Enhancing ℓ_1-minimization estimates of polynomial chaos expansions using basis selection

J.D. Jakeman [a,*], M.S. Eldred [a], K. Sargsyan [b]

[a] *Sandia National Laboratories, Albuquerque, NM 87185, United States*
[b] *Sandia National Laboratories, Livermore, CA 94550, United States*

ARTICLE INFO

ABSTRACT

In this paper we present a basis selection method that can be used with ℓ_1-minimization to adaptively determine the large coefficients of polynomial chaos expansions (PCE). The adaptive construction produces anisotropic basis sets that have more terms in important dimensions and limits the number of unimportant terms that increase mutual coherence and thus degrade the performance of ℓ_1-minimization. The important features and the accuracy of basis selection are demonstrated with a number of numerical examples. Specifically, we show that for a given computational budget, basis selection produces a more accurate PCE than would be obtained if the basis were fixed a priori. We also demonstrate that basis selection can be applied with non-uniform random variables and can leverage gradient information.

1. Introduction

Quantifying uncertainty in a computational model is essential to building the confidence of stakeholders in the predictions of that model. Sources of uncertainty in model predictions can be broadly grouped into two classes, uncertainty arising from model structure and uncertainty arising from the model parameterization. The effect of these uncertainties must be traced through the model and the effect on the model output (prediction) needs to be quantified. In this paper we will present a method for quantifying parametric uncertainty that utilizes the strengths of Polynomial Chaos Expansions (PCE) and ℓ_1-minimization.

When the computational cost of a simulation model is large, the most popular and effective means of quantifying parametric uncertainty is to construct an approximation of the response of the model output to variations in the model input. Once built, this surrogate can be interrogated cheaply, without further model evaluations, to obtain statistics of interest such as model output moments and distributions. Within the computational science community, the most widely adopted approximation methods used for Uncertainty Quantification (UQ) are based on generalized polynomial chaos expansions [25, 46], sparse grid interpolation [27,31] and Gaussian process models [37].

Polynomial chaos expansions represent a response surface as a linear combination of orthonormal multivariate polynomials. The choice of the orthonormal polynomials is related to the distribution of the model input variables. Provided sufficient smoothness conditions are met, PCEs exhibit high rates of convergence. Indeed in some cases even exponential

convergence can be obtained [1,46]. In this paper we will focus on PCEs as they allow one to leverage the advantages of ℓ_1-minimization for computing approximations from limited data.

The stochastic Galerkin [25,46] and stochastic collocation [2,33,41,45] methods are the two main approaches for approximating the PCE coefficients. The former is typically intrusive and so is only feasible when one has the ability to modify the code used to solve the governing equations of the model. Stochastic collocation, however, is a non-intrusive sampling based approach that allows the computational model to be treated as a black box. In this paper we focus on stochastic collocation which involves running the computational model with a set of realizations of the random parameters and constructing an approximation of corresponding model output.

Pseudo-spectral projection [16,17], sparse grid interpolation [23,27,31,35], probabilistic multi-element methods [22] are stochastic collocation methods which have been used effectively in many situations. These methods, however, all require structured samples and/or the ability to iteratively determine the collocation points.

Recently ℓ_1-minimization [11,12,18,19] and other sparse approximation methods [14,42] have been shown to be an effective method for approximating PCE coefficients from small number of and possibly arbitrarily positioned collocation nodes [7,20,32,36,39,49]. These methods are very effective when the number of non-zero terms in the PCE approximation of the model output is small (i.e. sparse) or the magnitude of the PCE coefficients decays rapidly (i.e. compressible).

The efficacy of ℓ_1-minimization when used to estimate PCE coefficients is dependent on the rate of the decay of the PCE coefficients, the characteristics of the stochastic collocation samples and the truncation of the PCE. The decay of the coefficients is dependent on the sparsity of the model response in the chosen basis. In this paper we take the typical approach [7,20] and adopt the Askey scheme [46], which selects the polynomial basis according to the distribution of the uncertain variables. The Askey scheme is desirable as it facilitates the analytical computation of moments and sensitivity indices from a PCE once it is built [40]. Using the Askey scheme the type of PCE basis is set and cannot be adjusted to enhance ℓ_1-recovery, however the truncation of the PCE and the sampling of the model inputs can both be controlled.

Recently some attention has been given to designing sampling strategies to increase the accuracy of sparse PCE [38, 47,48]. Almost no attention, however, has been given to the effect of the PCE truncation when using ℓ_1-minimization. Typically, when using ℓ_1-minimization, a total-degree truncation is applied to PCE. However the number of terms in this basis grows factorially with the number of model parameters. This fast growth in the number of basis terms significantly affects the ability of ℓ_1-minimization to accurately approximate PCE coefficients. To reduce the growth of a PCE basis in high dimensions a hyperbolic cross PCE truncation can be employed [7]. However, despite the slower growth of the hyperbolic truncation it can perform poorly when the 'true' PCE has large coefficients associated with interaction basis terms.

The goal of this paper is to present a basis selection algorithm that adaptively determines a set of PCE basis terms that enable accurate approximation of PCE coefficients using ℓ_1-minimization. Specifically, we aim to:

- Present an iterative algorithm for selecting a polynomial chaos basis set that, for a given computational budget, produces a more accurate PCE than would be obtained if the basis is fixed a priori.
- Demonstrate numerically that in high dimensions, for which high-order total-degree PCE bases are infeasible, basis selection allows the accurate identification of high-order terms that cannot be captured by a low-order total-degree basis.
- Demonstrate numerically that even for lower dimensional problems, for which high-order total-degree PCE bases are feasible, basis selection still produces more accurate results than a priori fixed basis sets.
- Show that basis selection can leverage function gradients, that for a given computational budget, will produce more accurate approximations than an approximation based solely on function values.
- Illustrate that basis selection can be applied with non-uniform random variables.

The remainder of this paper is organized as follows: Section 2 provides a brief summary of PCEs; Section 3 discuses how to use ℓ_1-minimization for building a PCE and the need to move away from a priori-fixed PCE truncations in higher dimensions; Section 4 proposes a new method for iteratively defining PCE truncations; the properties and effectiveness of the proposed method are demonstrated numerically in Section 5; and conclusions are presented in Section 6.

2. Polynomial chaos expansions

Polynomial Chaos methods represent both the model inputs $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{\tilde{d}})$ and model output $f(\boldsymbol{\theta})$ as an expansion of orthonormal polynomials of random variables $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_d)$. Specifically we represent the random inputs as

$$\theta_n \approx \sum_{i=1}^{N_{\theta_n}} \beta_i \phi_i(\boldsymbol{\xi}), \quad n = 1, \ldots, \tilde{d} \tag{1}$$

and the model output as

$$f(\boldsymbol{\theta}(\boldsymbol{\xi})) \approx \hat{f}(\boldsymbol{\xi}) = \sum_{i=1}^{N} \alpha_i \phi_i(\boldsymbol{\xi}). \tag{2}$$
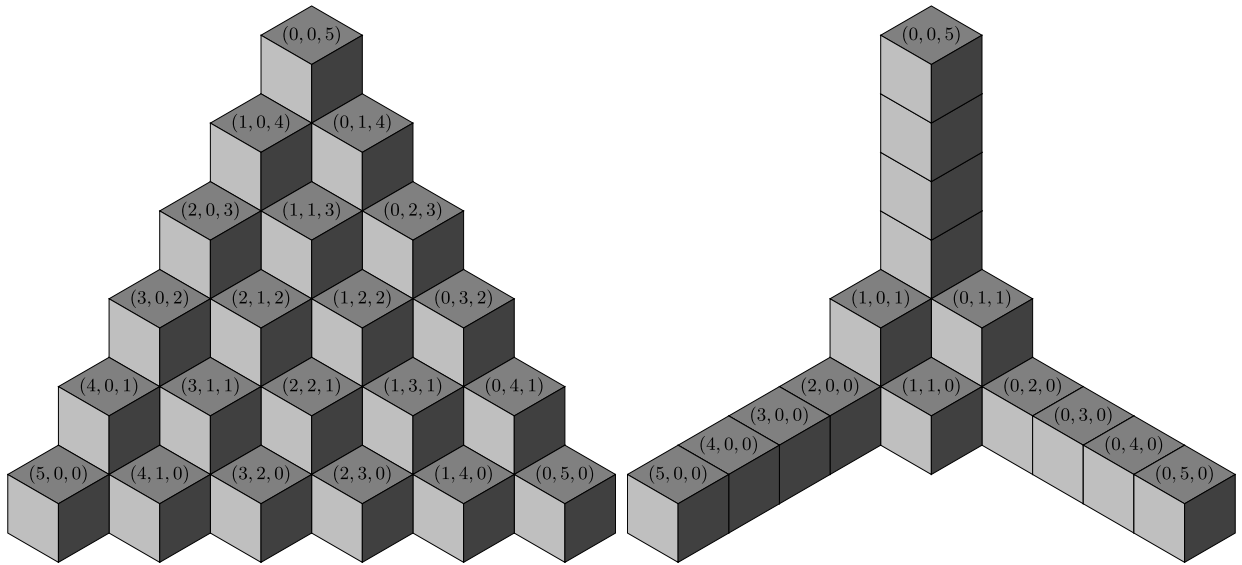
Fig. 1. (Left) A total-degree index set $\Lambda_{6,1}^3$. (Right) A hyperbolic index set $\Lambda_{6,1/2}^3$. Each cube represents a 3-dimensional index $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$ in Λ. The value of each $\boldsymbol{\lambda}$ is given on the top of each cube.

We refer to (1) and (2) as a polynomial chaos expansion (PCE). The PCE basis functions $\{\phi_i(\boldsymbol{\xi})\}$ are tensor products of orthonormal polynomials which are chosen to be orthonormal with respect to the probability distribution function $\rho(\boldsymbol{\xi})$ of the random vector $\boldsymbol{\xi}$. That is

$$\left(\phi_i(\boldsymbol{\xi}), \phi_j(\boldsymbol{\xi})\right) = \int_{I_{\boldsymbol{\xi}}} \phi_i(\boldsymbol{\xi})\phi_j(\boldsymbol{\xi})\rho(\boldsymbol{\xi})d\boldsymbol{\xi} = \delta_{ij}$$

where $I_{\boldsymbol{\xi}}$ is the range of the random variables.

The random variable (germ) $\boldsymbol{\xi}$ of the PCE is typically related to the distribution of the input variables. For example, if the one-dimensional input variable θ is uniform on $[a, b]$ then ξ is also chosen to be uniform on $[-1, 1]$ and ϕ are chosen to be Legendre polynomials such that $\theta = \beta_1 + \beta_2\xi = (b + a)/2 + \xi(b - a)/2$. For simplicity and without loss of generality, we will assume that $\boldsymbol{\xi}$ has the same distribution as $\boldsymbol{\theta}$ and thus we can use the two variables interchangeably (up to a linear transformation which we will ignore).

The rate of convergence is dependent on the regularity of the response surface. If $f(\boldsymbol{\xi})$ is analytical with respect to the random variables then (2) converges exponentially in $L_2(\rho(\boldsymbol{\xi}))$-sense [6].

