Journal of Computational Physics 230 (2011) 2345–2367

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

Adaptive sparse polynomial chaos expansion based on *least angle regression*

Géraud Blatman [a,b,*], Bruno Sudret [a,c]

[a] Clermont Université, IFMA, EA 3867, Laboratoire de Mécanique et Ingénieries, BP 10448, F-63000 Clermont-Ferrand, France
[b] EDF R&D, Département Matériaux et Mécanique des Composants, Site des Renardières, 77250 Moret-sur-Loing cedex, France
[c] Phimeca Engineering, Centre d'Affaires du Zénith, 34 rue de Sarliève, F-63800 Cournon d'Auvergne, France

ARTICLE INFO

Article history:
Received 15 December 2009
Received in revised form 7 December 2010
Accepted 14 December 2010
Available online 28 December 2010

Keywords:
Least angle regression
Adaptive stochastic finite elements
Sequential experimental design
Sparse polynomial chaos expansion
Sensitivity analysis
Response surfaces

ABSTRACT

Polynomial chaos (PC) expansions are used in stochastic finite element analysis to represent the random model response by a set of coefficients in a suitable (so-called *polynomial chaos*) basis. The number of terms to be computed grows dramatically with the size of the input random vector, which makes the computational cost of classical solution schemes (may it be intrusive (i.e. of Galerkin type) or non intrusive) unaffordable when the deterministic finite element model is expensive to evaluate.

To address such problems, the paper describes a non intrusive method that builds a *sparse* PC expansion. First, an original strategy for truncating the PC expansions, based on *hyperbolic index sets*, is proposed. Then an adaptive algorithm based on *least angle regression* (LAR) is devised for automatically detecting the significant coefficients of the PC expansion. Beside the sparsity of the basis, the experimental design used at each step of the algorithm is systematically complemented in order to avoid the overfitting phenomenon. The accuracy of the PC metamodel is checked using an estimate inspired by statistical learning theory, namely the *corrected leave-one-out error*. As a consequence, a rather small number of PC terms are eventually retained (*sparse* representation), which may be obtained at a reduced computational cost compared to the classical "full" PC approximation. The convergence of the algorithm is shown on an analytical function. Then the method is illustrated on three stochastic finite element problems. The first model features 10 input random variables, whereas the two others involve an input random field, which is discretized into 38 and 30 – 500 random variables, respectively.

1. Introduction

Mathematical models are widely used in many science disciplines, such as physics, biology and meteorology. They aim at better understanding and explaining real-world phenomena. These models are also extensively employed in an industrial context in order to analyze the behaviour of structures and products. This allows the engineers to design systems with ever increasing performance and reliability at an optimal cost.

In structural mechanics, mathematical models may range from simple analytical formulæ (e.g. simple applications in beam theory) to sets of partial differential equations (e.g. a general problem in continuum mechanics). Characterizing the behaviour of the system (e.g. identifying the stress or displacement fields of an elastic structure) may be not an easy task since a closed-form solution is generally not available. Then numerical solving schemes have to be employed, such as the finite difference or the finite element method.

* Corresponding author at: Clermont Université, IFMA, EA 3867, Laboratoire de Mécanique et Ingénieries, BP 10448, F-63000 Clermont-Ferrand, France.
E-mail address: geraud.blatman@edf.fr (G. Blatman).

0021-9991/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2010.12.021

From this viewpoint, the recent considerable improvements in *computer simulation* have allowed the analyst to handle models of ever increasing complexity. Therefore taking into account quite realistic constitutive laws as well as particularly fine finite element meshes become affordable. Nonetheless, in spite of this increase of the accuracy of the models, computer simulation never predicts exactly the behaviour of a real-world complex system. By and large such a discrepancy may be explained by an uncertainty in the input parameters.

Input parameter uncertainty can be taken into account in simulation by setting the problem in a probabilistic framework. This methodology relies upon the combination of a numerical model of the physical system under consideration and a probabilistic description of the input parameters. In this respect, the latter are represented by random variables or random fields. Therefore the response is also a random variable or vector resulting from the propagation of the random input through the model.

