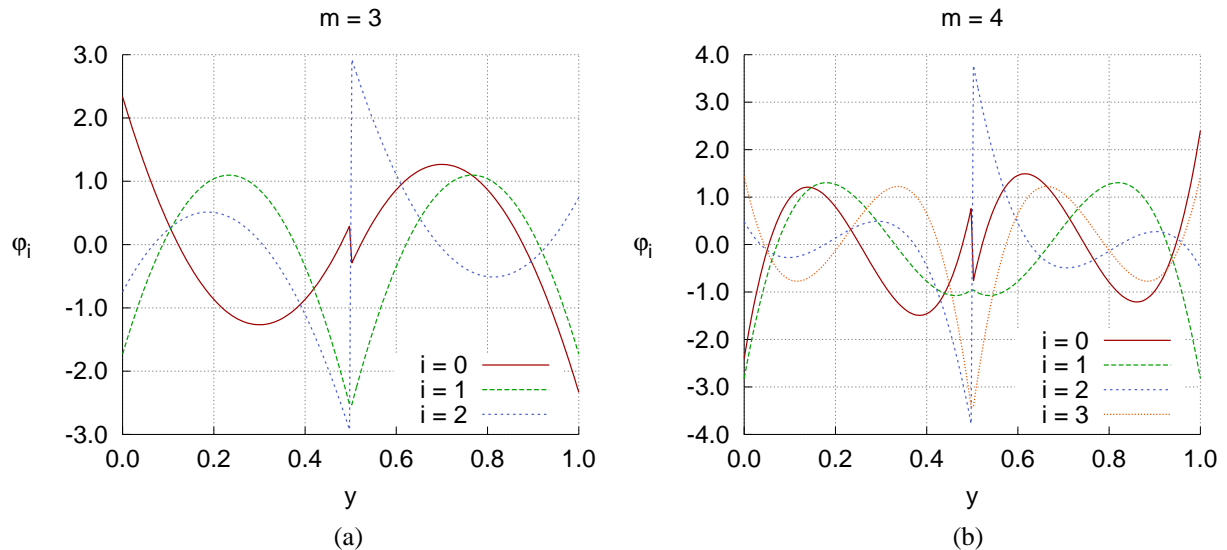


FIG. 1: Examples of Alpert multiwavelets with $m = 3$, (a), and $m = 4$, (b).

The complete set $\mathcal{S}_0^{\mathbf{m}}$ of scaling functions is given by

$$\mathcal{S}_0^{\mathbf{m}} = \{\phi_{\mathbf{i}}(\mathbf{y}) : \mathbf{i} \in \mathcal{I}\}.$$

Before writing the expression for the wavelet function, we introduce the following notation

$$\psi_i^0 = \phi_i, \quad \psi_i^1 = \varphi_i, \quad \psi_{\mathbf{i}}^{\mathbf{q}}(\mathbf{y}) = \psi_{i_1}^{q_1}(y_1) \dots \psi_{i_d}^{q_d}(y_d),$$

where \mathbf{q} is an element of the multi-index set $\mathcal{Q} = \{(q_1, \dots, q_d) : q_l \in \{0, 1\}, l = 1, \dots, d\}$. In other words, \mathbf{q} defines the products of univariate multiscaling and multiwavelet functions ϕ_i and ψ_i in $\psi_{\mathbf{i}}^{\mathbf{q}}(\mathbf{y})$. The set $\mathcal{W}_0^{\mathbf{m}}$ of functions spanning $\mathbf{W}_0^{\mathbf{m}}$ —the orthogonal complement of $\mathbf{V}_1^{\mathbf{m}}$ in $\mathbf{V}_1^{\mathbf{m}}$ —is defined as

$$\mathcal{W}_0^{\mathbf{m}} = \{\psi_{\mathbf{i}}^{\mathbf{q}}(\mathbf{y}) : \mathbf{i} \in \mathcal{I}, \mathbf{q} \in \mathcal{Q}\},$$

while

$$\mathcal{V}_1^{\mathbf{m}} = \mathcal{S}_0^{\mathbf{m}} \cup \mathcal{W}_0^{\mathbf{m}} = \{\psi_{\mathbf{i}}^{\mathbf{q}}(\mathbf{y}) : \mathbf{i} \in \mathcal{I}, \mathbf{q} \in \mathcal{Q} \cup \{(0, \dots, 0)\}\}.$$

Similar to the one-dimensional case, scaled and shifted analogues of the mother multiwavelets in $\mathcal{W}_0^{\mathbf{m}}$ are used to generate multiwavelet spaces of increasing resolution. At resolution $j > 0$, the shift multi-index $\mathbf{k} \in \mathcal{K} = \{(k_1, \dots, k_d) : 0 \leq k_l \leq 2^j - 1, l = 1, \dots, d\}$ is used to identify the partitions in a d -dimensional grid with 2^j subdivisions in each dimension. It follows that

$$\psi_{j,\mathbf{i},\mathbf{k}}^{\mathbf{q}}(\mathbf{y}) = 2^{jd/2} \psi_{i_1}^{q_1}(2^j y_1 - k_1) \dots \psi_{i_d}^{q_d}(2^j y_d - k_d)$$

and

$$\mathcal{W}_j^{\mathbf{m}} = \{\psi_{j,\mathbf{i},\mathbf{k}}^{\mathbf{q}}(\mathbf{y}) : \mathbf{i} \in \mathcal{I}, \mathbf{k} \in \mathcal{K}, \mathbf{q} \in \mathcal{Q}\}.$$

Given a vector $\mathbf{m} = (m_1, \dots, m_d)$ defining the maximum order of the univariate scaling functions, $u(\mathbf{y}) \in \mathbf{L}^2([0, 1]^d)$ can be written as

$$u(\mathbf{y}) = \sum_{\mathbf{i} \in \mathcal{I}} \alpha_{0,\mathbf{i},0} \psi_{0,\mathbf{i},0}^0(\mathbf{y}) + \sum_{j=0}^{\infty} \sum_{\mathbf{k} \in \mathcal{K}} \sum_{\mathbf{q} \in \mathcal{Q}} \sum_{\mathbf{i} \in \mathcal{I}} \alpha_{j,\mathbf{i},\mathbf{k}} \psi_{j,\mathbf{i},\mathbf{k}}^{\mathbf{q}}(\mathbf{y}) \quad (5)$$

Finally, we simplify the notation by rewriting (5) in the form

$$u(\mathbf{y}) = \sum_{i=1}^{\infty} \alpha_i \psi_i(\mathbf{y}), \quad (6)$$

in which we establish a one-to-one correspondence between elements of the basis sets $\{\psi_{0,\mathbf{i},0}^0, \psi_{j,\mathbf{i},\mathbf{k}}^{\mathbf{q}} : \mathbf{i} \in \mathcal{I}, \mathbf{k} \in \mathcal{K}, \mathbf{q} \in \mathcal{Q}, j = 0, \dots, \infty\}$ in (5) and $\{\psi_i : i = 1, \dots, \infty\}$ in (6).

As an example, in Fig. 2, we illustrate the multiwavelet approximation of the 2D function

$$u(y_1, y_2) = \begin{cases} -16(y_1 - 1/2)^2 - 16(y_2 - 1/2)^2 + 1, & \text{if } \sqrt{(y_1 - 1/2)^2 + (y_2 - 1/2)^2} \leq 1/4 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

using $\mathbf{m} = (0, 0)$, i.e., Haar basis, $\mathbf{m} = (1, 1)$, and $\mathbf{m} = (2, 2)$, and with the maximum resolution level $j_{\max} = 3$, i.e., 8×8 partitioning of the unit square.

Remark 1 (Implementation). The expression (5) or (6) may be used for efficient reconstruction and compact representation of a sparse stochastic function once the maximum degrees m_l , $l = 1, \dots, d$, the maximum resolution j_{\max} , and the number of input parameters d are provided. Four quantities are sufficient to identify a single member of the multidimensional multiwavelet basis set, namely the resolution j , and the multi-indices \mathbf{i} , \mathbf{k} , and \mathbf{q} . In practice, \mathbf{i} is determined by selecting one of all the possible permutations of the d indices i_l , $0 \leq i_l < m_l$. Similarly, \mathbf{k} is identified by mapping a given member of the set $\{0, \dots, 2^{jd} - 1\}$ into its coordinates in a d -dimensional uniform grid defined over $[0, 1]^d$ with 2^j uniform subdivisions in every dimension.

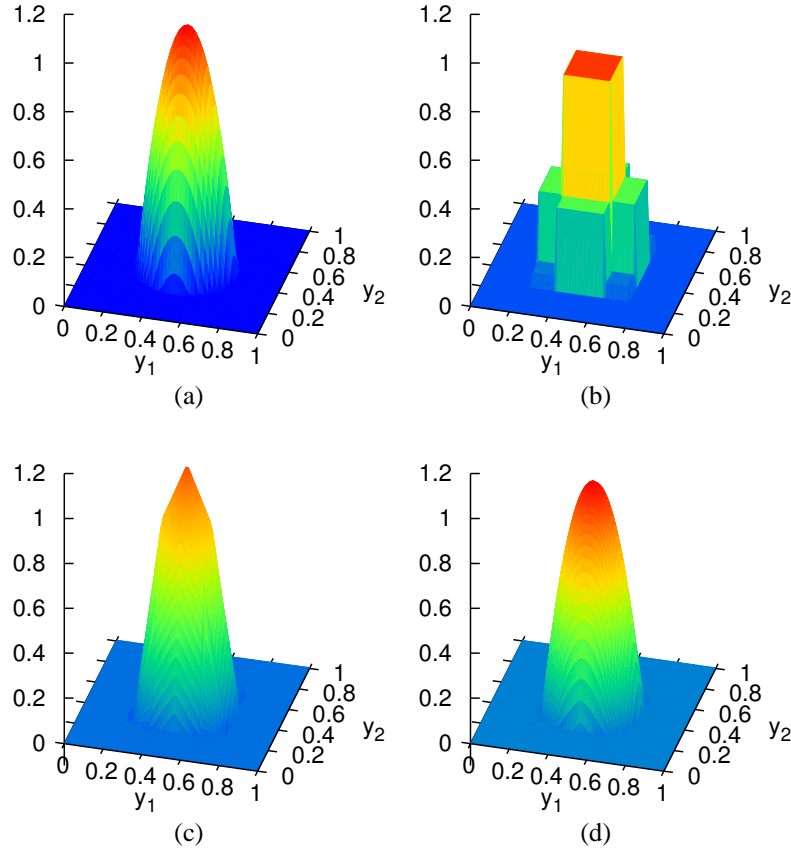


FIG. 2: Examples of 2D multiwavelet approximation for $u(y_1, y_2)$ given in Eq. (7). Original function, (a); Multiwavelet approximation with $\mathbf{m} = (0, 0)$, (b); Multiwavelet approximation with $\mathbf{m} = (1, 1)$, (c); Multiwavelet approximation with $\mathbf{m} = (2, 2)$, (d).

Remark 2 (Nonuniform measures). *In all the developments of this work, we assume that random inputs $y_i, i = 1, \dots, d$, are independent and uniformly distributed over $[0, 1]$. For cases where y_i 's are independent but not uniformly distributed, following [12], we may utilize a change of variable in order to apply the proposed framework. Specifically, let $\Phi(y_i)$ —assumed to be a continuous, monotonically increasing function of y_i —denote the cumulative distribution function of y_i . The cumulative Distribution function (CDF) of random variable $y_i = \Phi^{-1}(z_i)$, where z_i is uniformly distributed over $[0, 1]$, is Φ . This, therefore, suggests the change of variable $y_i \rightarrow \Phi^{-1}(z_i)$ and the construction of the expansion in (6) in terms of z_i instead of y_i .*

3.4 Approximation Error for Differentiable Functions

Theoretical error bounds and convergence rates for multiwavelet approximation are provided in [15]. For the one-dimensional case, consider an m -times differentiable function $u(y) \in C^m([0, 1])$. Using (3), we define $\mathbb{P}_j^m u(y) = \sum_i \langle u, \psi_i \rangle_{L^2([0, 1])} \psi_i(y)$ as the orthogonal projection of u onto \mathbf{V}_j^m . The following error estimate holds [15, Lemma 1.1]:

$$\|\mathbb{P}_j^m u - u\|_{L^2([0, 1])} \leq 2^{-jm} \frac{2}{4^m m!} \sup_{y \in [0, 1]} \left| \frac{d^m u}{dy^m} \right|. \quad (8)$$

The order of convergence is hence m for the one-dimensional case and generalizes to m/d in d dimensions [15]. Notice that the error bound in (8) and its multidimensional extension are only references in our case, as the solutions

of interest in this study are not necessarily m -times differentiable. Additionally, as we shall describe later, we do not compute the multiwavelet coefficients α_i in (3) via the projection $\alpha_i = \langle u, \psi_i \rangle_{\mathbf{L}^2([0,1])}$ as required in (8).

We next introduce the notion of sparsity in multiwavelet expansions and subsequently extend ideas from the field of compressive sampling [31–33], to compute the coefficients α_i in (3) or (6).

4. SPARSE MULTIWAVELET EXPANSION VIA COMPRESSIVE SAMPLING

In the present study, we are interested in approximating solutions u that admit sparse multiwavelet representations. More precisely, consider the multiwavelet expansion (6) corresponding to a given \mathbf{m} and truncated at some resolution j . We denote by P the cardinality of the resulting basis and call u *sparse* in $\{\psi_i, i = 1, \dots, P\}$ if $S = \#\{\alpha_i \neq 0 : i = 1, \dots, P\} \ll P$. In practice, many of the multiwavelet coefficients, while negligible, may not be precisely zero; these cases are considered as *approximately* sparse. We hereafter refer to both cases as sparse. Piecewise smooth functions exhibiting sharp gradients, bifurcations, or discontinuities, for example the solution of hyperbolic PDEs, may lend themselves to sparse multiwavelet representations.

Our interest in sparse multiwavelet representations stems from the recent developments in the field of compressive sampling which may enable exact construction of such representations using a number M of solution realizations that is considerably smaller than the cardinality P of the basis, i.e., $M \ll P$. This is in contrast to, for instance, standard least-squares regression methods which generally require $M \gg P$ for a stable approximation.