In practice the PCE (2) must be truncated. The most common approach is to set a degree p and retain only the multivariate polynomials of degree at most p. Rewriting (2) using the typical multi-dimensional index notation

$$f(\boldsymbol{\xi}) \approx \hat{f}(\boldsymbol{\xi}) = \sum_{\boldsymbol{\lambda} \in \Lambda} \alpha_{\boldsymbol{\lambda}} \phi_{\boldsymbol{\lambda}}(\boldsymbol{\xi}) \tag{3}$$

the total-degree basis of degree p is given by

$$\Lambda = \Lambda_{p,q}^d = \{\phi_{\boldsymbol{\lambda}} : \|\boldsymbol{\lambda}\|_q \le p\}, \quad \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d) \tag{4}$$

with $q = 1$. The number of terms in this total-degree basis

$$\operatorname{card} \Lambda_{p,1}^d \equiv P = \binom{d + p}{d}$$

grows factorially with dimension. This rapid growth limits the applicability of the total-degree basis to moderate dimensions or low degree polynomials in higher dimensions.

The authors of [7] propose using hyperbolic index sets, (4) with $q < 1$, to slow the growth of the PCE basis with dimensionality. The use of hyperbolic indices assumes that the contribution to variance from the interaction between the random variables decays rapidly as the number of variables involved in the interaction increases. Fig. 1 shows both a three-dimensional total-degree index set and a three-dimensional hyperbolic index set. It is clear that for a given degree p the hyperbolic index set has many less terms than the total-degree polynomial basis. However the smaller basis size requires omitting polynomial terms interaction terms, that is indices $\boldsymbol{\lambda}$ with at least two $\lambda_n > 0, n = 1, \ldots, d$. When a function has large non-zero PCE coefficients corresponding to these missing multivariate basis terms, the hyperbolic index set may be an inappropriate form of truncation. Ideally the basis set should be adapted to the function being approximated.

3. ℓ_1-minimization

The coefficients of a polynomial chaos expansion can be approximated effectively using ℓ_1-minimization. Specifically, given a small set of M unstructured realizations $\Xi = \{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_M\}$, with corresponding model outputs $\mathbf{f} = (f(\boldsymbol{\xi}_1), \ldots, f(\boldsymbol{\xi}_M))^T$, we would like to find a solution that satisfies

$$\boldsymbol{\Phi}\boldsymbol{\alpha} \approx \mathbf{f}$$

where $\boldsymbol{\alpha} = (\alpha_{\lambda_1}, \ldots, \alpha_{\lambda_N})^T$ denotes the vector of PCE coefficients and $\boldsymbol{\Phi}$ denotes the Vandermonde matrix with entries $\boldsymbol{\Phi}_{ij} = \phi_j(\boldsymbol{\xi}_i)$, $i = 1, \ldots, M$, $j = 1, \ldots, N$.

When the model $f(\boldsymbol{\xi})$ is high-dimensional and computationally expensive, and non-adaptive basis truncation rules are employed, the number of model simulations that can be generated is much smaller than the number of unknown PCE coefficients, i.e. $M \ll N$. Under these conditions, finding the PCE coefficients is ill-posed and we must impose some form of regularization to obtain a unique solution. One way to regularize is to enforce sparsity in a PCE.

A polynomial chaos expansion is defined as s-sparse when $\|\boldsymbol{\alpha}\|_0 \leq s$, i.e. the number of non-zero coefficients does not exceed s. ℓ_1-minimization provides a means of identifying *sparse* coefficient vectors from a limited amount of simulation data. In practice, not many simulation models will be truly sparse, but PCE are often *compressible*, that is the magnitude of the coefficients decays rapidly or alternatively most of the PCE variance is concentrated in a few terms. Compressible vectors are well represented by sparse vectors and thus the coefficients of compressible PCE can also be recovered accurately using ℓ_1-minimization.

ℓ_1-minimization attempts to find the dominant PCE coefficients by solving the following optimization problem

$$\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \quad \text{such that} \quad \|\boldsymbol{\Phi}\boldsymbol{\alpha} - \mathbf{f}\|_2 \leq \varepsilon \tag{5}$$

This ℓ_1-minimization problem is often referred to as Basis Pursuit Denoising. The problem obtained by setting $\varepsilon = 0$, to enforce interpolation, is termed Basis Pursuit. There is a close connection between (5) and Least Absolute Shrinkage Operator (LASSO) [42] well known in the statistics literature. Indeed these problems are equivalent under certain conditions [19].

3.1. ℓ_1-minimization algorithms

Numerous algorithms [5,7,13,34,43] exist for solving (5) which are all stable and accurate under certain well defined conditions. In this paper we will use the greedy algorithm Orthogonal Matching Pursuit (OMP) [14] to estimate PCE coefficients. We remark, however, that the basis selection procedure presented in this paper can be used in conjunction with most ℓ_1-minimization algorithms.

OMP requires stronger theoretical conditions than some of its counterparts [10]. For example, for a sufficiently small ratio of the number of samples to sparsity, basis pursuit can guarantee recovery of all s-sparse polynomials with high probability, whereas, for a fixed set of samples Ξ, OMP can guarantee recovery of at least one sparse polynomial but not all [29]. However despite this theoretical difference, in practice OMP can still obtain comparable accuracy to non-approximate algorithms such as BP [29]. Moreover OMP has a much faster execution speed than most algorithms, which makes OMP more amenable to cross validation which can be used to estimate optimal method parameters such as the tolerance ε of (5) as discussed in the following section.

3.1.1. Hyper-parameter estimation via cross validation

Accurately computing the coefficients of a polynomial chaos expansion requires determining a 'good' truncation set Λ and specifying the tolerance ε in the Basis Pursuit DeNoising problem (5). Cross validation has been shown to be effective at aiding these choices. Specifically cross validation has been used in the past to estimate the polynomial degree p of a hyperbolic expansion [7] and to estimate the tolerance ε of (5) [7,8,20,32,44].[1]

In this paper we will use $K = 10$ fold cross validation to choose the values of sets of hyper-parameters $\boldsymbol{\gamma}$. The number and type of hyper-parameters is dependent on the ℓ_1-minimization method used in conjunction with cross validation. As an example consider solving in (5) using an a priori fixed total-degree basis $\Lambda_{p,1}^d$. The hyper-parameters that can be estimated using cross validation are the degree p and the tolerance ε, that is $\boldsymbol{\gamma} = (p, \varepsilon)$.

Let $\zeta : \{1, \ldots, M\} \to \{1, \ldots, K\}$ be an indexing function that determines the partition of the training data. Furthermore let $\hat{f}^{-\zeta}$ be the PCE approximation built on the data with the ζ part removed, then the cross validation error is given by

$$e_{\text{cv}}(\boldsymbol{\gamma}) = \frac{1}{M} \sum_{k=1}^{K} e_{\zeta(k)}, \quad e_{\zeta(k)} = \sum_{j \in \zeta(k)} (y_j - \hat{f}^{-\zeta(k)}(x_j))^2 \tag{6}$$

To compute e_{cv} we divide the data pairs $(\Xi, \mathbf{f})$, based upon the randomly chosen partitions $\zeta(k)$, into K sets (folds) of equal size $(\Xi_k, \mathbf{f}_k)$, $k = 1, \ldots, K$. A PCE $\hat{f}^{-\zeta(k)}$, is then built on the training data $\Xi_t = \Xi \setminus \Xi_k$ with the k-th fold removed,

---

[1] The choice of ε can significantly affect the accuracy of the PCE obtained using (5). Reducing ε can lead to over-fitting, whilst higher values of ε can deteriorate the accuracy of the approximation.
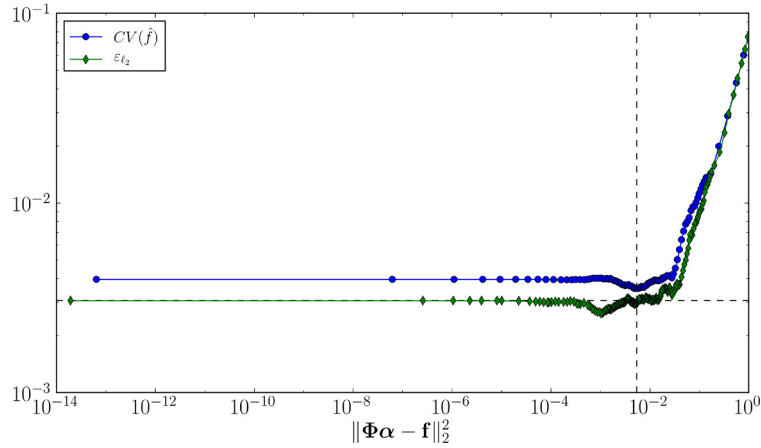
Fig. 2. The use of cross validation to select the truncation tolerance ε for (11) with coefficients $c^{(1)}$. The vertical line represents the tolerance chosen by cross validation and the horizontal line is the ε_{ℓ_2} error in the resulting PCE. The coefficients of the 3rd order total-degree basis were obtained using $M = 200$ uniform samples. Accuracy is measured using the ℓ_2 norm computed at 100,000 Latin-hypercube samples (see Section 5).

using the hyper-parameters $\boldsymbol{\gamma}$. The remaining data $\Xi_v = \Xi_k$ is then used to estimate the prediction error. To estimate the hyper-parameters $\boldsymbol{\gamma}$ we search over a set of possible values for $\boldsymbol{\gamma}$ and select $\boldsymbol{\gamma} = \arg\min_{\boldsymbol{\gamma}} e_{\mathrm{cv}}(\boldsymbol{\gamma})$.

Fig. 2 presents a typical example illustrating the change in the ℓ_2 error of a PCE with fixed degree, as the tolerance ε is decreased. The figure also plots the cross validation error which is a good indicator of the ℓ_2 error behavior. The vertical line represents the tolerance chosen by cross validation and the horizontal line is the ε_{ℓ_2} error in the resulting PCE. The result shown is typical. There is a bias (underestimation of ε_{ℓ_2}) in the cross validation estimate, yet despite this bias cross validation consistently chooses a tolerance that produces a near minimal error.

We remark that 10-fold cross validation can over-estimate the true approximation error, however we found that this was not a significant issue for the functions we tested, which is consistent with other studies [9,28]. One could use a larger number of folds, for example $K = M$ (leave-one-out cross validation), however any reduction in bias will likely be accompanied by greater variance [26]. Moreover setting K much larger than 10 significantly increases the computational effort required to perform cross validation.

3.2. Recoverability of ℓ_1-minimization

The ability of ℓ_1-minimization to accurately determine the large coefficients of the PCE is determined by the properties of the matrix $\boldsymbol{\Phi}$ and the sparsity of PCE representation of the model response $f(\boldsymbol{\xi})$. The sparsity of the model response is dependent on the PCE basis. In this paper we take the typical approach [7,20] and adopt the Askey scheme [46], which fixes the polynomial basis based on the distribution of the uncertain variables, thereby facilitating analytical computation of moments from the PCE and sensitivity analysis [40]. Using the Askey scheme the PCE basis and thus the sparsity of the model response is fixed, however the properties of the $\boldsymbol{\Phi}$ are influenced by the selection of the realizations $\{\boldsymbol{\xi}_i\}_{i=1}^M$ and the truncation Λ.