This random output may be represented explicitly in a suitable space spanned by the *polynomial chaos* (PC) basis [1,2] which is a series of multivariate polynomials that are orthogonal with respect to the joint distribution of the input variables. In this setting, characterizing the response PDF is equivalent to evaluating the PC coefficients, i.e. the coordinates of the random response in this basis. A considerable work has been accomplished in this direction for models governed by stochastic elliptic differential equations. In particular, the so-called *spectral stochastic finite elements method* (SSFEM) was pioneered in [1] for mechanical problems involving random fields. In this approach, the classical finite element discretization is combined with the Karhunen–Loève decomposition of the input random fields. Then the coefficients of the response PC expansion are obtained using a Galerkin scheme that leads to a system of *coupled* deterministic equations, hence the term *intrusive* for this approach [3]. Moreover, the method has been extended to models featuring particular types of non linearity, see e.g. [4–6].

As an alternative, *non intrusive* computational schemes emerged recently in stochastic finite element analysis. These methods allow the analyst to compute the stochastic model response by means of a set of calls to the existing deterministic model, i.e. without modifying the underlying computer code. Two approaches based on a PC representation of the response are usually distinguished:

- The *projection approach*: each PC coefficient is recast as a multidimensional integral [7,8] which can be computed either by simulation or quadrature.
- The *regression approach* [9–12], also known as *point collocation* [13]: the PC coefficients are estimated by minimizing the mean square error of the response approximation.

Note that a non intrusive scheme based on Lagrange interpolation of the model response, namely the *stochastic collocation* method, was also proposed in the past few years as an efficient alternative to SSFEM [14,15].

However, the required number of model evaluations (i.e. the computational cost) increases with the size of the truncated PC expansion which itself dramatically increases with the number of input variables, whatever the applied computational method, may it be intrusive or non intrusive. Several attempts to downsize the stochastic problems under consideration may be found in the literature. On the one hand, the *dimension-adaptive tensor product quadrature* technique outlined in [16] was applied to the stochastic collocation scheme by Ganapathysubramanian and Zabaras [17]. The method is intended to reducing the number of calls to the model when evaluating multivariate integrals by quadrature. On the other hand, various authors detailed generalized decomposition schemes of the random response of standard elliptic models. The so-called *reduced stochastic basis* method [18,19] has been developed, which consists in seeking the model response onto an *a priori* reduced stochastic basis, namely a basis of a low-dimensional Krylov subspace. More recently, the so-called *generalized spectral decomposition* (GSD) method has been investigated in [6]. It is aimed at building *iteratively* (and not *ab initio*) a reduced basis. The strategy has been recently extended to non linear problems [20].

In contrast to most of the literature, this work is focused on a methodology that can be applied to the uncertainty and reliability analyses of industrial systems, whatever the type of the governing equations. A non intrusive strategy is adopted so that a broad class of industrial problems (either linear or nonlinear) can be tackled. It is therefore crucial to minimize the number of model evaluations which may be time-consuming. In this purpose, a stepwise regression technique has been recently proposed in [21,22] in order to build up a *sparse* PC expansion, i.e. in which only a small number of significant basis functions are retained in the response PC approximation. The method provided a significant computational gain compared to usual full PC expansions for uncertainty and sensitivity analysis.

The present paper is aimed at improving the stepwise regression procedure in terms of convergence with respect to the number of model evaluations. To this end, a new strategy for truncating the PC expansions is proposed. In addition, an efficient recent method for variable selection in statistics, namely *least angle regression* (LAR) [23], is exploited for building up a sparse PC approximation at a reduced computational cost.

The remainder of this paper is organized as follows. In the next section one presents a general framework of the chaos representation of the random response. A new truncation scheme of the PC expansion is proposed in Section 3. Then an iterative LAR-based procedure is designed in Section 4 for building up iteratively a sparse PC approximation of the model response. A robust and conservative error estimate is developed in this purpose. The computational gain provided by the adaptive sparse PC expansions is illustrated in Section 6 by numerical studies, i.e. the analysis of an analytical function and three stochastic finite element models. The first model features 10 input random variables, whereas the two others involve an input random field, which is discretized into 38 and 100 random variables, respectively.