4.1 Rudiments of Compressive Sampling

Compressive sampling is a recent development in signal processing that breaks the traditional limits of the Shannon-Nyquist sampling rate for the reconstruction of sparse functions/signals. Consider the vector $\mathbf{u} = (u(\mathbf{y}^{(1)}), \dots, u(\mathbf{y}^{(M)}))^T \in \mathbb{R}^M$ containing realizations of $u(\mathbf{y}) \in \mathbf{L}^2([0, 1]^d)$ corresponding to M independent samples of \mathbf{y} . Assuming that u can be exactly expanded into a multiwavelet basis of the form (6), with some order vector \mathbf{m} and resolution j , \mathbf{u} is given by $\mathbf{u} = \Psi \boldsymbol{\alpha}$. Here, the so-called *measurement* matrix $\Psi \in \mathbb{R}^{M \times P}$ is such that $\Psi[i, j] = \psi_j(\mathbf{y}^{(i)})$ and $\boldsymbol{\alpha} \in \mathbb{R}^P$ is the vector of unknown expansion coefficients. If u , hence $\boldsymbol{\alpha}$, is *sufficiently* sparse, then $\mathbf{u} = \Psi \boldsymbol{\alpha}$ may admit a unique solution $\boldsymbol{\alpha}$. Compressive sampling aims at finding $\boldsymbol{\alpha}$ from an optimization problem

$$(P_s) : \quad \min_{\boldsymbol{\alpha} \in \mathbb{R}^P} \|\boldsymbol{\alpha}\|_s \quad \text{subject to} \quad \mathbf{u} = \Psi \boldsymbol{\alpha}, \quad (9)$$

in which $s = 0$ and the semi-norm $\|\boldsymbol{\alpha}\|_0 = \#\{\alpha_i : \alpha_i \neq 0\}$ is the number of nonzero components of $\boldsymbol{\alpha}$. A more general formulation results from allowing u to be approximately represented by a multiwavelet basis of the form (6),

$$(P_{s,\epsilon}) : \quad \min_{\boldsymbol{\alpha} \in \mathbb{R}^P} \|\boldsymbol{\alpha}\|_s \quad \text{subject to} \quad \|\mathbf{u} - \Psi \boldsymbol{\alpha}\|_2 \leq \epsilon, \quad (10)$$

with $s = 0$. In other words, in (10), we account for a *truncation* error of size $\|\mathbf{u} - \Psi \boldsymbol{\alpha}\|_2 \leq \epsilon$ for some $\epsilon \geq 0$. The sparsest solution $\boldsymbol{\alpha}$ to (P_0) , corresponding to minimizing the $\|\boldsymbol{\alpha}\|_0$, is generally NP-hard to compute [33]. To break this complexity, several heuristic techniques based on greedy pursuit, e.g., Orthogonal Matching Pursuit (OMP), and convex relaxation via ℓ_1 -minimization, i.e., $s = 1$, have been proposed [34–42]. Moreover, several indicators such as the *mutual coherence* [33] or the *restricted isometry constant* [32] have been introduced to establish guarantees on the uniqueness of the solution to (P_0) as well as the ability of the heuristic alternatives in recovering this solution. In particular, the mutual coherence of Ψ (e.g., see [33]) is defined as

$$\mu(\Psi) = \max_{i \neq j} \frac{|\boldsymbol{\psi}_i^T \boldsymbol{\psi}_j|}{\|\boldsymbol{\psi}_i\|_2 \|\boldsymbol{\psi}_j\|_2}, \quad (11)$$

where $\boldsymbol{\psi}_i \in \mathbb{R}^M$ is the i th column of Ψ . Note that $\mu(\Psi) \in [0, 1]$ in general and is strictly positive for $M < P$. Depending on sparsity $S = \|\boldsymbol{\alpha}\|_0$, $\mu(\Psi)$ provides a sufficient condition on the number M of measurements for a successful recovery of $\boldsymbol{\alpha}$ from (P_s) or $(P_{s,\epsilon})$, $s = \{0, 1\}$, as shown in [33]. Ideally, with the same M and S , one may anticipate more accurate solutions $\boldsymbol{\alpha}$ for smaller values of $\mu(\Psi)$.

We next discuss the methodologies we employ to compute the coefficient vector α , together with a brief introduction on a number of closely related approaches for solving (P_s) and $(P_{s,\epsilon})$ when $s = 0$.

4.2 Greedy Approaches

Greedy pursuit algorithms form a major class of techniques for solving (P_0) and $(P_{0,\epsilon})$ with tractable computational cost. Instead of performing an exhaustive search for the support of the sparse solution, these methods successively find one or more columns of Ψ , hence basis functions, that result in largest reduction in the residual, thus the name *greedy*. More specifically, the solution is sought for by repeating a two-stage procedure. In the first step, referred to as *sensing*, one or more indices in the support of α , $\{i : \alpha_i \neq 0\}$, are identified such that the current residual is reduced the most irrespective of already selected columns of Ψ . In a second stage, called *projection*, the coefficients associated with the updated column set are computed. A brief overview of several greedy approaches is presented next.

Matching Pursuit. The Matching Pursuit (MP) algorithm [43] is the basis of many sparse approximation techniques, including those employed in the present study. Let \mathcal{I}_{k-1} be the set of indices associated with the selected columns of Ψ at iteration $k-1 < M$. In MP the sensing and projection steps are closely related. Specifically, assuming the columns of Ψ have unit ℓ_2 norm, an index $i = \arg \max_{j \notin \mathcal{I}_{k-1}} |\mathbf{r}_{k-1}^T \psi_j|$ is first identified and the coefficient α_i is set to $\alpha_i = \mathbf{r}_{k-1}^T \psi_i$. The quantity $\alpha_i \psi_i$ is then subtracted from \mathbf{r}_{k-1} , providing an updated residual for the next iteration. The set of indices is also updated to $\mathcal{I}_k = \mathcal{I}_{k-1} \cup \{i\}$. A discussion on the infinite and finite dimensional properties of this algorithm is provided in [43]. In general, the residual vector produced by MP is only orthogonal to the last selected basis function, which may lead to slow convergence [44].

Orthogonal Matching Pursuit. Orthogonal Matching Pursuit (OMP) is a widely used algorithm for the solution of (P_0) and $(P_{0,\epsilon})$ [37, 44]. It improves on MP by solving a least-squares problem in the projection stage. Similar to MP, an index $i = \arg \max_{j \notin \mathcal{I}_{k-1}} |\mathbf{r}_{k-1}^T \psi_j|$ is identified and added to \mathcal{I}_{k-1} , i.e., $\mathcal{I}_k = \mathcal{I}_{k-1} \cup \{i\}$. Let $\Psi_{\mathcal{I}_k}$ be the submatrix of Ψ consisting of columns of Ψ with indices in \mathcal{I}_k . The entries of solution associated with \mathcal{I}_k , denoted by $\alpha_{\mathcal{I}_k}$, are computed by solving an overdetermined least-squares problem

$$(\Psi_{\mathcal{I}_k}^T \Psi_{\mathcal{I}_k}) \alpha_{\mathcal{I}_k} = \Psi_{\mathcal{I}_k}^T \mathbf{u}. \quad (12)$$

Notice that, unlike in MP, the residual $\mathbf{r}_k = \mathbf{u} - \Psi_{\mathcal{I}_k} \alpha_{\mathcal{I}_k}$ is made orthogonal to all the columns of Ψ with indices in \mathcal{I}_k . More details on the implementation of OMP may be found in Algorithm 1. Following [45], if a solution to (P_0) exists that satisfies $\|\alpha\|_0 \leq (1 + 1/\mu(\Psi))/2$, then OMP will identify it exactly.

CoSaMP [41] and StOMP [35] are greedy heuristics proposed as more efficient alternatives to OMP. Both approaches sense more than one basis function in each iteration, resulting in speed-up and, in some cases, accuracy enhancement over the standard OMP. They also accommodate fast matrix-vector products, both for sensing and for iteratively solving the least-squares problems.

Given the local nature of the employed multiresolution basis and as a result of operating in the undersampled case, it may occur that $\psi_{i^*} = \mathbf{0}$ for some $i^* \in \{1, \dots, P\}$. In this case i^* will not be included in the support set \mathcal{I}_{k-1} irrespective of the magnitude of α_{i^*} . This observation motivates the use of a more careful sampling technique as introduced in Section 5.

Algorithm 1. [Orthogonal Matching Pursuit (OMP)]

Inputs:

- Measurement matrix Ψ .
- Vector of realizations \mathbf{u} .
- Maximum allowable number of iterations k_{\max}
- Convergence tolerance ϵ .

Outputs:

- Solution vector α .

Initialize:

- $k \leftarrow 0, \alpha_0 \leftarrow \mathbf{0}, \mathbf{r}_0 \leftarrow \mathbf{u}, \mathcal{I}_0 \leftarrow \{\emptyset\}.$
- $\mathbf{N} \leftarrow \text{diag}(\|\psi_1\|_2, \dots, \|\psi_P\|_2)$

Set to unit norm columns $\Psi \leftarrow \Psi \mathbf{N}^{-1}$.

Iterate:

while $\|\mathbf{r}_k\|_2 > \epsilon$ and $k < k_{\max}$

 (*Sensing*)

 Find $i = \arg \max_{j \notin \mathcal{I}_k} |\mathbf{r}_k^T \boldsymbol{\psi}_j|$.

 (*Update support set*)

$k \leftarrow k + 1$.

$\mathcal{I}_k \leftarrow \mathcal{I}_{k-1} \cup \{i\}$.

 (*Least-squares projection*)

 Set $\Psi_{\mathcal{I}_k} \leftarrow [\boldsymbol{\psi}_{i_1} | \dots | \boldsymbol{\psi}_{i_k}]$, $i_j \in \mathcal{I}_k$.

 Solve $\boldsymbol{\alpha}_{\mathcal{I}_k} \leftarrow (\Psi_{\mathcal{I}_k}^T \Psi_{\mathcal{I}_k})^{-1} \Psi_{\mathcal{I}_k}^T \mathbf{u}$.

 (*Expand to full coefficient vector*)

$\boldsymbol{\alpha}_k(\mathcal{I}_k) \leftarrow \boldsymbol{\alpha}_{\mathcal{I}_k}$.

 (*Update residual*)

$\mathbf{r}_k \leftarrow \mathbf{u} - \Psi_{\mathcal{I}_k} \boldsymbol{\alpha}_{\mathcal{I}_k}$.

end while

Return:

$\mathbf{N} \boldsymbol{\alpha}_k$

Tree-based Orthogonal Matching Pursuit (TOMP). The observation that piecewise-smooth functions are characterized by a connected subtree representation in the wavelet space has lead to tree-based implementations of OMP, referred to as TOMP, as described in [36, 46, 47]. Basis functions are progressively added to the active set together with their ancestors, in an effort to perpetuate their connected subtree structure. The introduction of such an additional dependency *model* is a way to reduce the number of effective degrees of freedom in the expansion coefficients with the objective of enhancing the reconstruction accuracy. In the context of compressive sampling, other approaches for promoting *a priori* knowledge on the expansion coefficients have been developed, see, e.g., [48–50].

As discussed in Section 3, multiwavelet basis functions are characterized by a resolution index j and a shift multi-index $\mathbf{k} \in \mathcal{K}$. The number of basis functions at successive resolutions increases by a factor of 2^d , making the representation akin to a tree \mathcal{T} representation where every father has 2^d children and every node in the graph is associated with multiple basis functions sharing the same support. If we assume, for simplicity, a one-dimensional case with scalar wavelets, then a basis function ψ_i with scale and shift indices (j, k) is the *father* of basis functions with scale and shift indices $(j+1, 2k)$ and $(j+1, 2k+1)$. The set of indices of *ancestors* of ψ_i is denoted by \mathcal{A}_i and is obtained by recursively including fathers up to the root. Similarly, the set of indices of *descendants* of ψ_i , indicated by \mathcal{D}_i , is formed by including all the children of ψ_i up to the maximum scale index j_{\max} . A tree is *connected* if $i \in \mathcal{T}$ implies $\mathcal{A}_i \in \mathcal{T}$. A node i of \mathcal{T} is a *leaf* if $\mathcal{A}_i \in \mathcal{T}$ while $\mathcal{D}_i \notin \mathcal{T}$.

In the TOMP approach of [46], the sensing step at iteration k leads to a candidate set

$$\mathcal{C}_k = \{i : |\mathbf{r}_{k-1}^T \boldsymbol{\psi}_i| \geq \gamma \max_j |\mathbf{r}_{k-1}^T \boldsymbol{\psi}_j|, j \notin \mathcal{I}_{k-1}\},$$

in which $0 < \gamma \leq 1$ controls the number of indices added to \mathcal{C}_k . For each index $i \in \mathcal{C}_k$ at a time, the tentative support set $\mathcal{I}_k^{(i)} = \mathcal{I}_{k-1} \cup \{i, \mathcal{A}_i\}$ is formed. Subsequently, for each $\mathcal{I}_k^{(i)}$ a least-squares problem of the form (12) is solved and the set $\{i, \mathcal{A}_i\}$ resulting in the smallest residual is permanently added to the support set \mathcal{I}_{k-1} . For a given number of samples M , the iterations are stopped when $|\mathcal{I}_k^{(i)}| > \zeta M$. Empirical results of [36] suggest an optimal value of $\gamma = 0.975$, which we also use in the numerical experiments of the present study. Similar TOMP approaches, as described above, have been proposed in [36, 47]. There, to reduce the computation complexity of searching through all basis functions, only basis functions up to a fixed resolution below a selected basis function are included in the candidate set of [47].

Our implementation of TOMP parallels the approach of [46] with the main difference that, in our construction, multiple basis functions (multiwavelets) are present at the tree nodes. This suggests a *vector* tree node extension of the scalar tree nodes of [46], where, except for the selection of the ancestor sets that may be implemented in different

ways, all the other steps remain practically unchanged. To be more specific, multiple ancestors at resolution $j - 1$ correspond to a selected basis function at resolution $j \geq 1$. In our implementation, we choose to include all the ancestors. Algorithm 2 summarizes the details of the TOMP approach implemented in the present study.

Algorithm 2. [Tree-based Orthogonal Matching Pursuit (TOMP)]

Inputs:

Measurement matrix Ψ .
 Vector of realizations \mathbf{u} .
 Maximum allowable number of iterations k_{\max} .
 Tolerance ϵ .
 Coefficients γ, ζ .

Outputs:

Solution vector α .

Initialize:

$k \leftarrow 0, \mathbf{r}_0 \leftarrow \mathbf{u}, \mathcal{I}_0 \leftarrow \{\emptyset\}$.
 $\mathbf{N} \leftarrow \text{diag}(\|\psi_1\|_2, \dots, \|\psi_P\|_2)$.
 Set to unit norm columns $\Psi \leftarrow \Psi \mathbf{N}^{-1}$.