Mutual coherence is one measure often used to indicate the ability of ℓ_1-minimization to find a sparse solution. The mutual coherence of a matrix $\tilde{\boldsymbol{\Phi}} \in \mathbb{R}^{M \times N}$ with columns $\tilde{\phi}_j$ is

$$\mu(\boldsymbol{\Phi}) = \max_{1 < j < k \leq N} \frac{\left| \tilde{\phi}_j^T \tilde{\phi}_k \right|}{\left\| \tilde{\phi}_j \right\|_2 \left\| \tilde{\phi}_k \right\|_2} \tag{7}$$

and is a measure of the maximum correlation between any two columns in the matrix. ℓ_1-minimization will obtain a better estimate of the PCE coefficients if the mutual coherence of $\boldsymbol{\Phi}$ is small. Intuitively, if two columns are closely correlated the mutual coherence will be large and it will be impossible, in general, to distinguish whether the energy in the signal comes from one or the other.

The *restricted isometry property* (RIP) [13], quantified by the restricted isometry constant δ is another measure of the recoverability of the matrix $\boldsymbol{\Phi}$. For each $s = 1, 2, \ldots$ the isometry constant δ_s of a matrix $\boldsymbol{\Phi}$ is the smallest number such that

$$(1 - \delta_s) \|\boldsymbol{\alpha}_s\|_2^2 \leq \|\boldsymbol{\Phi} \boldsymbol{\alpha}_s\|_2^2 \leq (1 + \delta_s) \|\boldsymbol{\alpha}_s\|_2^2 \tag{8}$$

for all vectors $\boldsymbol{\alpha}_s$ with s non-zero entries. This is equivalent to requiring that the eigenvalues of all Grammian matrices $\boldsymbol{\Phi}_{\Lambda_s}^\top \boldsymbol{\Phi}_{\Lambda_s}$ lie between $[1 - \delta_s, 1 + \delta_s]$, where $\boldsymbol{\Phi}_{\Lambda_s}$ are $M \times s$ submatrices of $\boldsymbol{\Phi}$. The restricted isometry property measures the ability of $\boldsymbol{\Phi}$ to preserve the lengths of s-sparse vectors. The RIP can be intuitively thought of as a measure of s-wise coherence as opposed to mutual coherence which is a measure of pair-wise coherence.
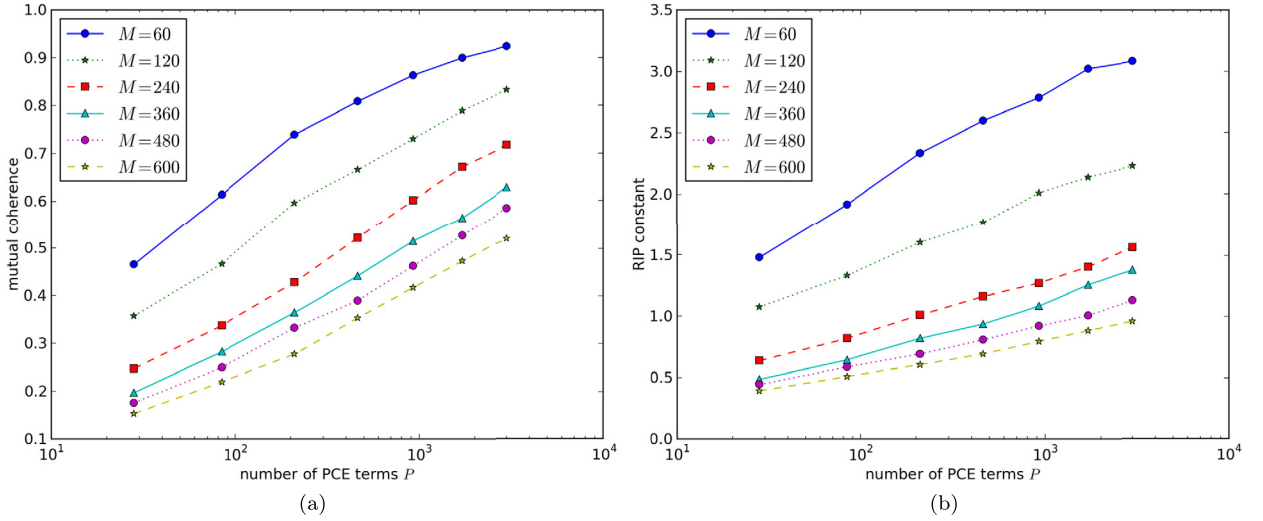
Fig. 3. Growth of mutual coherence and RIP constant δ_{10} with the number of terms P in a 6-dimensional total-degree multivariate Legendre PCE basis.
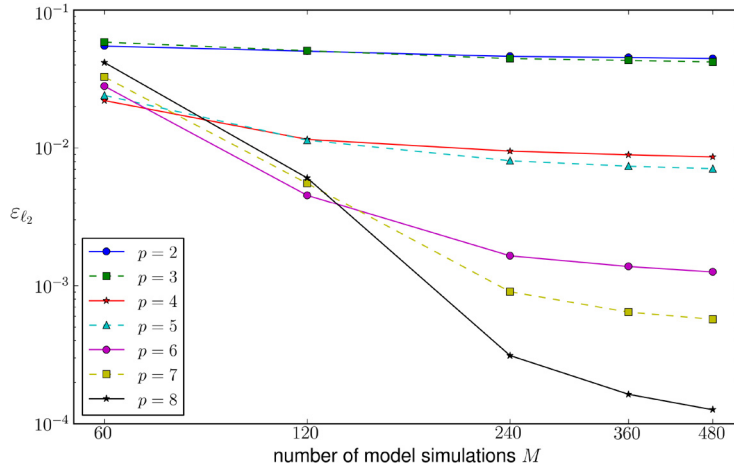

Fig. 4. The dependence of PCE accuracy on the polynomial degree p. The *degree* of the most accurate expansion is dependent on the number of LHS samples M used to construct the PCE. Results were obtained using orthogonal matching pursuit with cross validation to choose ε applied to the 6-dimensional random oscillator (12). Accuracy is measured using the ℓ_2 norm computed at 100,000 Latin-hypercube samples (see Section 5). The error reported is the average over 20 trials for each design size.

3.2.1. PCE truncation

Naively choosing a large degree p can cause a degradation in the accuracy of the PCE coefficients. Fig. 3 demonstrates that both the mutual coherence and the 10-sparse RIP constant δ_{10} of the Vandermonde matrix $\boldsymbol{\Phi}$ increases as the number of basis terms P increases.[2]

Increases in mutual coherence and RIP constant are correlated with a decrease in the accuracy of PCE coefficients recovered by ℓ_1-minimization. Fig. 4 demonstrates that, for a fixed number of samples, as the number of terms N and, consequently, the mutual coherence and RIP constant increase (see Fig. 3), the PCE recovered by ℓ_1-minimization becomes less accurate. Fig. 4 also illustrates that the accuracy of the PCE depends upon the degree of the basis used. As the number of samples M increases, ℓ_1-minimization is able to recover more dominant coefficients and a higher degree should be used. However for a given number of samples increasing the degree does not always lead to a reduction in error. For example when $M = 60$ the PCE of total-degree $p = 4$ has the smallest error and at $M = 120$ setting $p = 7$ produces the smallest error. It is not until $M = 240$ that the highest degree basis $p = 8$ produces the smallest error. These results are consistent with the theoretical results in [20] that assert that the number of samples M needed to recover a Legendre PCE of a certain sparsity s increases with the number of terms in the PCE basis.

---

[2] The RIP constant reported here is a lower bound found by computing the eigenvalues of 10,000 randomly selected submatrices $\boldsymbol{\Phi}_{\Lambda_s}$.
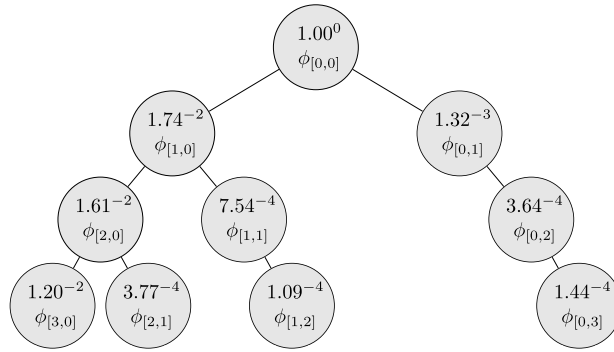
Fig. 5. Tree structure of the coefficients of a two dimensional PCE with $\Lambda_{3,1}^2$. For clarity we only depict one connection per node, but in d dimensions a node of a given degree p will be a child of up to d nodes of degree $p-1$. For example, not only is the basis $\phi_{[1,1]}$ a child of $\phi_{[1,0]}$ (as depicted) but it is also a child of $\phi_{[0,1]}$.

3.2.2. Sampling strategies and pre-conditioning

The sampling strategy used to choose the samples Ξ affects the mutual coherence and RIP of Φ and thus can impact the accuracy of the recovered polynomial chaos expansion. To date, the best sampling strategies for ℓ_1-minimization are random [38,47]. The nature of the random samples is dependent on the distribution of the random variables ξ, the number of random dimensions, and the degree of the PCE.

The accuracy of ℓ_1-minimization solutions of (5) can also be improved by the use of pre-conditioning. The pre-conditioned ℓ_1-minimization problem is given by

$$\alpha = \arg\min_{\alpha} \|\alpha\|_1 \quad \text{such that} \quad \|W\Phi\alpha - Wf\|_2 \leq \varepsilon \tag{9}$$

where $W \in \mathbb{R}^{M \times M}$ is a diagonal matrix with entries chosen to enhance the recovery properties of ℓ_1-minimization. When recovering s-sparse one-dimensional Legendre polynomials, randomly sampling $\Xi = \{\xi_m\}_{m=1}^M$ from the Chebyshev measure and choosing weights $w_{m,m} = (\pi/2)^{d/2}(1 - \xi_m^2)^{1/4}$, can result in significant increases in the accuracy of the coefficients recovered by ℓ_1-minimization [38]. In the multivariate setting, however, the benefit of pre-conditioning is less clear [48]. In this paper all numerical results presented are generated without pre-conditioning.

4. Iterative basis selection

When the coefficients of a PCE can be well approximated by a sparse vector, ℓ_1-minimization is extremely effective at recovering the coefficients of that PCE. It is possible, however, to further increase the efficacy of ℓ_1-minimization by leveraging realistic models of structural dependencies between the values and locations of the PCE coefficients. For example [3, 21,30] have successfully increased the performance of ℓ_1-minimization when recovering wavelet coefficients that exhibit a tree-like structure. In this vein, we propose an algorithm for identifying the large coefficients of PC expansions that form a semi-connected subtree of the PCE coefficient tree.