2. Polynomial chaos approximation

2.1. Polynomial chaos representation

2.1.1. Probabilistic modelling

Let us consider a physical model represented by a deterministic mapping $\boldsymbol{y} = \mathcal{M}(\boldsymbol{x})$. Here $\boldsymbol{x} = \{x_1, \ldots, x_M\}^{\mathsf{T}} \in \mathbb{R}^M$, $M \geqslant 1$ is the vector of the input variables (e.g. geometry, material properties, loading). Vector $\boldsymbol{y} = \{y_1, \ldots, y_Q\}^{\mathcal{T}} \in \mathbb{R}^Q$, $Q \geqslant 1$ is the set of quantities of interest provided by the model (e.g. nodal displacements, components of the strain or stress field, post-processed quantities such as energy, average values, etc.), referred to as the *model response* in the sequel.

As the input vector $\boldsymbol{x}$ is assumed to be affected by uncertainty, it is represented by a *random* vector $\boldsymbol{X}$ with prescribed probability density function (PDF) $f_{\boldsymbol{X}}(\boldsymbol{x})$. As a consequence the model response is also a random variable denoted by $Y = \mathcal{M}(\boldsymbol{X})$. This leads us to work in the probability space $(\mathbb{R}^M, \mathcal{B}_M, \mathbb{P}_X)$, where $\mathcal{B}_M$ is the Borel σ-algebra of $\mathbb{R}^M$ and $\mathbb{P}_{\boldsymbol{X}}$ is the probability measure of $\boldsymbol{X}$, i.e. $\mathbb{P}_{\boldsymbol{X}}(d\boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}$.

For the sake of simplicity, a *scalar* response Y is considered from now on, i.e. $Q = 1$. Note that in case of a vector-valued model response $\boldsymbol{Y}$, the following derivations hold componentwise. On the other hand, Y is assumed to have a finite variance, that is $\mathbb{V}[Y] < \infty$. In other words, Y is supposed to belong to the space $L^2 \equiv L^2(\mathbb{R}^M, \mathcal{B}_M, \mathbb{P}_X)$ of $\mathbb{P}_X$-square integrable functionals of $\boldsymbol{X}$. Let us define the following inner product:

$$\langle g(\boldsymbol{X}), h(\boldsymbol{X}) \rangle \equiv \mathbb{E}[g(\boldsymbol{X})\, h(\boldsymbol{X})], \tag{1}$$

which induces the norm:

$$\|g(\boldsymbol{X})\| \equiv \sqrt{\langle g(\boldsymbol{X}), g(\boldsymbol{X}) \rangle} \equiv \sqrt{\mathbb{E}[g(\boldsymbol{X})]}. \tag{2}$$

It is well-known that L^2 endowed with inner product $\langle \cdot, \cdot \rangle$ is a Hilbert space.

2.1.2. Independent random variables

The input random variables in $\boldsymbol{X}$ are supposed to be *independent*. Then it is shown that Y may be expanded onto an orthogonal polynomial basis as follows [2]:

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

where the series converges in the sense of the L^2-norm. The $a_{\boldsymbol{\alpha}}$'s are unknown deterministic coefficients, and the $\psi_{\boldsymbol{\alpha}}$'s are multivariate polynomials. The series in Eq. (3) is usually referred to as *polynomial chaos* (PC) *expansion*. The principles of the building of the PC basis are now recalled.

Let us first notice that due to the independence of the input random variables, the input joint PDF may be cast as:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{M} f_{X_i}(x_i), \tag{4}$$

where $f_{X_i}(x_i)$ is the marginal PDF of X_i. Let us consider a family $\{\pi_j^{(i)}, j \in \mathbb{N}\}$ of orthonormal polynomials with respect to f_{X_i}, i.e.