Iterate:

while $\|\mathbf{r}_k\|_2 > \epsilon$, $k < k_{\max}$, and $|\mathcal{I}_k| < \zeta M$.
 (*Sensing*)
 $\mathcal{C}_k = \{i : |\mathbf{r}_{k-1}^T \psi_i| \geq \gamma \max_j |\mathbf{r}_{k-1}^T \psi_j|, j \notin \mathcal{I}_{k-1}\}$.
 (*Update support set*)
 $k \leftarrow k + 1$.
 for all $i \in \mathcal{C}_k$.
 $\mathcal{I}_k^{(i)} \leftarrow \mathcal{I}_{k-1} \cup \{i, \mathcal{A}_i\}$.
 (*Least-squares projection*)
 $\alpha_{\mathcal{I}_k^{(i)}} \leftarrow (\Psi_{\mathcal{I}_k^{(i)}}^T \Psi_{\mathcal{I}_k^{(i)}})^{-1} \Psi_{\mathcal{I}_k^{(i)}}^T \mathbf{u}$.
 (*Tentative residual*)
 $\mathbf{r}_{\mathcal{I}_k^{(i)}} \leftarrow \mathbf{u} - \Psi_{\mathcal{I}_k^{(i)}} \alpha_{\mathcal{I}_k^{(i)}}$.
 end for
 $i_k \leftarrow \arg \min_{i \in \mathcal{C}_k} \|\mathbf{r}_{\mathcal{I}_k^{(i)}}\|_2$.
 $\mathcal{I}_k \leftarrow \mathcal{I}_k^{(i_k)}$.
 (*Update solution*)
 $\alpha_k(\mathcal{I}_k) \leftarrow \alpha_{\mathcal{I}_k^{(i_k)}}$.
 (*Update residual*)
 $\mathbf{r}_k \leftarrow \mathbf{r}_{\mathcal{I}_k^{(i_k)}}$.
 end while
Return:
 $\mathbf{N} \alpha_k$.

Finally, we remark that the accuracy of our sparse approximation depends on the truncation error ϵ in (10). An excessively small value of ϵ may result in overfitting the solution samples, while larger values of ϵ may lead to underfitting, hence, less accurate reconstructions. In [26, 51] a procedure based on cross validation is proposed to find an optimal value for ϵ . This technique is effective in reducing the over- or underfitting but requires multiple solutions of smaller-size problems, which affects the overall reconstruction time.

5. SAMPLING STRATEGY

The standard approach to obtain the measurements \mathbf{u} is to generate realizations $\mathbf{y}^{(i)}$ of the input independently from their joint probability measure $\rho(\mathbf{y})$, and evaluate the corresponding solution $u(\mathbf{y}^{(i)})$. However, for situations where u

exhibits, for instance, sharp gradients or discontinuities, such a sampling strategy may not necessarily lead to accurate multiwavelet approximations, especially if the number of realizations is limited. This is because the higher resolution basis functions, needed to capture the local structure of u , may not be sampled enough to be selected in the sensing stage of OMP or TOMP.

In [52], the optimality of the Chebyshev sampling is discussed for compressive sampling of a large class of global, bounded orthogonal polynomials. Except at resolution $j = 0$, multiwavelet functions are defined locally over the support of random inputs. As a result, depending on the problem at hand, sampling from the Chebyshev measure may depart from being optimal. To illustrate this empirically, we compare distributions of mutual coherence for measurement matrices generated from Legendre and multiwavelet expansions, by sampling uniformly or according to the Chebyshev distribution. Following that, an importance sampling approach is introduced that allows for local accumulations of samples and consequently enhancement of solution accuracy.

Remark 3 (Departure from standard compressive sampling). *Note that any adaptive sampling of solution is a departure from standard compressive sampling techniques, where the sampling is meant to be nonadaptive. Therefore, the approach presented in this work only overlaps with compressive sampling methods by sharing algorithms for sparse approximation.*

5.1 Sampling from Chebyshev Measure

The Chebyshev probability measure on $(0, 1)$ is defined as $\rho_C(y) = 2.0/(\pi\sqrt{1 - (2y - 1)^2})$. A uniform random variable $z \sim \mathcal{U}(0, 1)$ can be mapped to a random variable y with the Chebyshev measure using $y = (\sin[\pi(z - 0.5)] + 1)/2$. Let $\{L_i(y) : i = 0, \dots, P - 1\}$ denote the set of univariate Legendre polynomials of degree at most $P - 1$ defined over $(0, 1)$. In [53, Corollary 7.4], it is shown that under certain conditions the mutual coherence of the measurement matrix associated with $\{L_i(y)\}$ is bounded by $\mu(\Psi) \leq \sqrt{CM^{-1}K^2 \log(2^{3/4}P^2/\varepsilon)}$, with probability at least $1 - \varepsilon$. Here, $C \approx 26.24$ and K is a constant independent of P but dependent on the choice of sampling $\{L_i(y)\}$. In particular, K attains its minimum, $K = \sqrt{2}$, when $\{L_i(y)\}$ is sampled according to the Chebyshev measure and weighted appropriately [52, 53]. For other choices of sampling, $K > \sqrt{2}$, thus implying the *optimality* of the Chebyshev measure. This is a consequence of the observation that the smallest (uniform) upper bound on $\{L_i(y)\}$ follows the Chebyshev measure [52]. For multidimensional cases, the results in [52] may be extended to show that Chebyshev measure remains optimal as far as $d < P - 1$.

In an effort to extend the above sampling approach to the case of multiresolution basis functions, we note that, at the coarsest resolution and in a one-dimensional setting, $\phi_i(y) = L_i(y)$ and therefore the Chebyshev measure is optimal if we only include the scaling family in the expansion (2). However, the orthonormal mother multiwavelets $\varphi_i(y)$ are discontinuous functions growing unboundedly at $y \in \{0, 1/2, 1\}$ for increasing i . Due to their construction, the multiwavelet functions at higher resolutions and with any shift multi-index \mathbf{k} will also inherit this property from the mother multiwavelets. As a result, the Chebyshev measure does not necessarily provide the tightest envelope that uniformly bounds the multiwavelet functions, which in turn implies that no improvement may be expected using the Chebyshev sampling compared to the uniform counterpart.

To numerically investigate the performance of the Chebyshev sampling, we plot the CDF of the mutual coherence $\mu(\Psi)$ for matrices with Legendre and multiwavelet basis functions sampled uniformly and according to the Chebyshev measure. Legendre matrices are constructed using $d = 5$, $m = 5$, and $P = 252$, while $d = 3$, $j_{\max} = 0$, $m = 2$, and $P = 216$ is used for multiwavelets. The resulting curves are shown in Fig. 3(a) and 3(b). It can be seen that smaller values of $\mu(\Psi)$ are achieved for the Legendre basis by the Chebyshev sampling and for multiwavelet basis by the uniform sampling.

5.2 Importance Sampling

Importance sampling is a widely used variance reduction technique in Monte Carlo estimation [54]. Sampling is performed according to a probability measure different from that of the random inputs, with the purpose of promoting *important* regions of the input space.

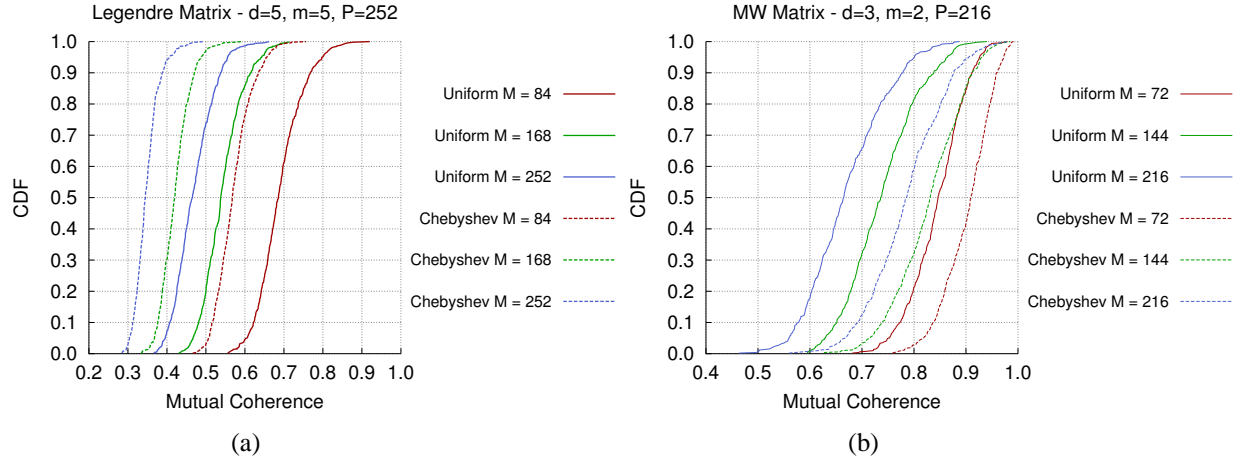


FIG. 3: CDFs of mutual coherence associated with random realizations of Legendre (a) and multiwavelet bases (b).

As stated at the beginning of Section 5, our motivation for developing an importance sampling strategy stems from the observation that, with the uniform sampling, higher resolution multiwavelets may be undersampled and, hence, incorrectly excluded from the active sets of OMP or TOMP. The idea is to concentrate samples at locations where large multiwavelet coefficients are observed, and precondition the basis functions to retain their orthogonality. We will illustrate numerically in Section 6.4 that this importance sampling strategy leads to accuracy enhancement when compared to the uniform sampling.

It must be emphasized that the proposed sampling and sparse regression approach is adaptive. Specifically, we begin by selecting an initial number of realizations M uniformly distributed over the sampling space in order to generate the realizations \mathbf{u} . The greedy solvers OMP and TOMP are then used to provide a *first* estimate for the multiwavelet coefficients. Based on these coefficients, a number of additional input samples and solution realizations are obtained, and the importance sampling measure is formed. This process is repeated until the maximum possible number of solution realizations is reached or a sufficiently small change in the coefficients is observed. For the interest of presentation, in what follows we assume that the approximate coefficients are available and we describe our approach for drawing additional samples of inputs as well as the construction of the importance sampling measure.

5.2.1 Importance-Driven Partitioning of Stochastic Space

We begin by discussing how a partition of the sampling space can be associated with a truncated multiwavelet representation. We then describe how samples are generated over such partition and, finally, how we build an importance sampling measure.

A partition of the stochastic space $[0, 1]^d$ is obtained from a multiwavelet tree representation by first *marking* the nodes in order to form a connected subtree—associated with large coefficients—and by identifying its leaves. Note that, in a multiwavelet representation, multiple basis functions share the same resolution level j and shift multi-index \mathbf{k} . This representation may be organized in a vector tree whose nodes are vectors of coefficients associated with these basis functions. In order to construct our importance-driven partitioning of $[0, 1]^d$, we convert this representation to a scalar tree before we proceed. Stated differently, we desire to assign only one representative coefficient for every node, i.e., pair (j, \mathbf{k}) , in the tree. To this end, we select the coefficient with the largest absolute value. Note that this is not the only possible choice; the ℓ_2 norm of those coefficients, for instance, may be a valid alternative. Our choice is, however, consistent with the assumption of sparsity of the multiwavelet coefficients. In what follows, envelope coefficients will be denoted with $\hat{\alpha}$ to avoid confusion with the coefficients in the multiwavelet expansion (6).

In the first step, all tree nodes with multiwavelet coefficients $\hat{\alpha}_i$ obeying $|\hat{\alpha}_i| > \hat{\alpha}_{tol}$, for some $\hat{\alpha}_{tol} > 0$, are identified and marked. In the second step, a scalar connected subtree $\mathcal{T}_c \subset \mathcal{T}$ is identified where every node with a marked descendant is also marked. Finally, the leaves (L in total) of \mathcal{T}_c —not necessarily at resolution j_{\max} —are

identified as unmarked vertices with marked parents or marked vertices at the resolution $j = j_{\max}$, and added to \mathcal{T}_c .

The support of these leaves form a set of disjoint subdomains $\{\mathcal{B}_i : i = 1, \dots, L\}$ and their union is $[0, 1]^d$, thus resulting in the desired partition of the stochastic space. These properties of $\{\mathcal{B}_i : i = 1, \dots, L\}$ may be verified using arguments based on tree traversal through parent-child edges. In particular, as leaves are selected from a connected tree, the path between an arbitrary leaf and the tree root does not contain other leaves, resulting in disjoint subdomains. Moreover, from the root of the tree, it is not possible to reach an *outside* node $i \notin \mathcal{T}_c$ without visiting a leaf first. The leaves, therefore, cover the entire $[0, 1]^d$. Figure 4 displays the identification of an instance of a subtree \mathcal{T}_c corresponding to $d = 1$, $j_{\max} = 3$, and $L = 4$.

5.2.2 Construction of Importance Sampling Measure

The partition $\{\mathcal{B}_i : i = 1, \dots, L\}$ drives the construction of the proposed importance sampling measure. In particular, let $\hat{\alpha}_i$ be the multiwavelet coefficient with largest magnitude corresponding to leaf node i , and \mathcal{A}_i the set of ancestors of ψ_i . For each leaf node $i = 1, \dots, L$, let

$$\beta_i = \max_{j \in \{i \cup \mathcal{A}_i\}} \frac{|\hat{\alpha}_j|}{|\text{supp}(\psi_j)|}, \quad i = 1, \dots, L, \quad (13)$$

denote the maximum ratio of the coefficient magnitudes to the support size of the corresponding basis functions on a branch with leaf i . We draw \tilde{M}_i additional samples uniformly over each subdomain \mathcal{B}_i such that \tilde{M}_i is proportional to the weights $\beta_i / \sum_{j=1}^L \beta_j$. Let M_i denote the total number of samples on each subdomain \mathcal{B}_i and $M = \sum_{i=1}^L M_i$ the total number of samples. We then define the empirical, importance sampling measure $\gamma(\mathbf{y}) : [0, 1]^d \rightarrow \mathbb{R}_{>0}$,

$$\gamma(\mathbf{y}) = \frac{\sum_{i=1}^L M_i \mathbb{I}_{\mathcal{B}_i}(\mathbf{y})}{\sum_{i=1}^L M_i |\mathcal{B}_i|}, \quad (14)$$

which, as we shall describe later, is needed to precondition the $(P_{0,\epsilon})$ problem in (10). Here, the indicator function $\mathbb{I}_{\mathcal{B}_i}(\mathbf{y}) = 1$ if $\mathbf{y} \in \mathcal{B}_i$ and zero otherwise. Notice that this importance sampling strategy gives higher priority to sampling within the subdomains corresponding to multiwavelets at higher resolutions.