The coefficients of polynomial chaos expansions often form a multi-dimensional tree. Given an ancestor basis term ϕ_λ of degree $\|\lambda\|_1$ we define the indices of its children as $\lambda + e_k$, $k = 1, \ldots, d$, where $e_k = (0, \ldots, 1, \ldots, 0)$ is the unit vector co-directional with the k-th dimension. An example of a typical PCE tree is depicted in Fig. 5. In this figure, as often in practice, the magnitude of the ancestors of a PCE coefficient is a reasonable indicator of the size of the child coefficient. In practice, some branches (connections) between levels of the tree may be missing. We refer to trees with missing branches as semi-connected trees.

In the following we present a method for estimating PCE coefficients that leverages the tree structure of PCE coefficients to increase the accuracy of coefficient estimates obtained by ℓ_1-minimization.

4.1. Algorithm

Typically ℓ_1-minimization is applied to an a priori chosen and fixed basis set Λ. However the accuracy of coefficients obtained by ℓ_1-minimization can be increased by adaptively selecting the PCE basis.

To choose a basis for ℓ_1-minimization we employ a four step iterative procedure involving restriction, expansion, identification and selection. Pseudo code for the iterative basis selection procedure is outlined in Algorithm 1 and a graphical version of the algorithm is depicted in Fig. 6. To emphasize the four stages of the algorithm, each stage is highlighted using the same colors in both Algorithm 1 and Fig. 6.

To initiate the basis selection algorithm, we first define a basis set $\Lambda^{(0)}$ and use ℓ_1-minimization to identify the largest coefficients $\alpha^{(0)}$. The choice of $\Lambda^{(0)}$ can sometimes affect the performance of the basis selection algorithm. For example, if $\Lambda^{(0)}$ contains insufficient terms, for a given sample size M, then the initial basis will not be able to adequately inform the
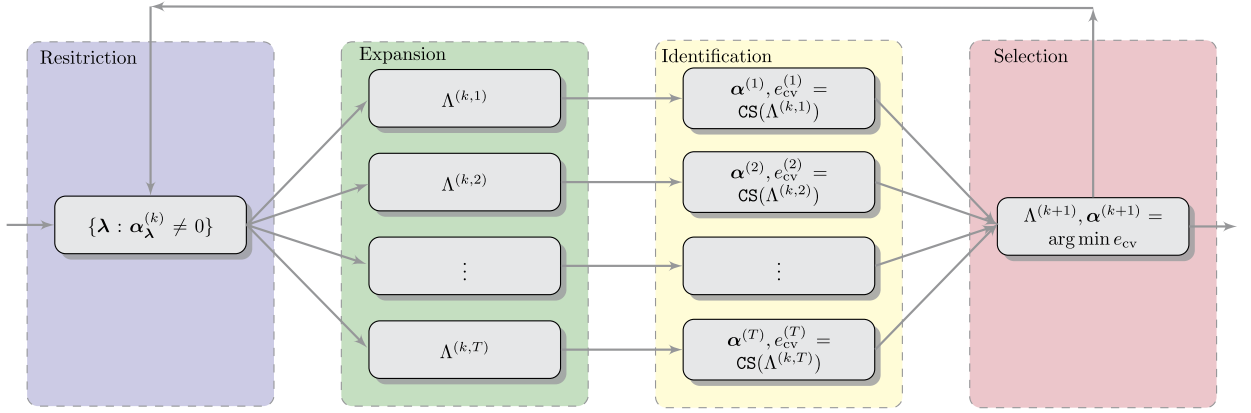
Fig. 6. Graphical depiction of the basis adaptation algorithm.

Algorithm 1: $\Lambda^\star, \boldsymbol{\alpha}^\star = \text{BASIS\_SELECTION}[\boldsymbol{\Phi}, \mathbf{f}, \varepsilon]$.

$\Lambda^\star = \Lambda^{(0)} = \Lambda^d_{p,1} = \arg\min_{\Lambda^d_{p,1} \in \{\Lambda^d_{1,1}, \Lambda^d_{2,1}, \dots\}} \left| |\Lambda^d_{p,1}| - 10M \right|$

$\boldsymbol{\alpha}^{(0)}, e^{(0)}_{\text{cv}} = \ell_1\text{-minimization}[\boldsymbol{\Phi}(\Lambda^{(0)}), \mathbf{f}]$

$T = 3, e^\star_{\text{cv}} = \infty, k = 1$

while *TRUE* **do**

    $e^{(k)}_{\text{cv}} = \infty$

    $\Lambda^{(k,0)} = \{\boldsymbol{\lambda} : \boldsymbol{\lambda} \in \Lambda^{(k-1)}, \boldsymbol{\alpha}^{(k)}_{\boldsymbol{\lambda}} \neq 0\}$

    for $t \in \{1, \dots, T\}$ **do**

        $\Lambda^{(k,t)} = \text{EXPAND}[\Lambda^{(k,t-1)}]$

        $\boldsymbol{\alpha}^{(k,t)}, e^{(k,t)}_{\text{cv}} = \ell_1\text{-minimization}[\boldsymbol{\Phi}(\Lambda^{(k,t)}), \mathbf{f}]$

        if $e^{(k,t)}_{\text{cv}} < e^{(k)}_{\text{cv}}$ **then**

            $e^{(k)}_{\text{cv}} = e^{(k,t)}_{\text{cv}}, \boldsymbol{\alpha}^{(k)} = \boldsymbol{\alpha}^{(k,t)}, \Lambda^{(k)} = \Lambda^{(k,t)}$

        end

    end

    if $e^{(k)}_{\text{cv}} > e^\star_{\text{cv}}$ **then**

        TERMINATE

    end

    $\boldsymbol{\alpha}^\star = \boldsymbol{\alpha}^{(k)}, \Lambda^\star = \Lambda^{(k)}, e^\star_{\text{cv}} = e^{(k)}_{\text{cv}}$

end

---

iterative procedure. Conversely, if the cardinality of the initial basis is to large, the increased degradation of the properties of the matrix $\boldsymbol{\Phi}$, such as mutual coherence (see Section 3.2), will result in poor recovery of the initial PCE coefficients which will limit the effectiveness of the basis selection procedure. We found a good choice to be $\Lambda^{(0)} = \Lambda_{p,1}$, where p is the degree that gives $|\Lambda^d_{p,1}|$ closest to $10M$, i.e. $\Lambda^d_{p,1} = \arg\min_{\Lambda^d_{p,1} \in \{\Lambda^d_{1,1}, \Lambda^d_{2,1}, \dots\}} \left| |\Lambda^d_{p,1}| - 10M \right|$.

Given a basis $\Lambda^{(k)}$ and corresponding coefficients $\boldsymbol{\alpha}^{(k)}$ we reduce the basis to a set $\Lambda^{(k)}_\varepsilon$ containing only the terms with non-zero coefficients. This restricted basis is then expanded T times using an algorithm which we will describe in Section 4.1.1. ℓ_1-minimization is then applied to each of the expanded basis sets $\Lambda^{(k,t)}$ for $t = 1, \dots, T$. Each time ℓ_1-minimization is used, we employ cross validation to choose ε. Therefore, at every basis set considered during the evolution of the algorithm we have a measure of the expected accuracy of the PCE coefficients. At each step in the algorithm the basis set that results in the lowest cross validation error is chosen. Typically the algorithm will choose the basis set that corresponds to $T = 1$ or $T = 2$.

4.1.1. Basis expansion

Define $\{\boldsymbol{\lambda} + \mathbf{e}_j : 1 \leq j \leq d\}$ the forward neighborhood of an index $\boldsymbol{\lambda}$ and similarly let $\{\boldsymbol{\lambda} - \mathbf{e}_j : 1 \leq j \leq d\}$ denote the backward neighborhood. To expand a basis set Λ we must first find the forward neighbors $\mathcal{F} = \{\boldsymbol{\lambda} + \mathbf{e}_j : \boldsymbol{\lambda} \in \Lambda, 1 \leq j \leq d\}$ of all indices $\boldsymbol{\lambda} \in \Lambda$. The expanded basis is then given by

$$\Lambda^+ = \Lambda \cup \mathcal{A}, \quad \mathcal{A} = \{\boldsymbol{\lambda} : \boldsymbol{\lambda} \in \mathcal{F}, \boldsymbol{\lambda} - \mathbf{e}_n \in \Lambda \text{ for } 1 \leq n \leq d, \lambda_k > 1\}$$

where we have used the following admissibility criteria

$$\boldsymbol{\lambda} - \mathbf{e}_n \in \Lambda \text{ for } 1 \leq n \leq d, \lambda_k > 1 \tag{10}$$
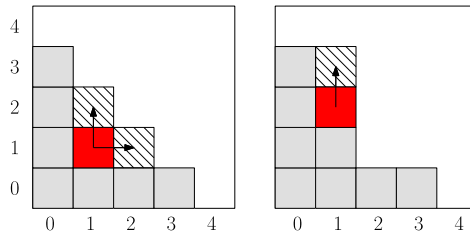
Fig. 7. Identification of the admissible indices of an index (red). The indices of the current basis Λ are gray and admissible indices are striped. An index is admissible only if its backwards neighbors exists in every dimension. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to target PCE basis indices that are likely to have large PCE coefficients. A forward neighbor is admissible only if its backward neighbors exist in all dimensions. If the backward neighbors do not exist then ℓ_1-minimization has previously identified that the coefficients of these backward neighbors are negligible.

The admissibility criterion is explained graphically in Fig. 7. In the left graphic, both children of the current index are admissible, because its backwards neighbors exist in every dimension. In the right graphic only the child in the vertical dimension is admissible, as not all parents of the horizontal child exist.

At the k-th iteration of Algorithm 1, ℓ_1-minimization is applied to $\Lambda^{(k-1)}$ and used to identify the significant coefficients of the PCE and their corresponding basis terms $\Lambda^{(k,0)}$. The set of non-zero coefficients $\Lambda^{(k,0)}$ identified by ℓ_1-minimization is then expanded. The EXPAND routine expands an index set by one polynomial degree, but sometimes it may be necessary to expand the basis $\Lambda^{(k)}$ more than once.[3] To generate these higher degree index sets EXPAND is applied recursively to $\Lambda^{(k,0)}$ up to a fixed number of T times. Specifically, the following sets are generated

$$\Lambda^{(k,t)} = \Lambda^{(k,t-1)} \cup \{\boldsymbol{\lambda} : \boldsymbol{\lambda} - \mathbf{e}_n \in \Lambda^{(k,t-1)}, 1 \leq n \leq d, \lambda_n > 1\}.$$

As the number of expansion steps T increases the number of terms in the expanded basis increases rapidly and degradation in the performance of ℓ_1-minimization can result (this is similar to what happens when increasing the degree of a total-degree basis). To avoid degradation of the solution, we use cross validation to choose the number of inner expansion steps $t \in [1, T]$.

5. Numerical examples

In this section we use several numerical tests to demonstrate the benefit of the basis selection method. In each example we seek a PCE approximation to a model output given a set of uncertain parameters with a known range or distribution.

