An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis

Stefano Marelli*, Bruno Sudret

Chair of Risk, Safety and Uncertainty Quantification, Institute of Structural Engineering, ETH Zürich, Stefano-Franscini-Platz 5, CH-8093 Zürich, Switzerland

ABSTRACT

Polynomial chaos expansions (PCE) have seen widespread use in the context of uncertainty quantification. However, their application to structural reliability problems has been hindered by the limited performance of PCE in the tails of the model response and due to the lack of local metamodel error estimates. We propose a new method to provide local metamodel error estimates based on bootstrap resampling and sparse PCE. An initial experimental design is iteratively updated based on the current estimation of the limit-state surface in an active learning algorithm. The greedy algorithm uses the bootstrap-based local error estimates for the polynomial chaos predictor to identify the best candidate set of points to enrich the experimental design. We demonstrate the effectiveness of this approach on a well-known analytical benchmark representing a series system, on a truss structure and on a complex realistic frame structure problem.

1. Introduction

Structural reliability analysis aims at computing the probability of failure of a system with respect to some performance criterion in the presence of uncertainty in its structural and operating parameters. Such uncertainty can be modelled by a random vector $X \in \mathbb{R}^M$ with prescribed joint probability density function f_X. The limit-state function g is defined over the support of X such that $\{x : g(x) \leqslant 0\}$ defines the failure domain, while defines the safe domain. The *limit state surface* implicitly defined by $g(x) = 0$ lies at the boundary between the two domains. The probability of failure of such a system can be defined as [1,2]:

$$P_F = \int_{\{x:g(x)\leqslant 0\}} f_X(x)dx. \tag{1}$$

A straightforward approach to compute the integral in Eq. (1) is to use of Monte Carlo Simulation (MCS). However, it is often the case that standard MCS approaches cannot be used in the presence of complex and computationally expensive engineering models, because of the large number of samples they require to estimate small probabilities (typically in the order of $\sim 10^{k+2}$ for $P_F \approx 10^{-k}$) with acceptable accuracy. Well-known methods based on local approximation of the limit-state function close to the failure domain (such as FORM [3] and SORM [4]) can be more efficient, yet they are usually based on linearisation and tend to fail in real-case scenarios with highly non-linear structural models.

In contrast, methods based on surrogate modelling have gradually gained momentum in the last few years. Due to the nature of the problem of estimating low probabilities, most recent methods combine active-learning-based greedy algorithms with Gaussian process surrogate models (Kriging). Among the first works to propose this approach, the earliest applications in this context were the efficient global reliability analysis method (EGRA) by Bichon et al. [5,6], and the active-learning reliability (AK-MCS) method based on Kriging by Echard et al. [7]. More recently, Kriging has been employed to devise quasi-optimal importance densities in [8,9]. Amongst other variations, polynomial-chaos-based Kriging has also been used as an alternative metamodelling technique [10] to overcome some of the limitations of pure Kriging-based methods. Additional works on the topic of Kriging and structural reliability can be found, including extensions of the original AK-MCS algorithm to more advanced sampling techniques [11,12], system reliability [13] and for the exploration of multiple-failure regions [14].

Polynomial chaos expansions (PCE) [15] are a well-established tool in the context of uncertainty quantification, with applications in uncertainty propagation [16], sensitivity analysis [17] and, to a lesser degree, structural reliability [18]. While often considered as an efficient surrogate modelling technique due to their global convergence behaviour, PCEs have been employed only seldom in reliability analysis (see, *e.g.* [19]) due to their lack of accuracy in the tails of the model response distribution, which are essential in this field.

In addition, most active-learning approaches with surrogates require some form of local error estimate to adaptively enrich a small set

---

of model evaluations close to the limit state surface. Kriging-based methods can rely on the Kriging variance for this task, but PCEs do not provide a natural equivalent.