5.2.3 Preconditioning

The importance sampling strategy discussed in Section 5.2.2 may lead to a large mutual coherence $\mu(\Psi)$. This is because the multiwavelet basis is orthogonal with respect to the uniform measure, while samples follow the generally

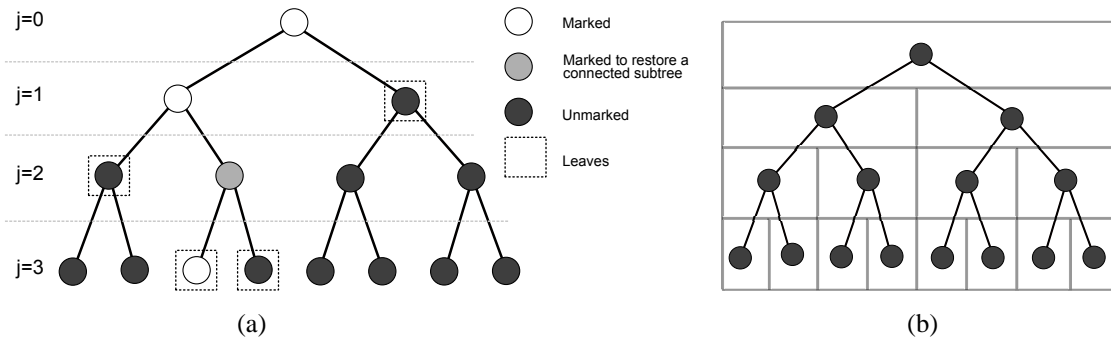


FIG. 4: Identification of leaves on a one-dimensional multiwavelet tree. Vertices associated with significant expansion coefficients are first marked. A connected subtree representation is identified by marking additional vertices. Leaves are finally selected as described in Section 5.2.1, (a). The same scalar tree is also illustrated in plot (b), with every node corresponding to an interval denoting the support of a multiwavelet basis function with largest coefficient.

nonuniform empirical measure $\gamma(\mathbf{y})$. To retain the asymptotic orthogonality of columns in the measurement matrix Ψ , we introduce the weighted basis functions $\hat{\psi}_i(\mathbf{y}) = \psi_i(\mathbf{y})/\sqrt{\gamma(\mathbf{y})}$ for which

$$\frac{1}{M} \sum_{k=1}^M \hat{\psi}_i(\mathbf{y}^{(k)}) \hat{\psi}_j(\mathbf{y}^{(k)}) \xrightarrow{\text{a.s.}} \int_{[0,1]^d} \frac{\psi_i(\mathbf{y})}{\sqrt{\gamma(\mathbf{y})}} \frac{\psi_j(\mathbf{y})}{\sqrt{\gamma(\mathbf{y})}} \gamma(\mathbf{y}) d\mathbf{y} = \delta_{ij} \quad \text{as } M \rightarrow \infty,$$

as a result of the strong law of large numbers. From a practical point of view, the weighting of the multiwavelet basis translates into forming the importance measure $\gamma(\mathbf{y})$ and using the preconditioned measurement matrix $\hat{\Psi} = \mathbf{W}\Psi$ and realization vector $\hat{\mathbf{u}} = \mathbf{W}\mathbf{u}$ in $(P_{0,\epsilon})$. Here, the *preconditioner* matrix \mathbf{W} is given by $\mathbf{W} = \text{diag}(1/\sqrt{\gamma(\mathbf{y}^{(1)})}, \dots, 1/\sqrt{\gamma(\mathbf{y}^{(M)})})$. An overview of the operations performed in the proposed importance sampling framework is outlined in Algorithm 3.

Algorithm 3. [Importance sampling]

Part 1—Identify a partition of $[0, 1]^d$ consistent with the MW representation [Section (5.2.1)].

Form a scalar tree \mathcal{T} using the coefficient $\hat{\alpha}_i$ with maximum absolute value among those associated with the same pair (j, \mathbf{k}) .

Given $\hat{\alpha}_{tol} > 0$, mark nodes with $|\hat{\alpha}_i| > \hat{\alpha}_{tol}$.

Identify a connected subtree \mathcal{T}_c of \mathcal{T} .

Identify the leaves and associated subdomains $\{\mathcal{B}_i, i = 1, \dots, L\}$.

Part 2—Sampling and importance measure computation [Section (5.2.2)]

Determine the coefficients $\beta_i, i = 1, \dots, L$, from (13).

Uniformly draw samples in each \mathcal{B}_i with a number proportional to weights $\beta_i / \sum_{j=1}^L \beta_j$.

Set the empirical measure $\gamma(\mathbf{y})$ from (14).

Part 3—Preconditioning [Section (5.2.3)]

Form the preconditioner matrix $\mathbf{W} = \text{diag}(1/\sqrt{\gamma(\mathbf{y}^{(1)})}, \dots, 1/\sqrt{\gamma(\mathbf{y}^{(M)})})$.

Apply \mathbf{W} to the matrix $\hat{\Psi} = \mathbf{W}\Psi$ and realization vector $\hat{\mathbf{u}} = \mathbf{W}\mathbf{u}$.

6. NUMERICAL TESTS

In the first benchmark problem, we consider two multiwavelet basis functions with overlapping support and show the effect of importance sampling and preconditioning on their empirical correlation. This exercise provides insight on the effect of sampling strategy on the mutual coherence $\mu(\Psi)$ of the measurement matrix Ψ . Following this, a test case is presented in Section 6.2, where two piecewise-smooth functions with different tree representations are approximated by solving $(P_{0,\epsilon})$ via OMP, TOMP, and using uniform and adaptive importance sampling. Moreover, the performance of the proposed multiresolution scheme for uncertainty propagation is assessed through its application to benchmark problems featuring solutions with discontinuities or sharp gradients with respect to random inputs. Specifically, a nonsmooth function, the Kraichnan-Orszag (KO) problem with one and two random initial conditions, and a dynamical system with passive vibration control operating under uncertainty are considered.

6.1 Empirical Correlation of Two Multiwavelet Basis Functions

To achieve a smaller mutual coherence $\mu(\Psi)$ —hence better reconstruction accuracy—we desire faster convergence of the empirical correlation of the multiwavelet basis. To investigate the impact of the sampling strategy on the empirical correlation, we consider two one-dimensional, multiwavelet basis functions $\psi_{j_1, i_1, k_1}^1, \psi_{j_2, i_2, k_2}^1$ at resolutions j_1, j_2 , translated by k_1, k_2 and with degrees i_1, i_2 , respectively. Here, we choose $j_1 = 0, k_1 = 0, i_1 = 2$ and $j_2 = 3, k_2 = 7, i_2 = 2$, resulting in intervals $\mathcal{B}_1 = [0, 7/8)$ and $\mathcal{B}_2 = [7/8, 1]$. An equal expansion coefficient is assumed for the two basis functions. Therefore, the coefficients β_1 and β_2 are inversely proportional to the sizes $|\mathcal{B}_1| = 7/8$ and $|\mathcal{B}_2| = 1/8$, respectively, and consequently $\tilde{M}_1 = 1/7\tilde{M}_2$.

Figure 5 shows the empirical correlation of the unweighted ψ_{j_1, i_1, k_1}^1 and ψ_{j_2, i_2, k_2}^1 with uniform as well as weighted ψ_{j_1, i_1, k_1}^1 and ψ_{j_2, i_2, k_2}^1 with importance sampling. Notice that the empirical correlation converges to zero fastest for the case of weighted ψ_{j_1, i_1, k_1}^1 and ψ_{j_2, i_2, k_2}^1 with importance sampling.

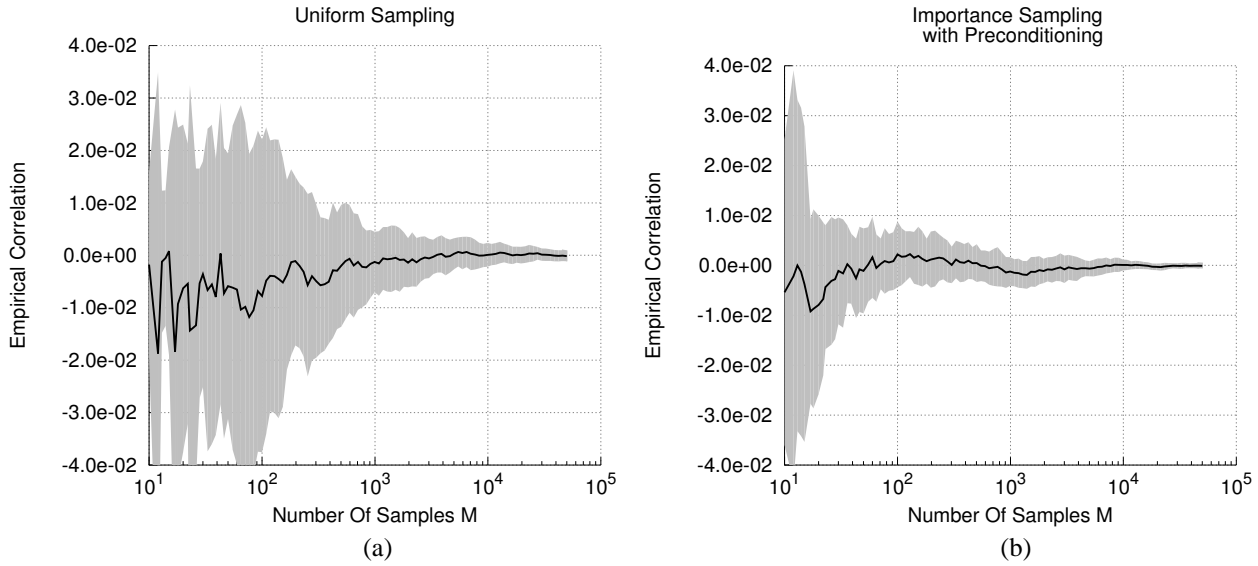


FIG. 5: Empirical correlation of basis functions of Section 6.1 vs. number of samples for uniform and importance sampling as described in Section 6.1. The gray areas indicate values up to one standard deviation away from the average empirical correlation. Products obtained using uniform sampling, (a), and importance sampling with basis preconditioning, (b).

6.2 Sparse Functions with Known Tree Representation

As a second numerical experiment, we study the accuracy of OMP and TOMP in reconstructing two piecewise-smooth functions with uniform and importance sampling.

The two functions $u_1, u_2 : [0, 1] \rightarrow \mathbb{R}$ are selected as follows:

$$u_1 = \begin{cases} \sin(40y) & \text{if } 0 \leq y < 0.25 \\ 10 \sin(15y) & \text{if } 0.25 \leq y \leq 1 \end{cases} \quad \text{and} \quad u_2 = \begin{cases} 1 & \text{if } 0 \leq y < 0.75 \\ 10 \sin(15y) & \text{if } 0.75 \leq y \leq 1, \end{cases}$$

where $y \sim \mathcal{U}[0, 1]$. Figure 6 shows u_1 and u_2 along with their scalar tree representations, where all supports having a coefficient with absolute value larger than $\hat{\alpha}_{tol} = 1.0 \times 10^{-3}$ are colored. Darker colors along the vertical direction denote coarser resolutions and different colors along the horizontal direction correspond to different shift values. Coefficients below this threshold were not included. As may be observed from Fig. 6, u_1 has an *unbalanced* tree representation, that is, the fine scale coefficients are large only on a small portion of the support $[0, 1]$. On the contrary, u_2 admits a *balanced* tree representation, where the large, fine scale coefficients are distributed *uniformly* over the entire support. In the former case, we anticipate considerable difference in the performance of the uniform and importance sampling strategies.

A one-dimensional multiwavelet basis with $j_{\max} = 6$ and $m = 2$ is employed, resulting in a basis with cardinality $P = 768$. The samples are progressively generated from $M = 10$ to $M = 730$ with increments of size 20. Both uniform and importance sampling of Section 5.2 are considered. Reconstructions are performed using OMP and TOMP by solving $(P_{0,\epsilon})$ with 500 independent replications for each sample size M .

In general, the best performance for TOMP and OMP is obtained for different values of the truncation error ϵ in (10). In order to find an *optimal* ϵ , we use cross validation over 11 values of ϵ in the range $[10^{-1}, 10^{-5}]$, with the number of testing to training samples set to $1/4$. For more details on the implementation of this cross validation approach, we refer the interested reader to the work in [26].

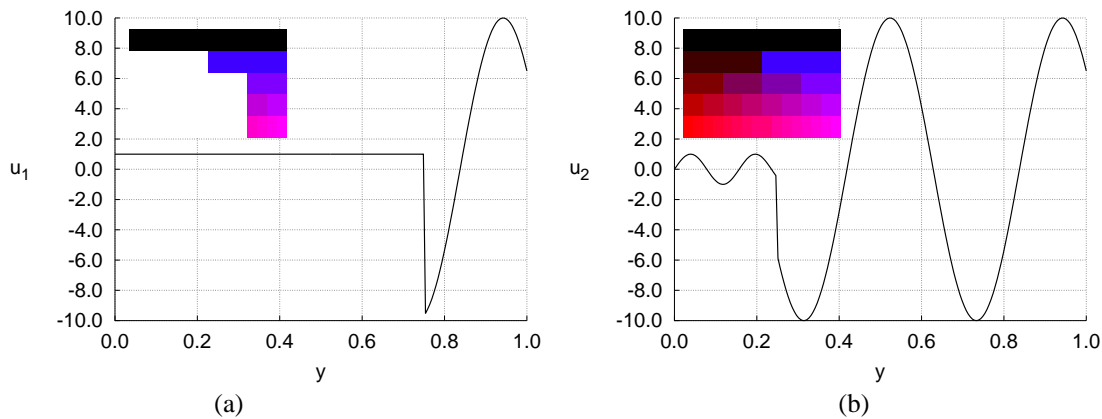


FIG. 6: Piecewise smooth function u_1 , (a), and u_2 , (b), introduced in Section 6.2. A graphical representation of the associated scalar tree is also shown at the top left corner of the two graphs. Darker colors along the vertical direction denote coarser resolutions and different colors along the horizontal direction correspond to different shift values.