$$\left\langle \pi_j^{(i)}(X_i), \pi_k^{(i)}(X_i) \right\rangle \equiv \mathbb{E}\left[\pi_j^{(i)}(X_i)\pi_k^{(i)}(X_i)\right] = \delta_{jk}, \tag{5}$$

where δ_{jk} equals 1 if $j = k$ and 0 otherwise. The polynomials are selected and enumerated in such a way that the degree of $\pi_j^{(i)}$ is j for $j > 0$ and $\pi_0^{(i)} \equiv 1$, $1 \leqslant i \leqslant M$. Upon tensorizing the M resulting families of univariate polynomials, one gets a basis of multivariate polynomials $\{\psi_{\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \mathbb{N}^M\}$ defined by

$$\psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \equiv \pi_{\alpha_1}^{(1)}(x_1) \times \cdots \times \pi_{\alpha_M}^{(M)}(x_M), \tag{6}$$

where the multidimensional index notation $\boldsymbol{\alpha} \equiv \{\alpha_1, \ldots, \alpha_M\}$ has been adopted. By construction, the family of multivariate polynomials in the input vector is an orthonormal basis of L^2:

$$\left\langle \psi_{\boldsymbol{\alpha}}(\boldsymbol{X}), \psi_{\boldsymbol{\beta}}(\boldsymbol{X}) \right\rangle \equiv \mathbb{E}\left[\psi_{\boldsymbol{\alpha}}(\boldsymbol{X})\, \psi_{\boldsymbol{\beta}}(\boldsymbol{X})\right] = \delta_{\boldsymbol{\alpha}\boldsymbol{\beta}}, \tag{7}$$

where $\delta_{\boldsymbol{\alpha}\boldsymbol{\beta}}$ equals 1 if $\boldsymbol{\alpha} = \boldsymbol{\beta}$ and 0 otherwise.

The PC expansion was originally formulated with standard Gaussian random variables and Hermite polynomials as the *finite-dimensional Wiener polynomial chaos* [24,1]. It was later extended to other types of random variables together with basis functions from the Askey family of hypergeometric polynomials [25–27]. The decomposition is then referred to as *generalized* PC expansion. In this setup, most common continuous distributions can be associated to a specific family of polynomials [28,25]. If other distribution types appear, then it is possible to employ a nonlinear mapping (namely an *isoprobabilistic transform*) such that the generalized PC expansion can be applied to the new variable. For instance, a lognormal variable will be recast as a function of a standard normal variable, which will be used in conjunction with Hermite polynomials. As an alternative, *ad hoc* orthogonal polynomial may be generated numerically for random variables with arbitrary distributions [29,30].

2.2. Case of an input random field

Many stochastic finite elements studies involve an input random field, e.g. spatially variable material properties in mechanics [1]. Let us denote by $H(\boldsymbol{z},\omega)$ such a random field, where $\boldsymbol{z}$ is a spatial variable in a bounded domain $\mathcal{D} \subset \mathbb{R}^d$ ($d \in \{1,2,3\}$) and ω is the elementary event of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ is a σ-algebra of Ω and $\mathbb{P}$ is a probability measure. The random field $H(\boldsymbol{z},\omega)$ is assumed to be square-integrable, with mean $\mu(\boldsymbol{z})$ and autocorrelation function $C_H(\boldsymbol{x},\boldsymbol{x}')$. Then it may be discretized using the *Karhunen–Loève expansion* [31]:

$$H(\boldsymbol{z},\omega) = \mu(\boldsymbol{z}) + \sum_{i=1}^{+\infty} \sqrt{\lambda_i}\xi_i(\omega)\varphi_i(\boldsymbol{x}). \tag{8}$$

The above series converges in the $\mathcal{L}^2$-norm. The $(\xi_i(\omega))_{i\in\mathbb{N}^*}$'s form a sequence of uncorrelated, zero-mean and unit-variance random variables, and the $(\lambda_i)_{i\in\mathbb{N}^*}$'s and the $(\varphi_i(\boldsymbol{x}))_{i\in\mathbb{N}^*}$'s are the solutions of the generalized eigenvalue problem:

$$\int_{\mathcal{D}} C_H(\boldsymbol{x},\boldsymbol{x}')\varphi_i(\boldsymbol{x}')d\boldsymbol{x}' = \lambda_i\varphi_i(\boldsymbol{x}), \quad \forall i \in \mathbb{N}^*. \tag{9}$$

The eigenvalues are indexed in decreasing order (i.e. $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_M \geqslant \cdots$). The Karhunen–Loève expansion is optimal in the sense of the mean-square norm.