We compare the approximations constructed using basis selection against those constructed using a non-adaptive strategy. The non-adaptive strategy consists of generating basis sets $\Lambda^d_{2,1}, \ldots, \Lambda^d_{p,1}$ where p is the degree that produces the basis set with a cardinality closest to 100,000. ℓ_1-minimization is then applied to this basis with a cross validation tolerance search to compute the non-zero polynomial coefficients. The resulting basis with the lowest cross validation error is chosen to be the final approximation.

We also compare the non-adaptive and basis selection methods against OMP using a basis oracle. Given an exact PCE representation of a scalar function and M function samples, the oracle basis consists of $N = M$ basis terms corresponding to the N largest coefficients of the exact PCE. For the examples that follow, the exact PCE are not known analytical and so they are computed to machine precision using a dimension adaptive sparse grid. The oracle basis, defined here, will be close to optimal and therefore will serve as a good estimate of the maximum accuracy that can be gained from the use of basis selection.

In order to construct a PCE approximation, the sample design has to be specified. By a design we mean the choice of sample size, M, and the selection $\Xi = \{\xi_i\}_{i=1}^M$. Here we opt for uniform random samples of size M. For small sample sizes M the selection of Ξ significantly affects the performance of any approximation method. Therefore, for each M, twenty different designs are used to build a PCE and the median first and third quartiles and maximum and minimum of the resulting errors are reported using box and whisker plots. We did investigate the utility of using samples drawn from the Chebyshev measure but found that there was no consistent benefit. Even in the cases for which a benefit was observed, the improvement was small relative to the benefit gained from using basis adaptation. This finding is consistent with [48].

To measure the performance of an approximation, we will use the ℓ_2 error (RMSE). Specifically given a set of $Q = 100,000$ Latin-hypercube samples $\Xi_{\text{test}} = \{\xi^{(i)}\}_{i=1}^Q \in I_\xi$ and samples of the true function $f(\xi^{(i)})$ and the PCE approximation $\hat{f}(\xi^{(i)})$ we compute

---

[3] The choice of $T > 1$ enables the basis selection algorithm to be applied to semi-connected tree structures as well as fully connected trees. Setting $T > 1$ allows us to prevent premature termination of the algorithm if most of the coefficients of the children of the current set $\Lambda^{(k)}$ are small but the coefficients of the children's children are not.

$$\varepsilon_{\ell_2} = \left(\frac{1}{Q} \sum_{i=1}^{Q} |\hat{f}(\boldsymbol{\xi}^{(i)}) - f(\boldsymbol{\xi}^{(i)})|^2 \right)^{1/2}$$

Note in all examples presented using Legendre polynomials we transform each d dimensional parameter domain $I_{\boldsymbol{\xi}}$, build points $\boldsymbol{\Xi}$ and test points $\boldsymbol{\Xi}_{\text{test}}$ to $[-1, 1]^d$.

5.1. Algebraic test function

Consider the algebraic corner-peak test function [24]

$$f_{\text{CP}}(\mathbf{x}) = \left(1 + \sum_{k=1}^{d} c_k \xi_k \right)^{-(d+1)}, \quad \boldsymbol{\xi} \in [0, 1]^d \tag{11}$$

This function provides a flexible test that can be used to identify the strengths of the proposed algorithm. Specifically, the coefficients c_k can be used to control the effective dimensionality and the compressibility of these functions. Here we will examine performance using three different choices of $\mathbf{c} = (c_1, \ldots, c_d)^T$, specifically

$$c_k^{(1)} = \frac{k - \frac{1}{2}}{d}, \quad c_k^{(2)} = \frac{1}{k^2} \quad \text{and} \quad c_k^{(3)} = \exp\left(\frac{k \log(10^{-8})}{d} \right), \quad k = 1, \ldots, d$$

normalizing such that $\sum_{k=1}^{d} c_k = 0.25$. The coefficients $\mathbf{c}^{(1)}$, $\mathbf{c}^{(2)}$ and $\mathbf{c}^{(3)}$ represent increasing levels of anisotropy and decreasing effective dimensionality. Anisotropy refers to the dependence of the function variability, often measured through variance, on individual parameter dimensions ξ_n. When a function is strongly anisotropic, the majority of the function variance can be attributed to a small set of dimensions. The size of this subset is referred to as the effective dimension.

The performance of the basis selection method is dependent on the compressibility of the model response in the chosen basis. Fig. 8 plots the ε_{ℓ_2} error in the polynomial approximations for increasing number of model evaluations. For all three levels of anisotropy the adaptive method produces an expansion no worse than the non-adaptive method, for the same sample size. When anisotropy is introduced the accuracy of basis selection increases relative to the non-adaptive method. The stronger the anisotropy the better the relative performance. Fig. 8 also plots the PCE obtained using OMP with an oracle basis. When very weak anisotropy is present $\mathbf{c}^{(1)}$, there is little that can be gained by using a well chosen basis, as evident by the lack of separation between the three convergence curves. However as the strength of the anisotropy is increased the effect of the oracle basis on accuracy becomes much more apparent. When strong anisotropy, $\mathbf{c}^{(3)}$, is present, basis selection is able to obtain the same accuracy as the oracle without a priori information on the truncation of the basis which is required by the oracle. In the moderately anisotropic $\mathbf{c}^{(2)}$ case basis selection does not perform as well as the oracle but does still perform better than the non-adaptive method.

To understand the correlation between anisotropy and the performance of the basis selection method we must consider the structure of the PCE coefficients induced by varying $\mathbf{c}$. Fig. 8(d) plots the decay of the PCE coefficients when sorted by magnitude. As anisotropy increases, so does the rate of decay of the sorted PCE coefficients.

It is the strength of decay that controls the performance of basis selection. When the rate of decay is high then the function is more compressible and thus better suited to being approximated using ℓ_1-minimization. Anisotropy will often result in compressible coefficients, but it is conceptually possible for models to be compressible without being anisotropic. For some problems such as the elliptic Poisson equation the decay of the PCE coefficients can be calculated a priori [1,4,15] but unfortunately in practice, the decay of the PCE coefficients of a model cannot be determined ahead of time. A practical means of identifying the coefficient decay regime would be very useful but is beyond the scope of this paper.

Not only does the rate of coefficient decay affect performance, but so to does the ability of the PCE basis Λ to represent the target function. For example if the 'true' PCE has large high degree terms with large coefficients but the basis Λ does not have these high degree terms then the PCE obtained using Λ will not be as accurate as a PCE obtained using a basis that included the important high degree terms.

Fig. 9 plots the exact PCE coefficients of the algebraic test function using $\mathbf{c}^{(1)}$ and $\mathbf{c}^{(3)}$. The 'exact' were coefficients obtained using a dimension-adaptive sparse grid with 100,000 samples, which resulted in an approximation error below 10^{-8}. In Fig. 9(a), the random variables contribute similarly to the total variance of f_{CP} and so the dominant coefficients are concentrated in the lower degree terms of the PCE, thus a total-degree basis set will perform as well as any alternative. In comparison, the importance of the dimensions of the function shown in Fig. 9(b) decay exponentially with dimension, which results in higher-degree terms with large coefficients in some dimensions. In this coefficient regime, if ℓ_1-minimization can only be applied with a low degree polynomial (which is true when using a total-degree basis), the accuracy of the resulting PCE, for a given number of samples M, will not be as high as a PCE constructed using a basis that includes the dominant high-degree terms. Basis selection will typically allow identification and recovery of more high-degree coefficients than would be possible if using a total-degree basis.
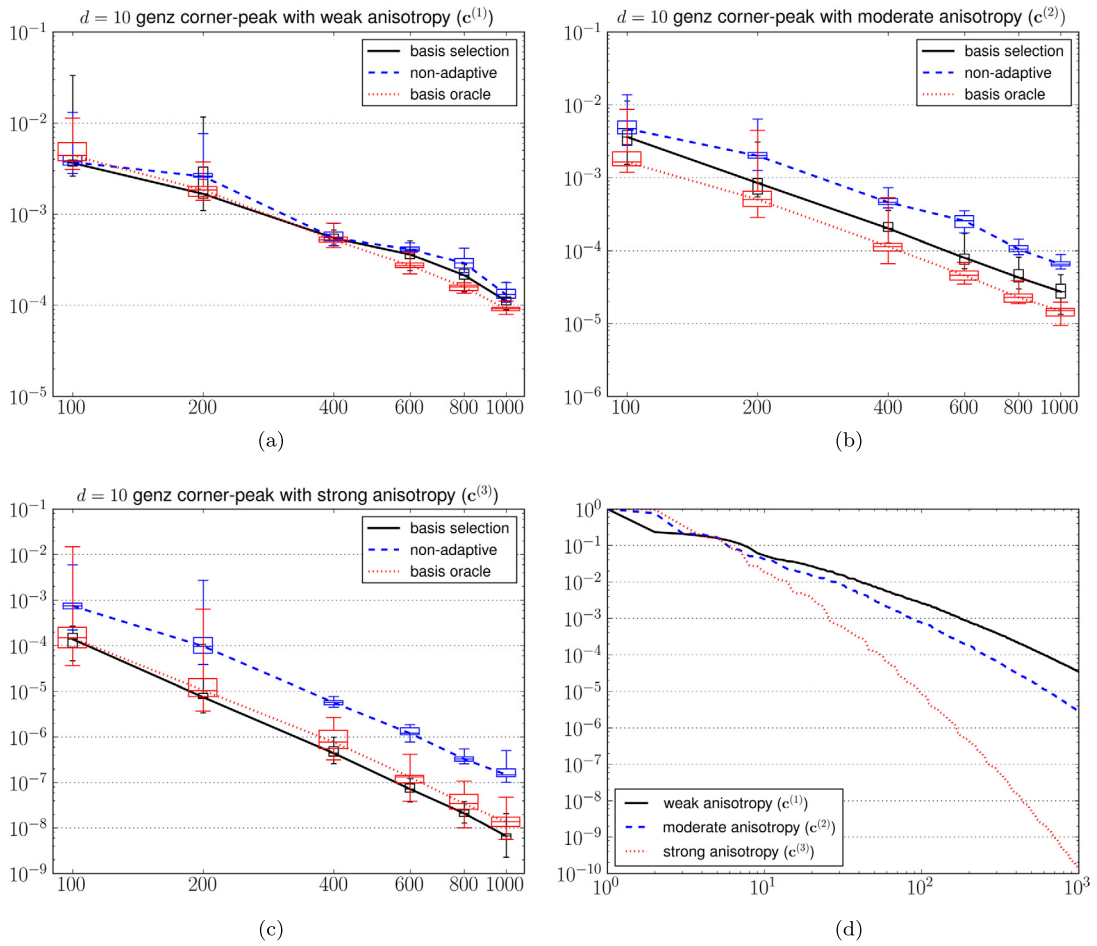
Fig. 8. Convergence of the RMSE, with respect to increasing design size M, in the Legendre PCE approximation of the Genz corner-peak function (11) for the three coefficient regimes: (a) $\mathbf{c}^{(1)}$, (b) $\mathbf{c}^{(2)}$, (c) $\mathbf{c}^{(3)}$. The error bars (whiskers) represent the minimum and maximum error over the 20 trials for each design size M and the boxes represent the 1st, 2nd (median) and 3rd quartiles. (d) PCE coefficients, sorted by magnitude, for the three coefficient regimes.
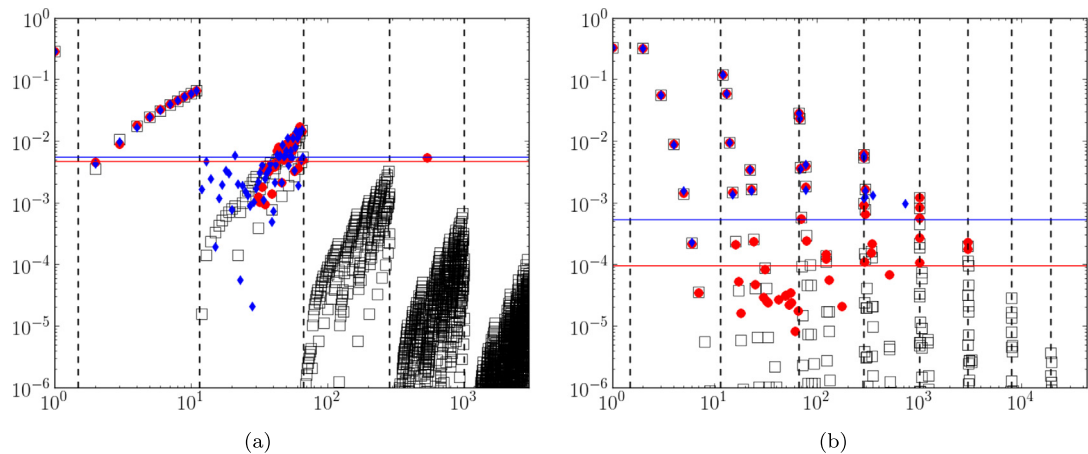