To compare the accuracy of OMP and TOMP in approximating u_1 and u_2 , we compute an average of the error

$$\varepsilon = \frac{\|\alpha_{MW} - \alpha_{ref}\|_2}{\|\alpha_{ref}\|_2}, \quad (15)$$

over 500 independent replications, where α_{MW} is the multiwavelet coefficients computed by OMP or TOMP and α_{ref} is obtained by quadrature integration. Specifically, for each multiwavelet coefficient α_i , sufficient number of Gauss-Legendre abscissas are used in the support of ψ_i to compute $\alpha_i = \langle u, \psi_i \rangle_{L^2([0,1])}$.

Figure 7 shows a consistently better performance achieved by TOMP as compared with OMP. The uniform and importance sampling give similar results for the case of balanced tree representation, i.e., corresponding to u_2 . The

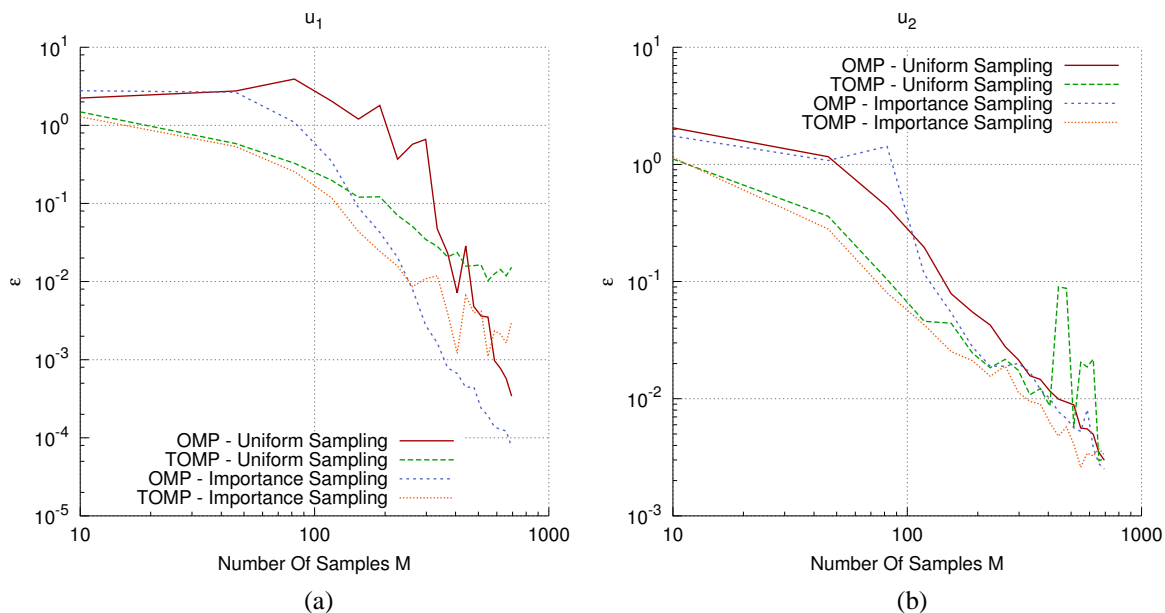


FIG. 7: Average reconstruction relative error ε in (15) for functions u_1 and u_2 defined in Section 6.2 over 500 independent replications. Average errors obtained with OMP, TOMP, uniform and adaptive sampling are displayed in (a) for u_1 and in (b) for u_2 .

combination of TOMP and importance sampling leads to smaller ε in (15) when the tree representation is unbalanced, i.e., for u_1 . In this case the TOMP algorithm is more efficient in finding the correct location of the large coefficients with small sample sizes as all nodes in the tree representation belong to the ancestor set of few leaves located at the maximum resolution. Finally, it can be noticed in Fig. 7 that error curves for OMP and TOMP eventually intersect, with TOMP achieving smaller error ε than OMP for cases of small sample sizes M .

6.3 A Discontinuous Function

In this example, we consider the discontinuous function

$$u(y_1, y_2) = \begin{cases} \sin(\pi y_1) \sin(\pi y_2) & \text{if } y_1 \leq 0.5, y_2 \leq 0.5, \\ 0 & \text{otherwise,} \end{cases} \quad (16)$$

previously examined in [55]. Here, the random variables $y_1, y_2 \sim \mathcal{U}[0, 1]$ are independent. A graphical representation of $u(y_1, y_2)$ is shown in Fig. 8(a).

Methods based on global polynomials, e.g., Legendre polynomials, result in poor approximation due to the presence of discontinuities across $y_1 = 1/2$ and $y_2 = 1/2$. Instead, this problem is well suited for the present multiresolution framework as it gives rise to a sparse expansion in the multiwavelet basis. This sparsity is observed from the scalar tree representation illustrated in Fig. 8(b), where only coefficients at the first two resolutions have significant magnitudes. A 2D representation of the multiwavelet tree leaves contoured by associated coefficient magnitudes is also illustrated in Fig. 8(c). All the weights $\beta_i / \sum_{j=1}^L \beta_j$ are equal in this case.

To assess the accuracy of the multiwavelet expansion of $u(y_1, y_2)$ in (16), we compute the discrete ℓ_p , $p \in \{1, 2, \infty\}$, semi-norms

$$\varepsilon_p = \left(\frac{1}{N} \sum_{i=1}^N |u_{MW}(\mathbf{y}^{(i)}) - u(\mathbf{y}^{(i)})|^p \right)^{1/p}, \quad p \in \{1, 2\}, \quad (17)$$

$$\varepsilon_\infty = \max_{i=1, \dots, N} |u_{MW}(\mathbf{y}^{(i)}) - u(\mathbf{y}^{(i)})|,$$

where u_{MW} denotes the approximant of u computed with the proposed multiwavelet expansion. Figure 9 shows the convergence of ε_1 , ε_2 , ε_∞ as a function of the number of solution realizations M , generated using the proposed importance sampling strategy. It can be observed that TOMP outperforms OMP in terms of the number of samples M needed to achieve a similar accuracy. Additionally, the accuracy of TOMP (or OMP) does not improve beyond a certain level even when larger sample sizes are used. This residual error is due to the finite resolution j and order vector \mathbf{m} used in the construction of multiwavelet basis, and may be reduced by increasing these parameters.

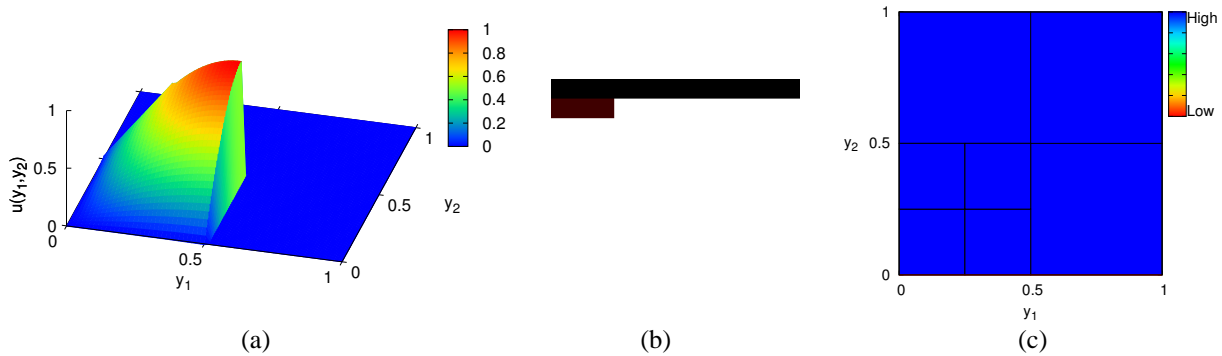


FIG. 8: Visualization of the nonsmooth function in (16), (a). The scalar multiwavelet tree is shown to highlight the resulting sparsity in the multiwavelet representation, (b). Partition of $[0, 1]^2$ colored by the weights $\beta_i / \sum_{j=1}^L \beta_j$, (c).

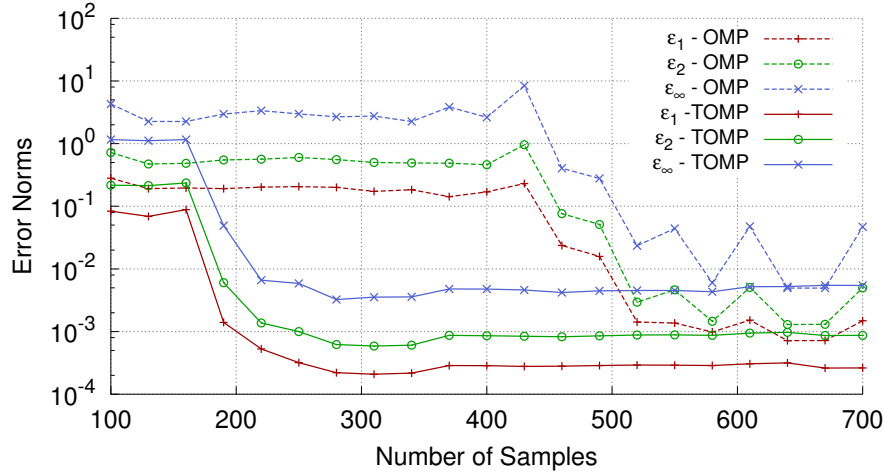


FIG. 9: Convergence of ε_1 , ε_2 , ε_∞ norms, described in (17), with $N = 1000$ samples generated independently from those used to generate the multiwavelet approximation of u in (16). The adaptive importance sampling strategy of Section 5.2 is used in both OMP and TOMP experiments.

The performance of the proposed framework in approximating the statistical moments of the function in (16) is presented in Fig. 10, showing a plot of the number of samples against relative errors of the form

$$\varepsilon_\mu = \frac{|\mathbb{E}[u_{MW}] - \mathbb{E}[u]|}{\mathbb{E}[u]} \quad \text{and} \quad \varepsilon_\sigma = \frac{|\sigma[u_{MW}] - \sigma[u]|}{\sigma[u]} \quad (18)$$

for the mean and standard deviation, respectively. In (18), $\mathbb{E}[u_{MW}]$ and $\sigma[u_{MW}]$ are the mean and standard deviation of u , respectively, resulting from the proposed multiwavelet expansion, while $\mathbb{E}[u]$ and $\sigma[u]$ are the corresponding

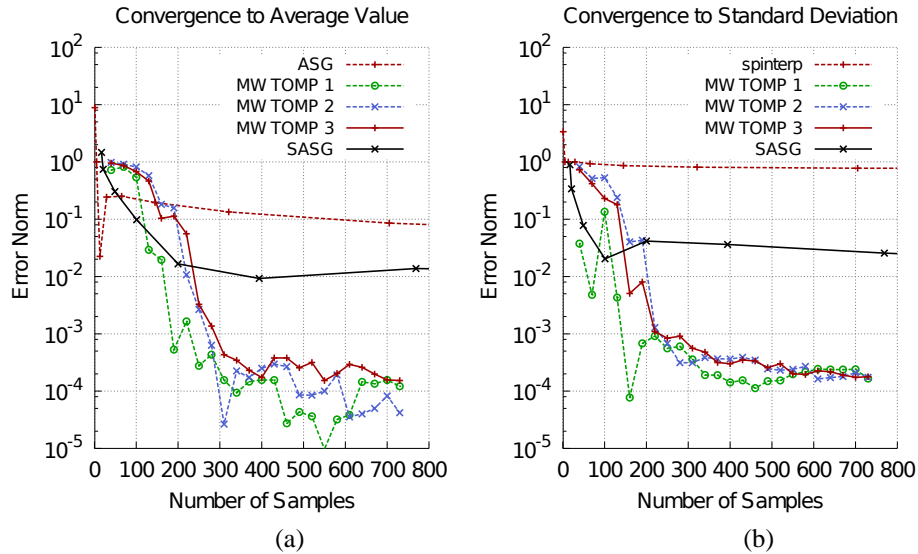


FIG. 10: Relative errors in average value (a) and standard deviation (b) for the nonsmooth function (16). The relative errors produced by the proposed methodology are compared with those computed using dimensionally adaptive sparse grids implemented in the SPINTERP MATLAB package [56] (ASG), and the spatially adaptive sparse grids implemented in SG++ [57] (SASG).

exact values computed analytically. Specifically, the results obtained from TOMP with $m = 2$ and $j_{max} = 4$ are compared to those computed using hierarchical sparse grid collocation implemented in the SPINTERP MATLAB package (ASG in Fig. 10) with dimensional adaptivity [56]. Moreover, the same quantities were also estimated using spatially adaptive sparse grid collocation as implemented in SG++ package [57] (SASG in Fig. 10). It can be seen how the proposed approach produces significantly better estimates of the statistical moments with similar sample sizes.

6.4 Kraichnan-Orszag (KO) Problem

The Kraichnan-Orszag (KO) problem is derived from the simplified inviscid Navier-Stokes equations [58] and is expressed as a coupled system of nonlinear ODEs. We here adopt a rotated version of the original KO problem [18]

$$\frac{du_1}{dt} = u_1 u_3, \quad \frac{du_2}{dt} = -u_2 u_3, \quad \frac{du_3}{dt} = -u_1^2 + u_2^2, \quad (19)$$

with initial conditions specified below.

In [18], the KO problem is used as a benchmark problem and analytical solutions are provided in terms of Jacobi's elliptic functions. The same example is used in [59] to test a treed Gaussian process model in the context of Bayesian uncertainty quantification. If the set of initial conditions is chosen such that the planes $u_1 = 0$ and $u_2 = 0$ are consistently crossed, it is shown in [18] that the accuracy of the global polynomial approximations (at the stochastic level) deteriorates rapidly in time.