In this paper, we leverage on the properties of regression-based sparse-PCE [20] to derive a local error estimator based on *bootstrap resampling*. We then use this estimator to construct an active-learning strategy that adaptively approximates the limit-state function with PCE by minimizing a *misclassification probability*-based learning function at every iteration. The method is then showcased on a standard benchmark functions representing a series system and on a complex realistic frame structure problem.

2. Methodology

2.1. Polynomial chaos expansions

Consider a finite variance model $Y = \mathcal{M}(\boldsymbol{X})$ representing the response of some quantity of interest (QoI) Y to the random input parameters $\boldsymbol{X} \in \mathbb{R}^M$, modelled by a joint probability distribution function (PDF) $f_{\boldsymbol{X}}$. Also consider the functional inner product defined by:

$$\langle g, h \rangle \equiv \int_{\boldsymbol{x} \in \Omega_{\boldsymbol{X}}} g(\boldsymbol{x})h(\boldsymbol{x})f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x} = \mathbb{E}\left[g(\boldsymbol{X})h(\boldsymbol{X}) \right] \tag{2}$$

where $\Omega_{\boldsymbol{X}}$ represents the input domain. Under the assumption of independence of the input variables, that is $f_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{M} f_{X_i}(x_i)$, one can represent $\mathcal{M}(\boldsymbol{X})$ as the following *generalised polynomial chaos expansion* (see, *e.g.* [15,16]):

$$Y = \mathcal{M}(\boldsymbol{X}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} y_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{X}), \tag{3}$$

where the $y_{\boldsymbol{\alpha}}$ are real coefficients and $\boldsymbol{\alpha}$ is a multi-index that identifies the degree of the multivariate polynomial $\Psi_{\boldsymbol{\alpha}}$ in each of the input variables X_i:

$$\Psi_{\boldsymbol{\alpha}} = \prod_{i=1}^{M} \phi_{\alpha_i}^{(i)}(X_i). \tag{4}$$

Here $\phi_{\alpha_i}^{(i)}$ is a polynomial of degree α_i that belongs to the family of orthogonal polynomials w.r.t. the marginal PDF f_{X_i}. For more details on the construction of such polynomials for both standard and arbitrary distributions, the reader is referred to [16].

In the presence of a complex dependence structure between the input variables, it is always possible to construct isoprobabilistic transforms (*e.g.* Rosenblatt or Nataf transforms, see *e.g.* [21]) to decorrelate the input variables prior to the expansion, even in the case of complex dependence modelled by vine copulas [22]. For the sake of notational simplicity and without loss of generality, we will hereafter assume independent input variables.

In practical applications, the series expansion in Eq. (3) is traditionally truncated based on the maximal degree p of the expansion, thus yielding a set of basis elements identified by the multi-indices $\boldsymbol{\alpha} \in \mathcal{A}: \sum_{i=1}^{M} \alpha_i \leqslant p$, with $\text{card}(\mathcal{A}) \equiv P = \binom{M + p}{p}$, or using more advanced truncation schemes that favour sparsity, *e.g.* hyperbolic truncation [23]. The corresponding expansion coefficients $y_{\boldsymbol{\alpha}}$ can then be calculated efficiently via least-square analysis based on an existing sample of the input random vector $\mathcal{X} = \{\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(N)}\}$, known as the *experimental design* (ED), and the corresponding model responses $\mathcal{Y} = \{y^{(1)}, \cdots, y^{(N)}\}$ as follows:

$$y_{\boldsymbol{\alpha}} = \text{argmin} \frac{1}{N} \sum_{i=1}^{N} \left[y^{(i)} - \sum_{\boldsymbol{\alpha} \in \mathcal{A}} y_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}^{(i)}) \right]^2. \tag{5}$$

When the number of unknown coefficients P is high (*e.g.* for high-dimensional inputs or high-degree expansions), regression strategies that favour sparsity are needed to avoid over-fitting in the presence of a limited-size experimental design and to make the analysis at all feasible

with a reasonable sample size N. Amongst them, *least angle regression* (LARS, [24]), based on a regularized version of Eq. (5), has proven to be very effective in tackling realistic engineering problems even in relatively high dimensions (*i.e.* $M \sim 100$). In this paper, we adopt the full degree-adaptive, sparse PCE based on hybrid-LARS introduced in [20], as implemented in the UQLAB Matlab software [25,26].