Fig. 9. Comparison of PCE coefficients of $\mathbf{c}^{(1)}$ (a) and $\mathbf{c}^{(3)}$ (b). The black squares represent the 'exact' coefficients, the red circles the non-zero coefficients of the basis selection method and the blue diamonds are the non-zero coefficients recovered by the non-adaptive approach. The indices on the x axis are sorted lexicographically by degree. The dashed vertical lines separate the PCE terms into degrees. The horizontal lines represent the ℓ_2 error in the basis selection and non-adapted PCE. The identification of more terms accurately with basis selection results in a smaller error. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
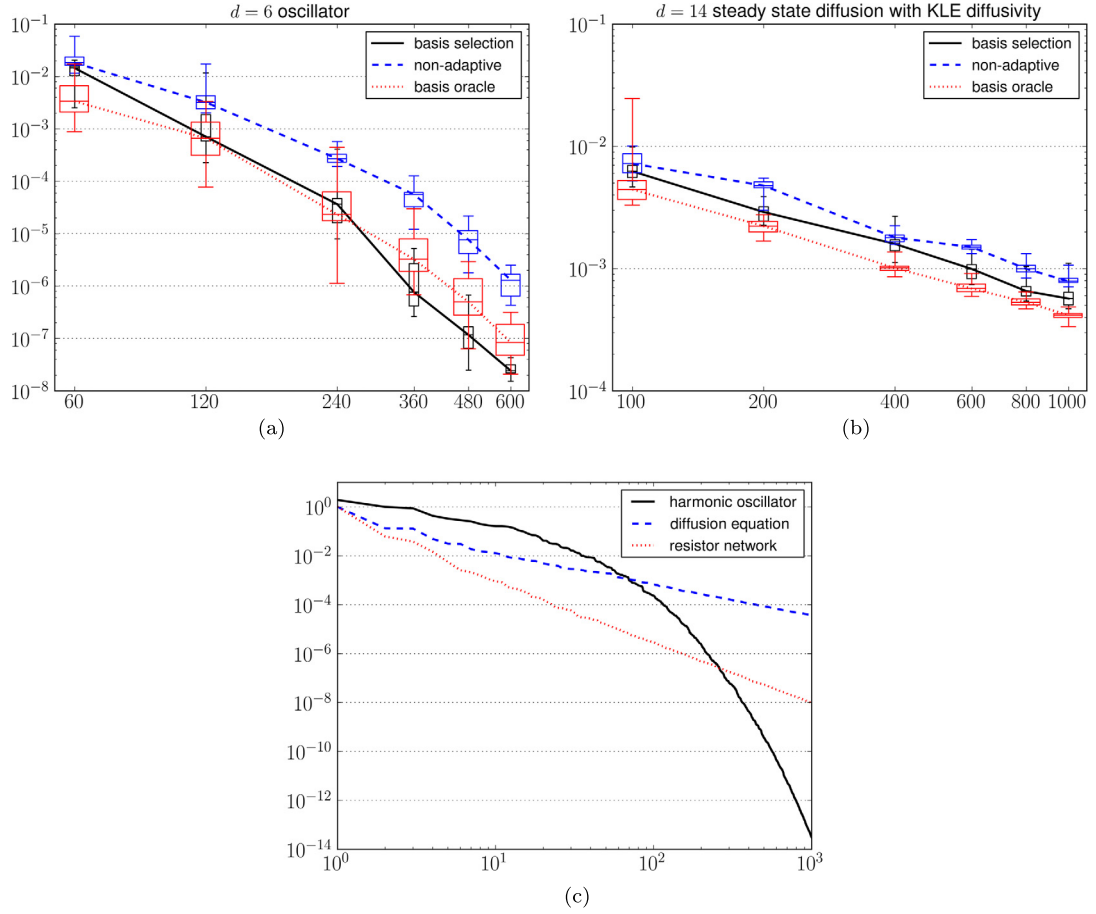
Fig. 10. Convergence of the RMSE, with respect to increasing design size M, in the Legendre PCE approximation of (a) the harmonic oscillator (12) and (b) the solution $u(1/3, \boldsymbol{\xi})$ of the diffusion equation (14). The error bars (whiskers) represent the minimum and maximum error over the 20 trials for each design size M and the boxes represent the 1st, 2nd (median) and 3rd quartiles. (c) PCE coefficients, sorted by magnitude, for the harmonic oscillator, the diffusion equation and the resistor network.

5.2. Random oscillator

In this section we investigate the performance of basis selection to quantify uncertainty in a damped linear oscillator subject to external forcing with six unknown parameters. That is,

$$\frac{d^2x}{dt^2}(t, \boldsymbol{\xi}) + \gamma \frac{dx}{dt} + kx = f \cos(\omega t), \tag{12}$$

subject to the initial conditions

$$x(0) = x_0, \quad \dot{x}(0) = x_1, \tag{13}$$

where we assume the damping coefficient γ, spring constant k, forcing amplitude f and frequency ω, and the initial conditions x_0 and x_1 are all uncertain. We solve (12) analytically to avoid consideration of discretization errors in our study.

Defining $\boldsymbol{\xi} = (\gamma, k, f, \omega, x_0, x_1)$ let $\xi_1 \in [0.08, 0.12]$, $\xi_2 \in [0.03, 0.04]$, $\xi_3 \in [0.08, 0.12]$, $\xi_4 \in [0.8, 1.2]$, $\xi_5 \in [0.45, 0.55]$, $\xi_6 \in [-0.05, 0.05]$. For any parameter realization in $I_{\boldsymbol{\xi}}$ the harmonic oscillator will be underdamped. In the following, we choose our quantity of interest to be the position $x(t)$ of the oscillator at $t = 20$ seconds.

Fig. 10(a) depicts the error in the Legendre PCE for increasing design sizes M. The basis selection method clearly outperforms the non-adaptive approach and produces comparable results to the oracle. Again the improvement in performance is associated with a rapid decay of the exact PCE coefficients (see Fig. 10).

5.3. Diffusion equation

In this section, we consider the heterogeneous diffusion equation in one-spatial dimension subject to uncertainty in the diffusivity coefficient. This problem has been used as a benchmark in other works [20,49]. Attention is restricted to
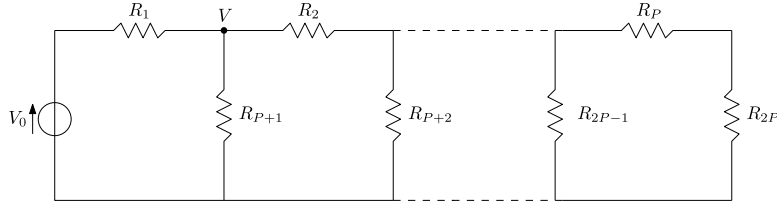
Fig. 11. Resistor network comprised of $d = 2P$ resistances R_i, $i = 1, \ldots, d$ of uncertain ohmage and the network is driven by a voltage source providing a known potential V_0. We are interested in determining voltage at V.

one-dimensional physical space to avoid unnecessary complexity. The procedure described here can easily be extended to higher physical dimensions. Consider the following problem with $d \geq 1$ random dimensions:

$$-\frac{d}{dx}\left[a(x, \boldsymbol{\xi}) \frac{du}{dx}(x, \boldsymbol{\xi})\right] = 1, \quad (x, \boldsymbol{\xi}) \in (0, 1) \times I_{\boldsymbol{\xi}} \tag{14}$$

subject to the physical boundary conditions

$$u(0, \boldsymbol{\xi}) = 0, \quad u(1, \boldsymbol{\xi}) = 0. \tag{15}$$

Furthermore, assume that the random diffusivity satisfies

$$a(x, \boldsymbol{\xi}) = \bar{a} + \sigma_a \sum_{k=1}^{d} \sqrt{\lambda_k} \phi_k(x) \xi_k, \tag{16}$$

where $\{\lambda_k\}_{k=1}^{d}$ and $\{\phi_k(x)\}_{k=1}^{d}$ are, respectively, the eigenvalues and eigenfunctions of the squared exponential covariance kernel

$$C_a(x_1, x_2) = \exp\left[-\frac{(x_1 - x_2)^2}{l_c^2}\right].$$

The variability of the diffusivity field (16) is controlled by σ_a and the correlation length l_c which determines the decay of the eigenvalues λ_k. Here we approximate the solution $u(1/3, \boldsymbol{\xi})$ with $\bar{a} = 0.1$, $d = 14$, $\sigma_a = 0.03$, $l_c = 1/5$, while the uncertain inputs $\xi_k \in [-1, 1]$, $k = 1, \ldots, d$ are independent and uniformly distributed random variables. We solve the model (14) using quadratic finite elements with a high enough spatial resolution to neglect discretization errors in our analysis.

Fig. 10(b) plots the error in Legendre PCE approximations built using increasing design sizes M. There is negligible difference between basis selection and the non-adaptive strategy, but there is also negligible difference between these methods and the oracle, indicating there is not much improvement that can in principle be gained from basis selection. The negligible improvement is due to the fact that the 'exact' PCE is not very compressible, as can be seen from Fig. 10. The lack of compressibility means that many coefficients are of similar magnitude and thus ℓ_1-minimization in any form is not very effective.