6.4.1 Results for the $d = 1$ KO Problem at $t = 30$ s

We assume initial conditions for (19) to be random and specified as

$$u_1(t = 0) = 1, \quad u_2(t = 0) = 0.2y - 0.1, \quad u_3(t = 0) = 0, \quad (20)$$

where y is uniformly distributed on $[0, 1]$. The stochastic response is reconstructed using a multiwavelet dictionary with $m = 3$ and a resolution up to $j_{max} = 7$. Figure 11(a) illustrates the time history of the standard deviation of u_1

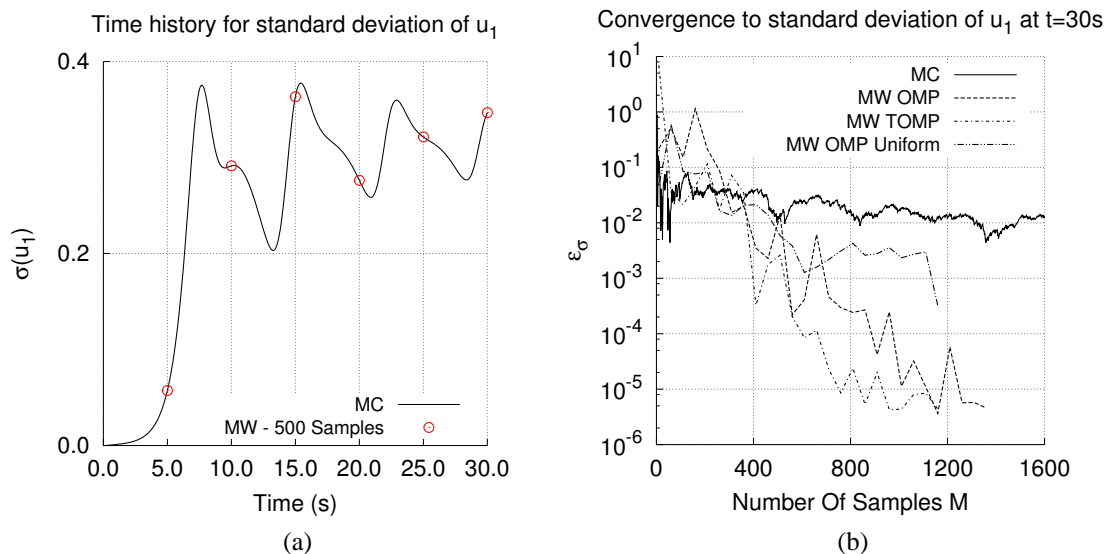


FIG. 11: Results for the 1D KO problem. Estimates for the standard deviation of u_1 at various times are obtained using $M = 500$ realizations of u_1 , (a); Convergence of the standard deviation of u_1 ($t = 30$ s) (as a function of M) computed using OMP and TOMP, (b).

computed using 2.0×10^6 Monte Carlo simulations; the estimates provided by the proposed approach are also shown for time intervals of 5 s and obtained using $M = 500$ realizations of u_1 .

For the stochastic response at $t = 30$ s, the relative standard deviation error

$$\varepsilon_\sigma = \frac{|\sigma_{MW} - \sigma_{ref}|}{\sigma_{ref}} \quad (21)$$

is plotted against the number of samples M in Fig. 11(b). The quantity σ_{MW} is an estimate for the standard deviation of u_1 calculated from the proposed multiwavelet expansion and importance sampling approach, as $\sigma_{MW} = (\sum_{i=2}^P \alpha_i^2)^{1/2}$, while σ_{ref} is a reference value obtained from the Monte Carlo simulation. To illustrate the effect of the importance sampling, we also present results using OMP with uniform sampling.

Figure 12 provides more details on the proposed multiresolution approximation of u_1 using the importance sampling strategy. In particular, the approximation of $u_1(y_1)$ at $t = 30$ s is shown in Fig. 12(a), where the multiwavelet approximation with $M = 500$ coincides with the reference solution. The evolution of the adaptive importance sampling measure $\gamma(y_1)$ is shown in Fig. 12(b). From an initially uniform measure, a large peak in $\gamma(y_1)$ close to $y_1 = 0.5$, i.e., the region with the highest response gradients, is observed. Finally, Figs. 12(c) and 12(d) show the partition of $[0, 1]$ for $M = 500$, with colors reflecting the weights w_i , and the associated scalar multiwavelet tree representation, respectively.

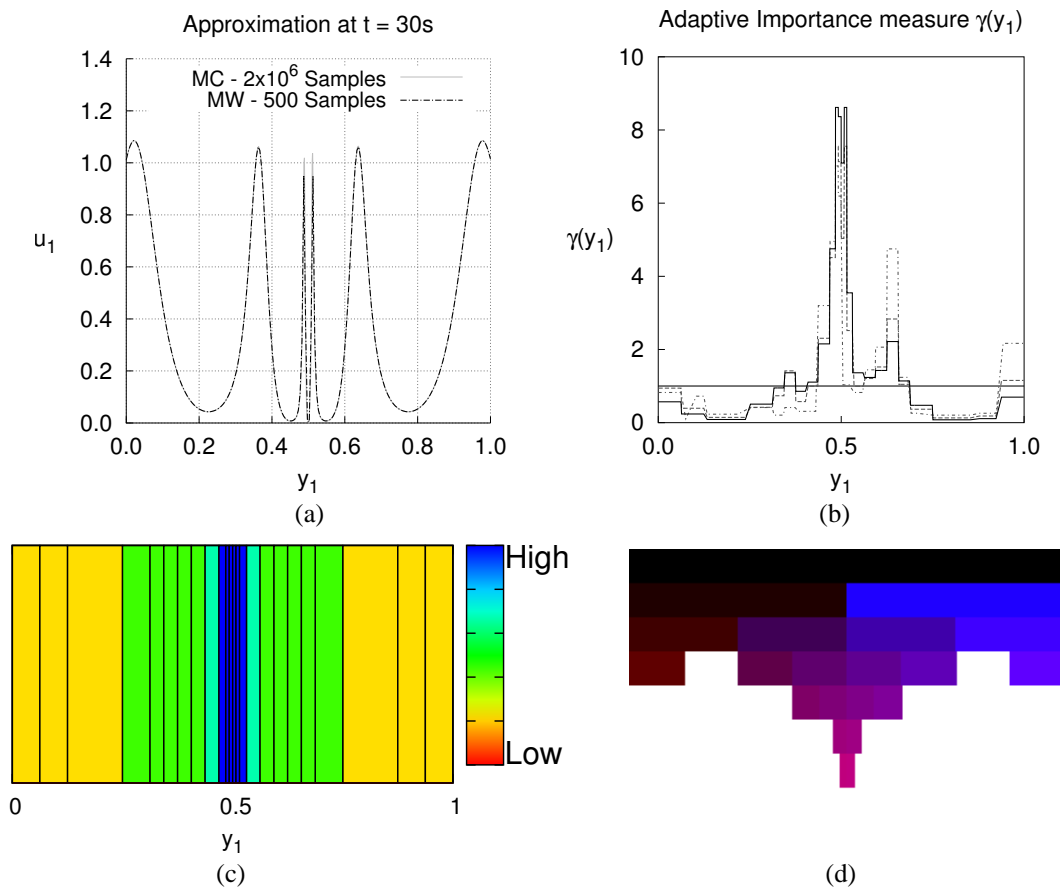


FIG. 12: Approximation of the stochastic response of u_1 at $t = 30$ s for the 1D KO problem, (a); Evolution of the adaptive importance sampling measure for increasing number of samples, (b); Partition of $[0, 1]$ colored by the weights $\beta_i / \sum_{j=1}^L \beta_j$, (c); Scalar multiwavelet tree representation, (d).

6.4.2 Results for $d = 2$ KO Problem at $t = 10$ s

The initial conditions of the Kraichnan-Orszag problem are again assumed to be uncertain but this time functions of two random variables

$$u_1(t=0) = 1, \quad u_2(t=0) = 0.2y_1 - 0.1, \quad u_3(t=0) = 2y_2 - 1, \quad (22)$$

where y_1 and y_2 are independent and uniformly distributed on $[0, 1]$. A two-dimensional multiwavelet approximation of u_1 at $t = 10$ s is generated with $m = 2$ and a maximum resolution $j_{\max} = 4$, resulting in a basis of cardinality $P = 9216$. Figure 13(a) shows the convergence of the standard deviation of u_1 with uniform and adaptive importance sampling. The uniform sampling result are obtained using the OMP algorithm.

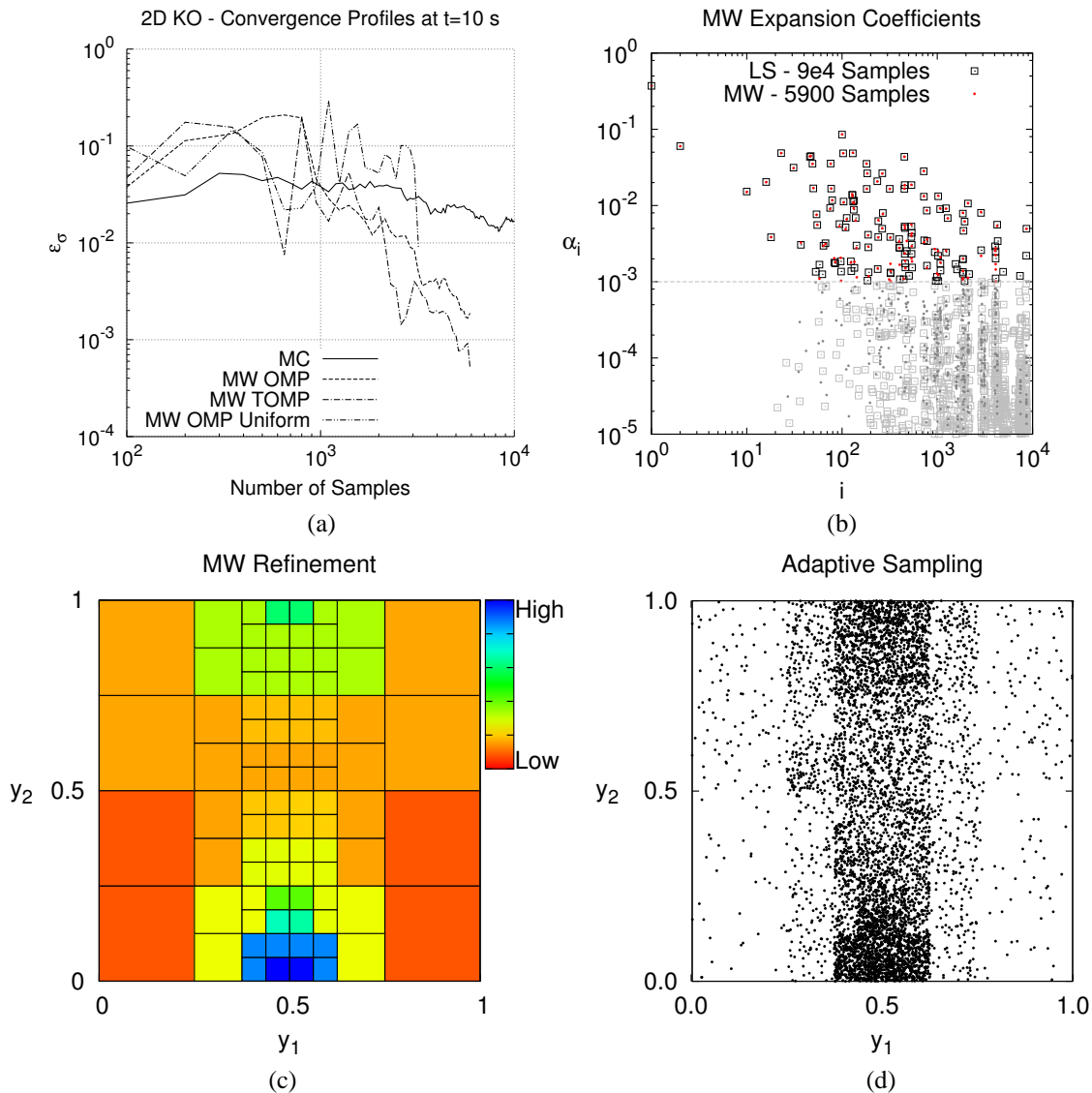


FIG. 13: Results for the 2D KO problem. Standard deviation error (21) for the standard Monte Carlo sampling as well as OMP and TOMP, (a); Multiwavelet expansion coefficients using $M = 5.9 \times 10^3$ samples are compared to those of least-squares regression based on $M = 9.0 \times 10^4$ samples, (b); Partition of $[0, 1]^2$ colored by the weights $\beta_i / \sum_{j=1}^L \beta_j$, (c); Samples generated based on the adaptive importance sampling approach of Section 5.2 ($M = 5.9 \times 10^3$), (d).

In Fig. 13(b), the expansion coefficients computed by the proposed strategy with $M = 5.9 \times 10^3$ samples are compared to those obtained using a standard least-squares regression. In the latter approach the multiwavelet coefficients are given by $\alpha_{LS} = (\Psi^T \Psi)^{-1} \Psi^T \mathbf{u}$ in which we used $M = 9.0 \times 10^4$ samples of u_1 . As may be observed from this plot, the proposed sparse approximation technique recovers the large multiwavelet coefficients accurately using $M < P$ solution realizations. Finally, Figs. 13(c) and 13(d) show the partition of $[0, 1]^2$ with the color scheme based on the w_i , together with $M = 5.9 \times 10^3$ adaptively chosen samples.

6.5 Application: Passive Vibration Control Using Tuned Mass Damper Devices

Vibrations produced by harmonic or stochastic excitations may result in excessive acceleration levels for structures with impacts on serviceability. If resonance occurs, effects of applied forces may be significantly amplified. In this case, the available system damping plays a crucial role. Horizontal acceleration levels higher than 0.5% of g (the gravitational acceleration) may be perceived by the occupants of a given structural system, while 5% of g can be considered an upper bound for serviceability related to human perception. Passive vibration control may provide a cost effective remedy against excessive structural vibrations levels compared to expensive active control systems. Tuned Mass Damper (TMD) devices are among the typical choices for vibration reduction. Their introduction follows from a relatively simple observation on a two Degrees Of Freedom (DOF) spring-mass system: the steady-state undamped response of the principal mass subject to a harmonic excitation can be minimized by applying a TMD device tuned both to the forcing and system frequencies. The efficiency of a TMD can be defined, in this case, based on the reduction obtained in the peak acceleration response of the principal mass. Perfect efficiency, i.e., zero peak acceleration response, is possible under idealized conditions; however, practical efficiency of TMDs is limited by the variations of the actual system/loading conditions from those used in the TMD design. Real forcing, for example, may be characterized by a broad frequency spectrum and generally has variable magnitude. This may make a TMD device, designed for a particular frequency, less effective to prevent excessive vibration for other frequencies.

In the present study, we examine the effect of such uncertainties—relative to nominal conditions—on the efficiency of an example TMD device. To do this, we employ the present multiwavelet regression approach in order to generate stochastic representation of the efficiency metric.