2.2. Bootstrap-based local error estimation in PCE

2.2.1. Bootstrap in least-square regression

Adopting a least-square regression strategy to calculate the coefficients in Eq. (5) allows one to use the *bootstrap* resampling method [27] to obtain information on the variability in the estimated coefficients caused by the finite size of the experimental design. Suppose that a set of estimators θ is a function of a finite-size sample $\mathcal{X} = \{\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(N)}\}$ drawn from the random vector $\boldsymbol{X}$. Then the *bootstrap* method consists in drawing B new sample sets $\{\mathcal{X}^{(1)}, \cdots, \mathcal{X}^{(B)}\}$ from the original $\mathcal{X}$ by *resampling with substitution*. This is achieved by randomly assembling B−times N realizations $\boldsymbol{x}^{(i)} \in \mathcal{X}$, possibly including repeatedly the same realization multiple times within each sample. The set of estimated quantities can then be re-calculated from each of the B samples, thus yielding a set of estimators $\Theta = \{\theta^{(1)}, \cdots, \theta^{(B)}\}$. This set of estimators can then be used to directly assess the variability of θ due to the finite size of the experimental design $\mathcal{X}$, at no additional costs, *e.g.* by calculating statistics, or directly using each realization separately. An application of the bootstrap method combined with PCE to provide confidence bounds in the estimated P_F in structural reliability applications can be found in *e.g.* [19,28].

2.2.2. Bootstrap-PCE

We propose to use the bootstrap technique to provide local error estimates to the PCE predictions. The rationale is the following: the PCE coefficients $\boldsymbol{y}_{\boldsymbol{\alpha}}$ in Eq. (5) are estimated from the experimental design $\mathcal{X}$, therefore they can be resampled through bootstrap. This can be achieved by first generating a set of bootstrap-resampled experimental designs $\{\mathcal{X}^{(b)}, \mathcal{Y}^{(b)}, b = 1, \cdots, B\}$. For each of the generated designs, one can calculate a corresponding set of coefficients $\boldsymbol{y}_{\boldsymbol{\alpha}}^{(b)}$, effectively resulting in a set of B different PCEs. Correspondingly, the response of each PCE can be evaluated at a point $\boldsymbol{x}$ as follows:

$$Y_{PC}^{(b)}(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} y_{\boldsymbol{\alpha}}^{(b)} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}), \tag{6}$$

thus yielding a full response sample at each point $\{Y_{PC}^{(b)}(\boldsymbol{x}), b = 1, \cdots, B\}$. Therefore, empirical quantiles can be employed to provide local error bounds on the PCE prediction at each point, as well as to any derived quantity (*e.g.* P_F or sensitivity indices, see *e.g.* [28,29]).

This bootstrap-resampling strategy in Eq. (6) yields in fact a family of B surrogate models that can be interpreted as *trajectories*. Fig. 1 showcases how such trajectories can be directly employed to assess confidence bounds on point-wise predictions on a simple 1D test function given by:

$$f(x) = x\sin(x), \qquad x \in [0, 2\pi], \tag{7}$$

where the single random variable is assumed to be uniformly distributed within the bounds $X \sim \mathcal{U}(0, 2\pi)$, and where $B = 100$ bootstrap samples have been used.

This process of bootstrap-based trajectory resampling to provide better estimates of point-wise confidence bounds has been recently explored in the Gaussian process modelling literature, see *e.g.*, [30,31].

We refer to this approach as to *bootstrap-PCE*, or bPCE in short.

2.2.3. Fast bPCE

Because the training of a PCE model with sparse least-square analysis may be time consuming, especially in high dimension and/or when an already large experimental design is available (*i.e.* $N \sim 10^3$), and because in this particular application we do not need very accurate