5.4. Resistor network

As our last example, consider the electrical resistor network shown in Fig. 11. The network is comprised of $d = 2P$ resistances of uncertain ohmage and the network is driven by a voltage source providing a known potential V_0. We are interested in determining how the voltage V shown in the figure depends on the $d = 2P$ resistances, which we take as random parameters uniformly distributed in the interval $\xi_k \in [1 - \varepsilon, 1 + \varepsilon]$, $k = 1, \ldots, d$. This function is anisotropic. The effect of the resistors on the voltage will decay with distance (in terms of the number of preceding resistors) from the point V. In this example we set $d = 40$ ($P = 20$) and $d = 80$ ($P = 40$), take the maximum perturbation to be $\varepsilon = 0.1$ and set the reference potential $V_0 = 1$.

Fig. 12 shows the error in the Legendre PCE for increasing design sizes M. In both cases the basis selection method produces a PCE that is significantly more accurate than the PCE produced by the non-adaptive strategy. The basis selection method provides comparable results to the approximately optimal oracle.

5.5. Gradient-enhanced ℓ_1-minimization

Typical ℓ_1-minimization, when used for PCE approximation, attempts to find solutions to

$$\boldsymbol{\Phi} \boldsymbol{\alpha} \approx \mathbf{f}$$

where denotes the Vandermonde matrix with entries $\boldsymbol{\Phi}_{ij} = \phi_j(\boldsymbol{\xi}_i)$, $i = 1, \ldots, M$, $j = 1, \ldots, N$. If gradients of the model f with respect to the random variables $\boldsymbol{\xi}$ are known, then one can enhance the accuracy of the PCE by finding a solution to
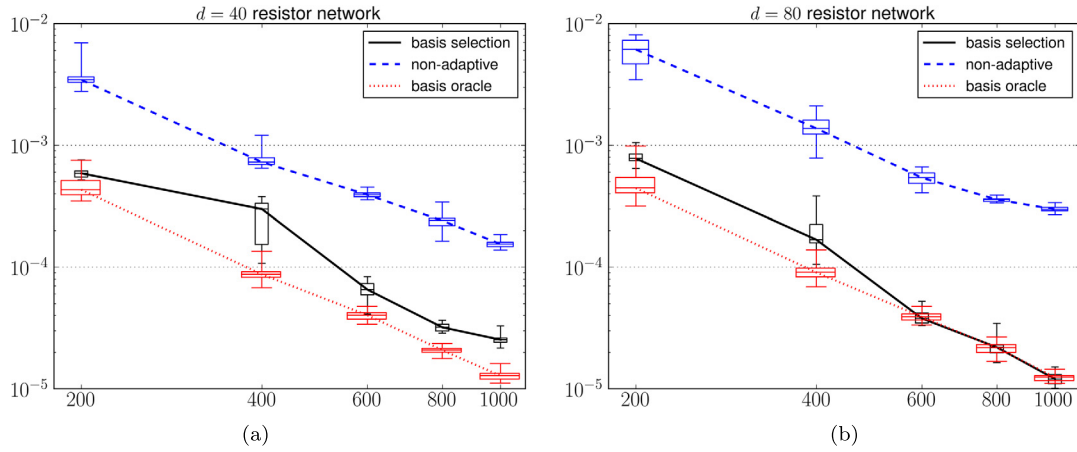
Fig. 12. Convergence of the RMSE, with respect to increasing design size M, in the Legendre PCE approximation of the resistor network. The error bars (whiskers) represent the minimum and maximum error over the 20 trials for each design size M and the boxes represent the 1st, 2nd (median) and 3rd quartiles.

$$\boldsymbol{\Phi}_\partial \boldsymbol{\alpha} \approx \mathbf{f}_\partial \quad \text{where} \quad \boldsymbol{\Phi}_\partial = \begin{bmatrix} \boldsymbol{\Phi} \\ \frac{\partial \boldsymbol{\Phi}}{\partial \xi_1} \\ \vdots \\ \frac{\partial \boldsymbol{\Phi}}{\partial \xi_d} \end{bmatrix}, \quad \mathbf{f}_\partial = \begin{bmatrix} \mathbf{f} \\ \frac{\partial \mathbf{f}}{\partial \xi_1} \\ \vdots \\ \frac{\partial \mathbf{f}}{\partial \xi_d} \end{bmatrix}$$

and $(\frac{\partial \boldsymbol{\Phi}}{\partial \xi_n})_{ij} = \frac{\partial \phi_j}{\xi_n}(\boldsymbol{\xi}_i)$ and $(\frac{\partial \mathbf{f}}{\partial \xi_n})_i = \frac{\partial f}{\xi_n}(\boldsymbol{\xi}_i)$, $i = 1, \dots, M$, $j = 1, \dots, N$, $n = 1, \dots, d$. To find a solution we again use basis pursuit denoising and solve

$$\boldsymbol{\alpha}^\star = \underset{\boldsymbol{\alpha}}{\arg\min} \, \|\boldsymbol{\alpha}\|_1 \quad \text{such that} \quad \|\boldsymbol{\Phi}_\partial \boldsymbol{\alpha} - \mathbf{f}_\partial\|_2 \le \varepsilon$$

This gradient based formulation consists of $M(d+1)$ equations that match both function values and gradients, in comparison to (5) which consists of only M equations that match function values.

Fig. 13 demonstrates the utility of using gradient data to build PCE approximations of the corner-peak function (11) with $d = 10$. Unlike the previous figures in this paper, the horizontal axis is no longer the number of model runs but rather the computational cost. We assume that running the model to only obtain function values costs one computational unit and running the model to obtain both function values and all gradients components requires two units. For example, adjoint methods for differential equations can be used to obtain all gradients at a cost less than or equal to the cost of one forward model run.

Despite the extra computational cost required to obtain gradients, the use of gradients improves both the PCE resulting from both the non-adaptive and basis selection methods. Similar to the results presented in Section 5.1 the results shown here demonstrate that basis selection is more accurate than the non-adaptive strategy. Again the relative benefit is dependent on the rate of decay of the PCE coefficients.

Basis selection is able to make effective use of gradient information. For a design $\boldsymbol{\Xi}$ with M samples, the size of the gradient enhanced Vandermonde matrix is $M(d+1) \times N$. We see that for a given accuracy gradient-based PCE requires a factor of 4 fewer samples than the PCE based on the function values only. This is close to the optimal reduction factor of $d/2 = 5$ that can be obtained using gradients, assuming that each gradient component is as informative as a function value and the cost of computing function values with gradients is twice the cost of just computing function values.

5.6. Non-uniform model inputs

Throughout this paper, we have discussed basis selection when applied to Legendre polynomials and uniform variables. However, basis selection can also be applied to other variable/polynomial combinations. Let us once again consider the resistor network, but now let $\boldsymbol{\xi}$ be Gaussian variables with mean 1.0 and standard deviation 0.005.[4] We now draw random samples from the aforementioned Gaussian distribution to form $\boldsymbol{\Xi}$ and run the model at each sample to obtain $f(\boldsymbol{\Xi})$. Fig. 14 demonstrates that the advantages of basis selection are also present when we compute PCE approximations with non-uniform random variables.

Note that the error bars depicted in Fig. 14 tend to increase in size with the number of samples M. This increase is likely caused by the poor numerical conditioning of the Vandermonde matrix in (5) which increases rapidly, with M, when using

---

[4] The standard deviation is made sufficiently small to make the chance of negative resistances practically zero.
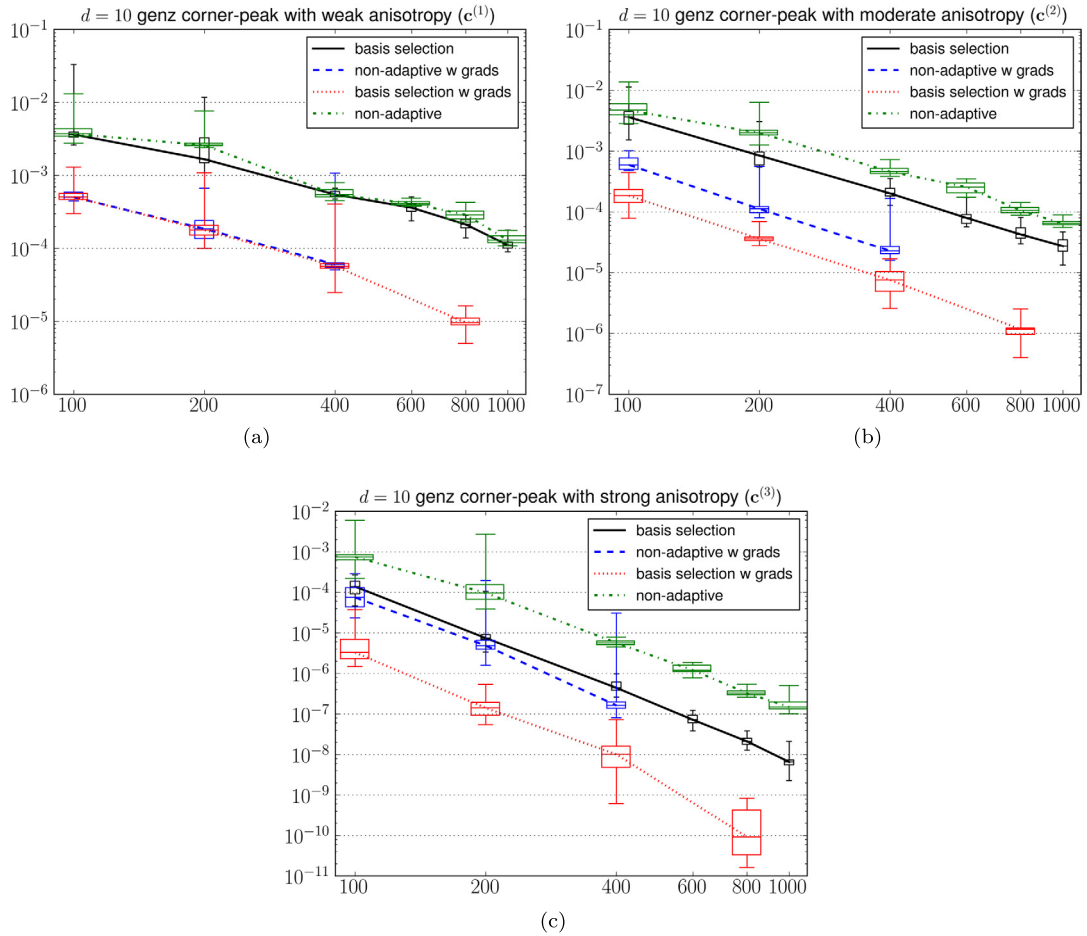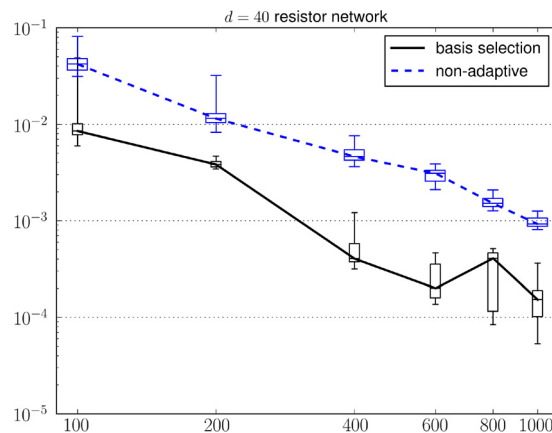
Fig. 13. Convergence of the RMSE, with respect to increasing computation units, in the gradient-enhanced Legendre PCE approximation of the corner-peak function (11) for the three coefficient regimes: (a) $\mathbf{c}^{(1)}$, (b) $\mathbf{c}^{(2)}$, (c) $\mathbf{c}^{(3)}$. The error bars (whiskers) represent the minimum and maximum error over the 20 trials for each design size M and the boxes represent the 1st, 2nd (median) and 3rd quartiles. We have assumed computing functional values at a design point is one unit and computing all gradient components is another unit. For example, the gradient enhanced basis selection approximation (dotted line) with 50 design points costs 100 units (horizontal axis value).