6.5.1 Two DOF System with Passive Vibration Control

The motion of the two DOF system displayed in Fig. 14 is characterized by the triplets $(\ddot{x}_1(t), \dot{x}_1(t), x_1(t))$ and $(\ddot{x}_2(t), \dot{x}_2(t), x_2(t))$ providing the evolution in time of the principal and TMD mass in terms of acceleration, velocity, and displacement, respectively. Assuming a linear elastic material and small oscillations, the equations of motion of the main system with installed TMD device can be written as

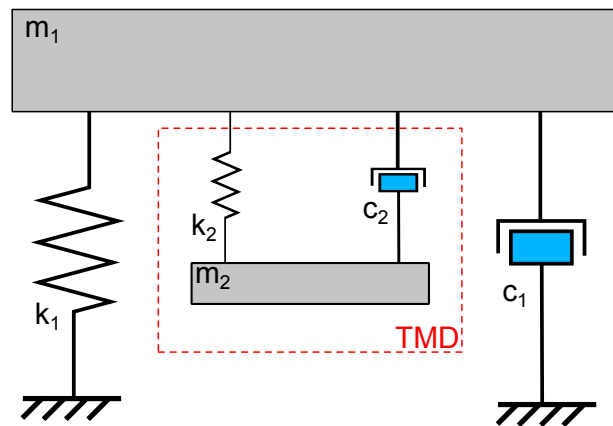


FIG. 14: Schematic representation of a two DOF dynamical system characterized by a principal system (subscript “1”) and an attached TMD device (subscript “2”).

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} + \mathbf{K}\mathbf{x} = \mathbf{f}(t), \quad (23)$$

where

$$\ddot{\mathbf{x}}^T = [\ddot{x}_1, \ddot{x}_2], \quad \dot{\mathbf{x}}^T = [\dot{x}_1, \dot{x}_2], \quad \mathbf{x}^T = [x_1, x_2], \quad \mathbf{f}(t)^T = [f_1(t), f_2(t)], \quad (24)$$

and

$$\mathbf{M} = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} c_1 + c_2 & -c_2 \\ -c_2 & c_2 \end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_2 \end{bmatrix}. \quad (25)$$

We consider characterization of the damping coefficients c_1 and c_2 in terms of *damping ratios* ξ_1 and ξ_2 ,

$$c_1 = 2 \xi_1 \sqrt{m_1 k_1}, \quad c_2 = 2 \xi_2 \sqrt{m_2 k_2}.$$

An instance of displacement, velocity, and acceleration time history of the principal mass m_1 (with TMD device installed) subject to a unit step load applied at $t = 0.5$ s, is illustrated in Fig. 15. In particular, reductions in acceleration amplitudes are observed in the presence of TMD.

A better understanding of the attenuation mechanism of TMD devices may be reached through frequency analysis. Consider the same dynamical system as in (23) where the integration in time has been extended to $T_{tot} = 10.0$ s. A family of harmonic excitations of the form

$$\mathbf{f}(t) = [F^* \sin(2\pi f), 0]^T \quad (26)$$

is considered here, with amplitude F^* and frequency range $f \in [4, 6]$ Hz. A graph of the maximum acceleration in the principal system versus the external excitation frequency is depicted in Fig. 16 for the following configurations:

- undamped principal system with no TMD device installed,
- undamped principal system with undamped TMD device installed,
- $\xi_1 = 0.01$ damped principal system with undamped TMD device installed, and
- $\xi_1 = 0.01$ damped principal system and $\xi_2 = 0.10$ damped TMD device installed.

After installing the TMD device, a single-peak infinite acceleration response, typical of a resonance state of the single DOF system, is replaced by two nearby peaks of lower magnitude. Note that maximum values of acceleration shown in Fig. 16 are obtained from transient responses integrated over a limited time duration of 10.0 s. The effect of $\xi_1 = 0.01$ damping in the principal system also results in a significant reduction in the acceleration response relative to the new peaks, as expected for resonance conditions. An increased damping ratio $\xi_2 = 0.10$ ratio of the TMD device further reduces the peak acceleration response.

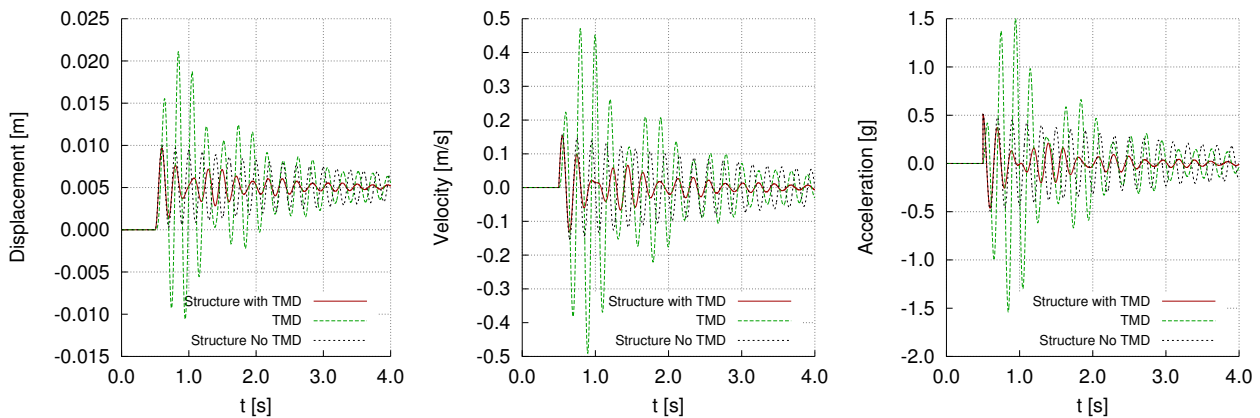


FIG. 15: Transient dynamic simulation of the two DOF system of Fig. 14, showing reduction in the principal system (i.e., mass m_1) response after installation of the TMD device. At $t = 0.5$ s a unit step load is applied to the mass m_1 .

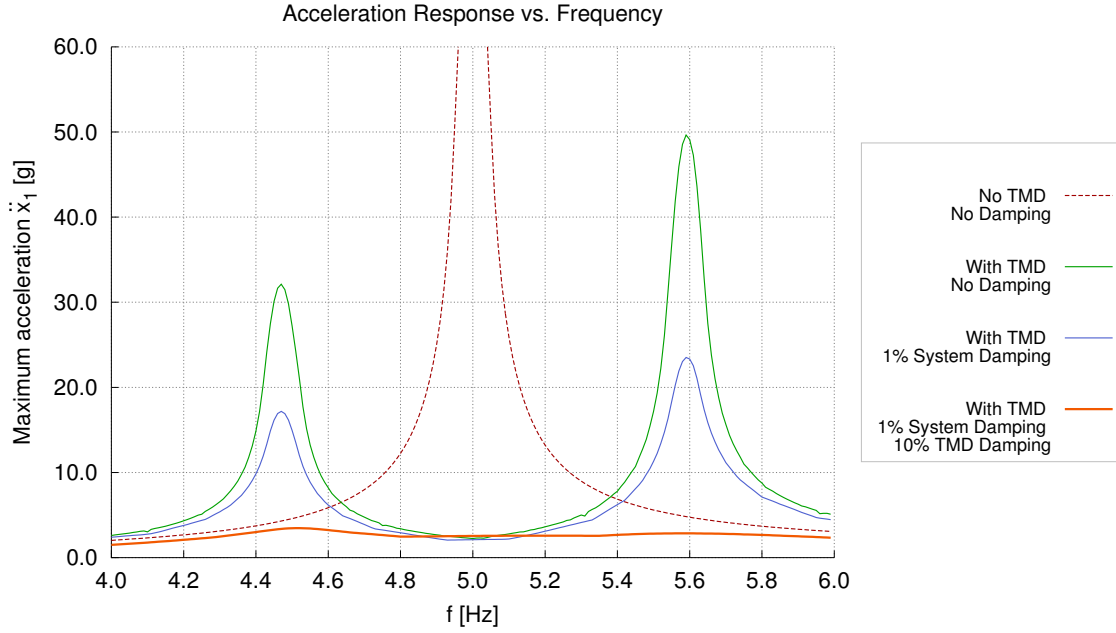


FIG. 16: Acceleration response of a two DOF dynamical system with and without the TMD device installed for a range of forcing frequencies f . The effect of variations in the damping ratios ξ_1, ξ_2 of the principal system and the TMD device is also explored.

6.5.2 Uncertainty Quantification of TMD Efficiency

In our numerical experiments, we consider the forcing frequency f in (26) and the damping ratio of the principal system ξ_1 to be random variables. Randomness in the forcing frequency may be the result of environmental, e.g., due to wind, or anthropic, e.g., induced by walking or running, variations. In particular, we set $f = 3.5 + 2.0 y_1$ and $\xi_1 = 0.005 + 0.005 y_2$ where $y_1, y_2 \sim \mathcal{U}[0, 1]$ are statistically independent. Due to the low level of damping in the system, steep gradients are likely to occur in the stochastic response, thus justifying the application of multiresolution approaches over, for instance, standard Polynomial Chaos (PC) techniques. For each realization of (y_1, y_2) , we evaluate the efficiency e of the TMD defined by

$$e = \frac{\max_t |\ddot{x}_1(t)|}{\max_t |\tilde{\ddot{x}}_1(t)|} - 1, \quad (27)$$

where $\max_t |\ddot{x}_1(t)|$ is the maximum acceleration of the principal system over the time interval $[0, 10 \text{ s}]$ without any vibration control device. Additionally, $\max_t |\tilde{\ddot{x}}_1(t)|$ is the corresponding value with the TMD device installed. Note that when $e \leq 0$, TMD results in amplification or no reduction in the maximum acceleration of the principal mass; therefore, positive large values of e are desirable. We compute the cumulative distribution function (CDF) of the TMD efficiency e using three approaches: the Monte Carlo sampling method, a Legendre PC expansion whose coefficients are computed by numerical integration using tensor-product Clenshaw-Curtis quadrature nodes, and the proposed multiwavelet approach with $j_{\max} = 2$, $m = 2$, $\epsilon = 5.0 \times 10^{-4}$, and adaptive importance sampling.

To assess the accuracy of our approximations, we first generate a reference CDF of e using $M = 1.0 \times 10^4$ Monte Carlo samples. We use $M = 100$ samples for the construction of multiwavelet representation of e . To compute the coefficients of the Legendre PC expansion, we use $M = 10 \times 10$ and $M = 14 \times 14$ nodes obtained by the tensorization of one-dimensional Clenshaw-Curtis abscissas.

The surface plot of the TMD efficiency e as a function of y_1, y_2 is illustrated in Fig. 17(a), and the resulting CDFs of e are shown in Fig. 17(c). As expected, areas of steep gradients close to the resonance can be observed. It can be

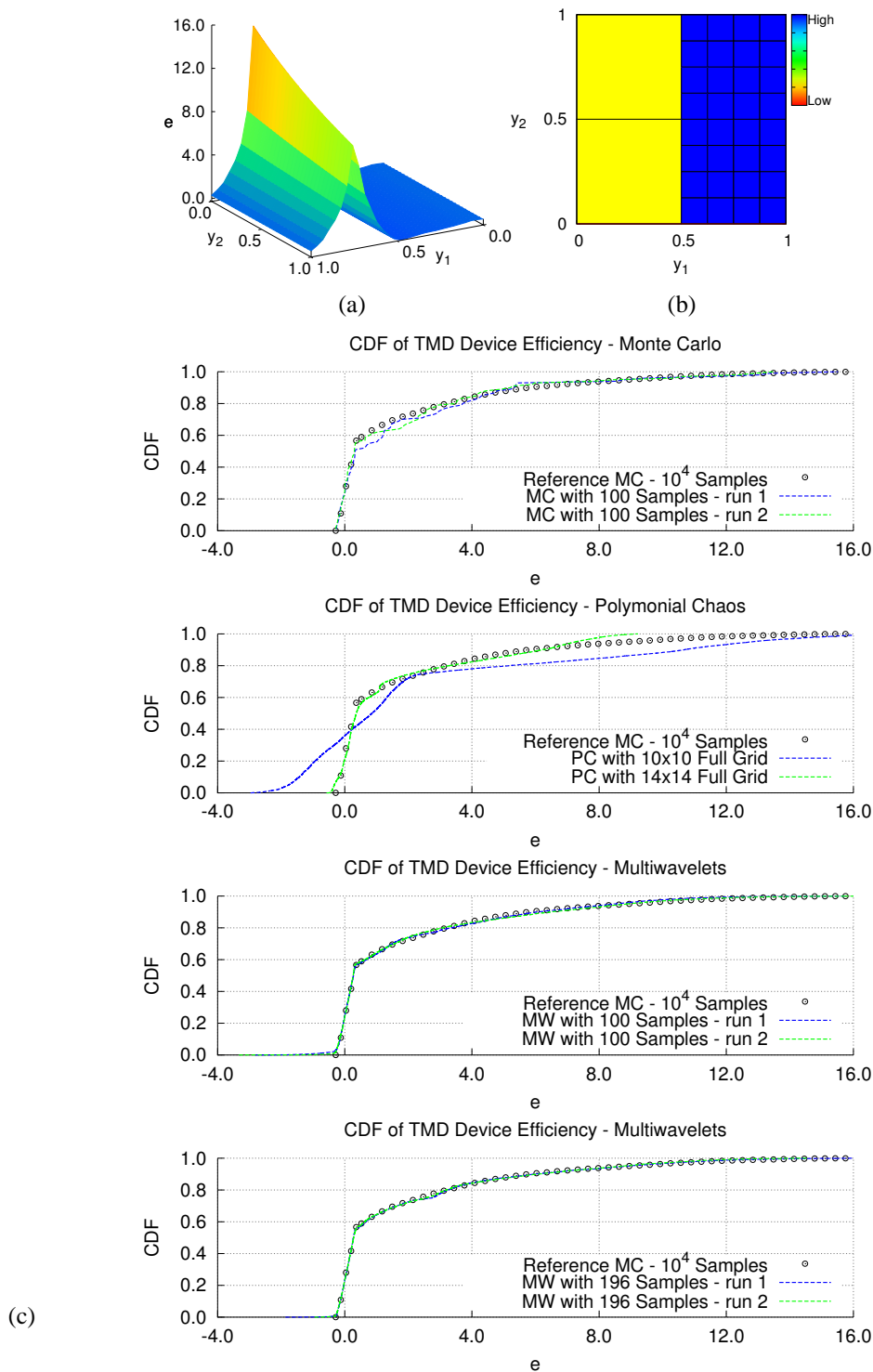


FIG. 17: Representation of the TMD efficiency e , defined in (27), for the two DOF system, (a). Partition of $[0, 1]^2$ colored by the weights $\beta_i / \sum_{j=1}^L \beta_j$, (b). Comparison of the CDFs of e computed with the Monte Carlo, polynomial chaos and proposed approaches, (c).

seen from the plots in Fig. 17(c) that our multiresolution method outperforms both the polynomial chaos and Monte Carlo approaches in estimating the CDF of efficiency e .