Fig. 14. Convergence of the RMSE, with respect to increasing design size M, in the Hermite PCE approximation of the $d = 40$ resistor network. The error bars (whiskers) represent the minimum and maximum error over the 20 trials for each design size M and the boxes represent the 1st, 2nd (median) and 3rd quartiles.

Hermite polynomials and normally distributed random variables. Development of sampling and pre-conditioning strategies for normal variables is an important area of future research, but is beyond the scope of this paper.

6. Conclusions

In this paper we present a basis selection method that can be used with ℓ_1-minimization to adaptively determine the large coefficients of polynomial chaos expansions (PCE). The method attempts to identify structure in the coefficients of a PCE and only applies ℓ_1-minimization to those terms believed to have large coefficients. The adaptive construction produces anisotropic basis sets that have more terms in important dimensions and limits the number of unimportant terms which increase mutual coherence and thus degrade the performance of ℓ_1-minimization. The basis selection method produces, for a given computational budget, a more accurate PCE than would be obtained if the basis is fixed a priori. The important features and the accuracy of basis selection are demonstrated with a number of numerical examples. Specifically we show that in high dimensions, for which high-order total-degree PCE bases are infeasible, basis selection allows accurate identification of high-order terms that cannot be captured by low-order total-degree expansions. We demonstrate that even for lower dimensional problems, for which high-order total-degree PCE bases are feasible, basis selection still produces more accurate results than basis sets that are fixed priori. Finally, we demonstrate that basis selection can effectively leverage function gradients and be applied to PCE of non-uniform random variables.

Acknowledgements

References

[1] I. Babuska, R. Tempone, G. Zouraris, Galerkin finite element approximations of stochastic elliptic partial differential equations, SIAM J. Numer. Anal. 42 (2) (2004) 800–825.
[2] I.M. Babuska, F. Nobile, R. Tempone, A stochastic collocation method for elliptic partial differential equations with random input data, SIAM J. Numer. Anal. 45 (3) (2007) 1005–1034.
[3] R.G. Baraniuk, V. Cevher, M.F. Duarte, C. Hegde, Model-based compressive sensing, IEEE Trans. Inf. Theory 56 (4) (2010) 1982–2001.
[4] J. Beck, R. Tempone, F. Nobile, L. Tamellini, On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods, Math. Models Methods Appl. Sci. 22 (09) (2012) 1250023.
[5] S. Becker, J. Bobin, E. Candés, Nesta: a fast and accurate first-order method for sparse recovery, SIAM J. Imaging Sci. 4 (1) (2011) 1–39.
[6] Marcel Bieri, Christoph Schwab, Sparse high order {FEM} for elliptic sPDEs, in: {HOFEM07} International Workshop on High-Order Finite Element Methods, 2007, Comput. Methods Appl. Mech. Eng. 198 (1314) (2009) 1149–1170.
[7] Géraud Blatman, Bruno Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, J. Comput. Phys. 230 (6) (2011) 2345–2367.
[8] Petros Boufounos, Marco F. Duarte, Richard G. Baraniuk, Sparse signal reconstruction from noisy compressive measurements using cross validation, in: IEEE/SP 14th Workshop on Statistical Signal Processing, 2007, SSP '07, Aug. 2007, pp. 299–303.
[9] Leo Breiman, Philip Spector, Submodel selection and evaluation in regression. The x-random case, Int. Stat. Rev. 60 (3) (1992) 291–319.
[10] T.T. Cai, Lie Wang, Orthogonal matching pursuit for sparse signal recovery with noise, IEEE Trans. Inf. Theory 57 (7) (July 2011) 4680–4688.
[11] E.J. Candes, T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies?, IEEE Trans. Inf. Theory 52 (12) (Dec 2006) 5406–5425.
[12] Emmanuel J. Candés, Justin K. Romberg, Terence Tao, Stable signal recovery from incomplete and inaccurate measurements, Commun. Pure Appl. Math. 59 (8) (2006) 1207–1223.
[13] Emmanuel J. Candes, Justin K. Romberg, Terence Tao, Stable signal recovery from incomplete and inaccurate measurements, Commun. Pure Appl. Math. 59 (8) (2006) 1207–1223.
[14] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (1) (January 2001) 129–159.
[15] A. Cohen, R. DeVore, C. Schwab, Convergence rates of best n-term Galerkin approximations for a class of elliptic sPDEs, Found. Comput. Math. 10 (6) (2010) 615–646.
[16] Patrick R. Conrad, Youssef M. Marzouk, Adaptive Smolyak pseudospectral approximations, SIAM J. Sci. Comput. 35 (6) (2013).
[17] P.G. Constantine, M.S. Eldred, E.T. Phipps, Sparse pseudospectral approximation method, Comput. Methods Appl. Mech. Eng. 229–232 (2012) 1–12.
[18] D.L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (4) (April 2006) 1289–1306.
[19] D.L. Donoho, M. Elad, V.N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, IEEE Trans. Inf. Theory 52 (1) (Jan. 2006) 6–18.
[20] A. Doostan, H. Owhadi, A non-adapted sparse approximation of PDEs with stochastic inputs, J. Comput. Phys. 230 (8) (2011) 3015–3034.
[21] Marco F. Duarte, Michael B. Wakin, Richard G. Baraniuk, Fast reconstruction of piecewise smooth signals from random projections, in: Online Proceedings of the Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS), Rennes, France, 2005.
[22] J. Foo, G.E. Karniadakis, Multi-element probabilistic collocation method in high dimensions, J. Comput. Phys. 229 (5) (2010) 1536–1557.
[23] B. Ganapathysubramanian, N. Zabaras, Sparse grid collocation schemes for stochastic natural convection problems, J. Comput. Phys. 225 (1) (2007) 652–685.
[24] A. Genz, A package for testing multiple integration subroutines, in: P. Keas, G. Fairweather (Eds.), Numerical Integration, D. Riedel, 1987, pp. 337–340.
[25] R.G. Ghanem, P.D. Spanos, Stochastic Finite Elements: A Spectral Approach, Springer-Verlag New York, Inc., New York, NY, USA, 1991.
[26] T. Hastie, R. Tibshirani, J.H. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction: With 200 Full-Color Illustrations, Springer-Verlag, New York, 2001.
[27] J.D. Jakeman, S.G. Roberts, Local and dimension adaptive stochastic collocation for uncertainty quantification, in: Jochen Garcke, Michael Griebel (Eds.), Sparse Grids and Applications, in: Lecture Notes in Computational Science and Engineering, vol. 88, Springer, Berlin, Heidelberg, 2013, pp. 181–203.

[28] Ron Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Proceedings of the 14th International Joint Conference on Artificial Intelligence – vol. 2, IJCAI'95, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1995, pp. 1137–1143.

[29] Stefan Kunis, Holger Rauhut, Random sampling of sparse trigonometric polynomials, II. Orthogonal matching pursuit versus basis pursuit, Found. Comput. Math. 8 (6) (2008) 737–763.

[30] C. La, M.N. Do, Tree-based orthogonal matching pursuit algorithm for signal reconstruction, in: 2006 IEEE International Conference on Image Processing, 2006, pp. 1277–1280.

[31] X. Ma, N. Zabaras, An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations, J. Comput. Phys. 228 (2009) 3084–3113.

[32] L. Mathelin, K.A. Gallivan, A compressed sensing approach for partial differential equations with random input data, Commun. Comput. Phys. 12 (2012) 919–954.

[33] L. Mathelin, M. Hussaini, T. Zang, Stochastic approaches to uncertainty quantification in CFD simulations, Numer. Algorithms 38 (1–3) (Mar. 2005) 209–236.

[34] Deanna Needell, Joel A. Tropp, Cosamp: iterative signal recovery from incomplete and inaccurate samples, Commun. ACM 53 (12) (December 2010) 93–100.

[35] F. Nobile, R. Tempone, C.G. Webster, An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data, SIAM J. Numer. Anal. 46 (5) (2008) 2411–2442.

[36] Ji Peng, Jerrad Hampton, Alireza Doostan, A weighted-minimization approach for sparse polynomial chaos expansions, J. Comput. Phys. 267 (0) (2014) 92–111.

[37] Carl Edward Rasmussen, Christopher K.I. Williams, Gaussian Processes for Machine Learning, Adaptive Computation and Machine Learning, The MIT Press, 2005.

[38] Holger Rauhut, Rachel Ward, Sparse Legendre expansions via ℓ_1-minimization, J. Approx. Theory 164 (5) (2012) 517–533.

[39] Khachik Sargsyan, Cosmin Safta, Habib N. Najm, Bert J. Debusschere, Daniel Ricciuto, Peter Thornton, Dimensionality reduction for complex models via Bayesian Compressive Sensing, Int. J. Uncertain. Quantificat. 4 (1) (2014) 63–93.

[40] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, Reliab. Eng. Syst. Saf. 93 (7) (Jul. 2008) 964–979.

[41] M. Tatang, W. Pan, R. Prinn, G. McRae, An efficient method for parametric uncertainty analysis of numerical geophysical model, J. Geophys. Res. 102 (D18) (1997) 21925–21932.

[42] R. Tibshirani, Regression shrinkage and selection via the Lasso, J. R. Stat. Soc., Ser. B, Methodol. 58 (1) (1996) 267–288.

[43] E. van den Berg, M.P. Friedlander, Probing the Pareto frontier for basis pursuit solutions, SIAM J. Sci. Comput. 31 (2) (2008) 890–912.

[44] R. Ward, Compressed sensing with cross validation, IEEE Trans. Inf. Theory 55 (12) (Dec. 2009) 5773–5782.

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

[46] D. Xiu, G.E. Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, SIAM J. Sci. Comput. 24 (2) (2002) 619–644.

[47] Z. Xu, T. Zhou, On sparse interpolation and the design of deterministic interpolation points, SIAM J. Sci. Comput. (2015), http://dx.doi.org/10.1137/13094596X, in press.

[48] L. Yan, L. Guo, D. Xiu, Stochastic collocation algorithms using ℓ_1-minimization, Int. J. Uncertain. Quantificat. 2 (3) (2012) 279–293.

[49] X. Yang, G.E. Karniadakis, Reweighted minimization method for stochastic elliptic differential equations, J. Comput. Phys. 248 (2013) 87–108.