7. CONCLUSIONS AND FUTURE WORK

A novel framework for nonintrusive, i.e., sampling-based, uncertainty propagation has been proposed that consolidates the flexibility of multiresolution representations, in capturing piecewise-smooth stochastic responses of physical systems, with the efficiency of sparse approximation techniques. Within this framework, two existing greedy algorithms, namely Orthogonal Matching Pursuit (OMP) and Tree-based OMP (TOMP), have been employed and extended to reconstruct stochastic functions that lend themselves to sparse multiwavelet expansions. In order to enhance the recovery of dominant coefficients, the latter approach exploits the compact tree structure typically exhibited by the multiwavelet coefficients. In addition to the standard sampling of inputs, i.e., according to their probability measure, an adaptive importance sampling strategy has been proposed to further improve the reconstruction accuracy when the solution of interest exhibits sharp gradients or discontinuities.

The accuracy of the proposed multiresolution framework has been demonstrated through its application to a number of benchmark problems as well as a passively controlled dynamical system under uncertainty. In particular, it has been shown numerically that the TOMP algorithm outperforms the standard OMP solver which does not account for any structure in the multiwavelet coefficients. Additionally, it has been illustrated that the proposed adaptive importance sampling strategy achieves higher accuracy as compared to the standard sampling of inputs. The proposed approach improves previous work on nonintrusive construction of multiresolution expansions by exploiting the sparsity of piecewise-smooth stochastic solutions in multiwavelet basis and, thereby, reducing the number of required solution realizations.

In the present study, adaptivity has been investigated with the specific focus on informing the selection of input samples, while fixing the multiwavelet basis. Future work will attempt to extend this approach by adaptively refining the approximation basis. Additionally, theoretical results are needed to certify the advantage of the proposed importance sampling strategy over the standard counterpart. Another possible extension of the present approach consists of associating sparsity-promoting prior probability distribution to the multiwavelet coefficients, providing an alternative way to monitor the local accuracy of the expansion, and estimating the coefficients via Bayesian update. Such a sparse approximation has been previously introduced in [60, 61] and dubbed Bayesian Compressive Sensing. Recent applications of this method to polynomial chaos expansions have also been considered in [62, 63].

ACKNOWLEDGMENTS

The authors gratefully acknowledge various suggestions and comments from the anonymous reviewers that contributed to improve the quality of this manuscript. G. Iaccarino and D. Schiavazzi's work is supported under sub-contract no. B597952, with Lawrence Livermore National Security under prime contract no. DE-AC52-07NA27344 from the Department of Energy National Nuclear Security Administration for the management and operation of the Lawrence Livermore National laboratory. A. Doostan's work is supported by the Department of Energy under Advanced Scientific Computing Research Early Career Research award DE-SC0006402 as well as NSF under award CMMI-1201207.

REFERENCES

1. Xiu, D. and Karniadakis, G., The Wiener–Askey polynomial chaos for stochastic differential equations, *SIAM J. Sci. Comput.*, 24(2):619–644, February 2002.
2. Wiener, N., The homogeneous chaos, *Am. J. Math.*, 60(4):897–936, 1938.
3. Cameron, R. and Martin, W., The orthogonal development of non-linear functionals in series of Fourier–Hermite functionals, *Ann. Math.*, 48(2):385–392, 1947.
4. Ghanem, R. and Spanos, P., *Stochastic Finite Elements: A Spectral Approach*, Springer Verlag, Berlin, 1991.

5. Babuška, I., Tempone, R., and Zouraris, G., Galerkin finite element approximations of stochastic elliptic partial differential equations, *SIAM J. Num. Anal.*, 42(2):800–825, 2004.
6. Cohen, A., Devore, R., and Schwab, C., Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs, *Anal. Appl.*, 9(01):11–47, 2011.
7. Beck, J., Tempone, R., Nobile, F., and Tamellini, L., On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods, *Math. Models Methods Appl. Sci.*, 22(09):1250023, 2012.
8. Xiu, D. and Hesthaven, J., High-order collocation methods for differential equations with random inputs, *SIAM J. Sci. Comput.*, 27(3):1118–1139, 2006.
9. Babuška, I., Nobile, F., and Tempone, R., A stochastic collocation method for elliptic partial differential equations with random input data, *SIAM J. Num. Anal.*, 45(3):1005–1034, 2007.
10. Nobile, F., Tempone, R., and Webster, C., A sparse grid stochastic collocation method for partial differential equations with random input data, *SIAM J. Num. Anal.*, 46(5):2309–2345, 2008.
11. Le Maître, O., Knio, O., Najm, H., and Ghanem, R., Uncertainty propagation using Wiener–Haar expansions, *J. Comput. Phys.*, 197(1):28–57, 2004.
12. Le Maître, O., Najm, H., Ghanem, R., and Knio, O., Multi-resolution analysis of Wiener-type uncertainty propagation schemes, *J. Comput. Phys.*, 197(2):502–531, 2004.
13. Reagan, M., Najm, H., Debusschere, B., Le Maître, O., Knio, O., and Ghanem, R., Spectral stochastic uncertainty quantification in chemical systems, *Combust. Theory Model.*, 8(3):607–632, 2004.
14. Alpert, B., Fast numerical linear algebra, *Wavelets: A Tutorial in Theory and Applications*, vol. 2, Academic Press, Inc., 1992.
15. Alpert, B., A class of bases in l^2 for the sparse representation of integral operators, *SIAM J. Math. Anal.*, 24:246–262, 1993.
16. Alpert, B., Beylkin, G., Coifman, R., and Rokhlin, V., Wavelet-like bases for the fast solution of second-kind integral equations, *SIAM J. Sci. Comput.*, 14:159–184, 1993.
17. Alpert, B., Beylkin, G., Gines, D., and Vozovoi, L., Adaptive solution of partial differential equations in multiwavelet bases, *J. Comput. Phys.*, 182(1):149–190, 2002.
18. Wan, X. and Karniadakis, G., An adaptive multi-element generalized polynomial chaos method for stochastic differential equations, *J. Comput. Phys.*, 209(2):617–642, 2005.
19. Wan, X. and Karniadakis, G., Multi-element generalized polynomial chaos for arbitrary probability measures, *SIAM J. Sci. Comput.*, 28(3):901–928, 2007.
20. Chantasmis, T., Doostan, A., and Iaccarino, G., Padé–Legendre approximants for uncertainty analysis with discontinuous response surfaces, *J. Comput. Phys.*, 228(19):7159–7180, 2009.
21. Ma, X. and Zabarar, N., An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations, *J. Comput. Phys.*, 228(8):3084–3113, 2009.
22. Witteveen, J. and Iaccarino, G., Simplex stochastic collocation with random sampling and extrapolation for non-hypercube probability spaces, *SIAM J. Sci. Comput.*, 34:A814–A838, 2012.
23. Abgrall, R., Congedo, P., Corre, C., and Galera, S., A simple semi-intrusive method for uncertainty quantification of shocked flows, comparison with a non-intrusive polynomial chaos method, In *ECCOMAS CFD 2010—Fifth European Conference on Computational Fluid Dynamics*, Lisbon, Portugal, 14–17 June, 2010.
24. Tryoen, J., Le Maître, O., and Ern, A., Adaptive anisotropic spectral stochastic methods for uncertain scalar conservation laws, *SIAM J. Sci. Comput.*, 34(5):A2459–A2481, 2012.
25. Doostan, A., Owahdi, H., Lashgari, A., and Iaccarino, G., Non-adapted sparse approximation of PDEs with stochastic inputs, Tech. Rep. Annual Research Brief, Center for Turbulence Research, Stanford University, 2009.
26. Doostan, A. and Owahdi, H., A non-adapted sparse approximation of PDEs with stochastic inputs, *J. Comput. Phys.*, 230(8):3015–3034, 2011.
27. Blatman, G. and Sudret, B., Adaptive sparse polynomial chaos expansion based on least angle regression, *J. Comput. Phys.*, 230:2345–2367, 2011.
28. Mathelin, L. and Gallivan, K., A compressed sensing approach for partial differential equations with random input data, *Commun. Comput. Phys.*, 12:919–954, 2012.

29. Yan, L., Guo, L., and Xiu, D., Stochastic collocation algorithms using ℓ_1 -minimization, *Int. J. Uncertainty Quantification*, 2(3):279–293, 2012.
30. Yang, X. and Karniadakis, G. E., Reweighted ℓ_1 minimization method for stochastic elliptic differential equations, *J. Comput. Phys.*, 248:87–108, 2013.
31. Donoho, D., Compressed sensing, *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
32. Candes, E. and Tao, T., Decoding by linear programming, *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.
33. Bruckstein, A., Donoho, D., and Elad, M., From sparse solutions of systems of equations to sparse modeling of signals and images, *SIAM Rev.*, 51(1):34–81, 2009.
34. Chen, S., Donoho, D., and Saunders, M., Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.*, 20(1):33–61, 1998.
35. Donoho, D., Tsai, Y., Drori, I., and Starck, J., Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit, *IEEE Trans. Inform. Theory*, 58(2):1094–1121, 2012.
36. La, C. and Do, M., Tree-based orthogonal matching pursuit algorithm for signal reconstruction, In *IEEE International Conference on Image Processing*, Atlanta, Georgia, Oct. 8–11, 2006.
37. Tropp, J. and Gilbert, A., Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Inform. Theory*, 53(12):4655–4666, 2007.
38. Candes, E., Wakin, M., and Boyd, S., Enhancing sparsity by reweighted ℓ_1 minimization, *J. Fourier Anal. Appl.*, 14(5):877–905, 2008.
39. Van Den Berg, E. and Friedlander, M., Probing the Pareto frontier for basis pursuit solutions, *SIAM J. Sci. Comput.*, 31(2):890–912, 2008.
40. Daubechies, I., DeVore, R., Fornasier, M., and Güntürk, C., Iteratively reweighted least squares minimization for sparse recovery, *Commun. Pure Appl. Math.*, 63(1):1–38, 2010.
41. Needell, D. and Tropp, J., CoSaMP: Iterative signal recovery from incomplete and inaccurate samples, *Appl. Comput. Harmonic Anal.*, 26(3):301–321, 2009.
42. Beck, A. and Teboulle, M., A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.
43. Mallat, S. and Zhang, Z., Matching pursuits with time-frequency dictionaries, *IEEE Trans. Signal Processing*, 41(12):3397–3415, 1993.
44. Pati, Y. C., Rezaiifar, R., and Krishnaprasad, P., Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, In *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, California, pp. 40–44, Nov. 1–3, 1993.
45. Tropp, J. A., Greed is good: Algorithmic results for sparse approximation, *IEEE Trans. Inform. Theory*, 50(10):2231–2242, 2004.
46. La, C. and Do, M., Signal reconstruction using sparse tree representations, In *Proceedings—Wavelets XI at SPIE Optics and Photonics*, San Diego, California, July 22–26, 2005.
47. Duarte, M., Wakin, M., and Baraniuk, R., Fast reconstruction of piecewise smooth signals from incoherent projections, *SPARS 05*, Rennes, France, Nov. 16–18, 2005.
48. Escoda, O., Granai, L., and Vanderghenst, P., On the use of a priori information for sparse signal approximations, *IEEE Trans. Signal Proces.*, 9:3468–3482, 2006.
49. Khajehnejad, M. A., Xu, W., Avestimehr, A. S., and Hassibi, B., Improved sparse recovery thresholds with two-step reweighted ℓ_1 minimization, In *IEEE International Symposium on Information Theory Proceedings (ISIT)*, Austin, Texas, June 13–18, pp. 1603–1607, 2010.
50. Peng, J., Hampton, J., and Doostan, A., A weighted ℓ_1 -minimization approach for sparse polynomial chaos expansions, *J. Comput. Phys.*, 267:92–111, 2014.
51. Boufounos, P., Duarte, M., and Baraniuk, R., Sparse signal reconstruction from noisy compressive measurements using cross validation, In *IEEE/SP 14th Workshop on Statistical Signal Processing*, Madison, Wisconsin, Aug. 26–29, pp. 299–303, 2007.
52. Rauhut, H. and Ward, R., Sparse Legendre expansions via ℓ_1 -minimization, *J. Approx. Theory*, 164(5):517–533, 2012.
53. Rauhut, H., Compressive sensing and structured random matrices, *Theor. Foundations Num. Methods Sparse Recovery*, 9:1–

- 92, 2010.
54. Srinivasan, R., *Importance Sampling: Applications in Communications and Detection*, Springer, Berlin, 2002.
 55. Agarwal, N. and Aluru, N., A domain adaptive stochastic collocation approach for analysis of MEMS under uncertainties, *J. Comput. Phys.*, 228(20):7662–7688, 2009.
 56. Klimke, A. and Wohlmuth, B., Algorithm 847: Spinterp: piecewise multilinear hierarchical sparse grid interpolation in MATLAB, *ACM Trans. Math. Software (TOMS)*, 31(4):561–579, 2005.
 57. Pflüger, D., Peherstorfer, B., and Bungartz, H., Spatially adaptive sparse grids for high-dimensional data-driven problems, *J. Complexity*, 26(5):508–522, 2010.
 58. Kraichnan, R., Direct-interaction approximation for a system of several interacting simple shear waves, *Phys. Fluids*, 6:1603–1609, 1963.
 59. Bilonis, I. and Zabaras, N., Multi-output local Gaussian process regression: Applications to uncertainty quantification, *J. Comput. Phys.*, 231(17):5718–5746, 2012.
 60. Ji, S., Xue, Y., and Carin, L., Bayesian compressive sensing, *IEEE Trans. Signal Proces.*, 56(6):2346–2356, 2008.
 61. He, L. and Carin, L., Exploiting structure in wavelet-based Bayesian compressive sensing, *IEEE Trans. Signal Proces.*, 57(9):3488–3497, 2009.
 62. Karagiannis, G. and Lin, G., Selection of polynomial chaos bases via Bayesian model uncertainty methods with applications to sparse approximation of PDEs with stochastic inputs, *J. Comput. Phys.*, 259:114–134, 2014.
 63. Sargsyan, K., Safta, C., Najm, H., Debusschere, B., Ricciuto, D., and Thornton, P., Dimensionality reduction for complex models via Bayesian compressive sensing, *Int. J. Uncertainty Quantification*, 4(1):63–93, 2014.