For computational purpose the series in Eq. (8) is truncated after M terms, the value of which being selected *a priori* with respect to a target accuracy of discretization [32]. Although the problem in Eq. (9) admits a closed-form solution for particular choices of $C_H(\boldsymbol{x},\boldsymbol{x}')$, it generally requires the implementation of a numerical solving scheme. In this purpose, a Galerkin scheme may be used together with the approximation of the autocorrelation function C_H onto a suitable basis, e.g. a finite element-like basis [1] or spectral bases such as orthogonal polynomials [33] and wavelets [34].

If the random field $H(\boldsymbol{z},\omega)$ is Gaussian, the ξ_i's form a set of *independent* standard Gaussian random variables. Then the model response may be expanded onto a basis made of normalized Hermite polynomials as shown in Section 2.1.2. A particular class of non Gaussian random fields $H(\boldsymbol{z},\omega)$ may be cast as a non-linear transformation of a Gaussian random field. Such random fields are known as *translation fields* [35]. For instance, input parameters such as material properties are often modelled by *lognormal* random fields:

$$H(\boldsymbol{z},\omega) = \mathrm{e}^{N(\boldsymbol{z},\omega)}, \tag{10}$$

where $N(\boldsymbol{z},\omega)$ is a Gaussian random field. $H(\boldsymbol{z},\omega)$ may be recast in terms of independent Gaussian random variables by substituting $N(\boldsymbol{z},\omega)$ for its Karhunen–Loève expansion in Eq. (10). The reader is referred to [36] for a comprehensive overview of the methods for simulating non-Gaussian random fields.

2.3. Estimation of the coefficients by regression

For computational purpose the PC expansion in Eq. (3) has to be truncated after P terms. One generally retains those polynomials $\psi_{\boldsymbol{\alpha}}$ with total degree up to p:

$$\mathcal{M}_p(\boldsymbol{X}) = \sum_{0\leqslant|\boldsymbol{\alpha}|\leqslant p} a_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(\boldsymbol{X}) \equiv \boldsymbol{a}^{\mathcal{T}}\boldsymbol{\psi}(\boldsymbol{X}), \quad |\boldsymbol{\alpha}| \equiv \sum_{i=1}^{M}\alpha_i, \tag{11}$$

where $\boldsymbol{a}$ (resp. $\boldsymbol{\psi}$) gathers the coefficients $\{a_{\boldsymbol{\alpha}}, 0 \leqslant |\boldsymbol{\alpha}| \leqslant p\}$ (resp. the basis polynomials $\{\psi_{\boldsymbol{\alpha}}, 0 \leqslant |\boldsymbol{\alpha}| \leqslant p\}$).

The coefficients may be estimated by determining the L^2-projection of the response $\mathcal{M}(\boldsymbol{X})$ onto the space spanned by the polynomials $\{\psi_{\boldsymbol{\alpha}}(\boldsymbol{X}), |\boldsymbol{\alpha}| \leqslant p\}$ as follows [11,37]:

$$\widehat{\boldsymbol{a}} = \arg\min_{\boldsymbol{a}\in\mathbb{R}^p} \mathbb{E}\left[\left(\mathcal{M}(\boldsymbol{X}) - \boldsymbol{a}^{\mathcal{T}}\boldsymbol{\psi}(\boldsymbol{X})\right)^2\right]. \tag{12}$$

The solution reads:

$$\widehat{\boldsymbol{a}} = \left(\mathbb{E}\left[\boldsymbol{\psi}(\boldsymbol{X})\boldsymbol{\psi}^{\mathcal{T}}(\boldsymbol{X})\right]\right)^{-1}\mathbb{E}[\boldsymbol{\psi}(\boldsymbol{X})\mathcal{M}(\boldsymbol{X})]. \tag{13}$$

By orthonormality of the PC basis, this expression reduces to:

$$\widehat{\boldsymbol{a}} = \mathbb{E}[\boldsymbol{\psi}(\boldsymbol{X})\mathcal{M}(\boldsymbol{X})]. \tag{14}$$

Two kinds of coefficient estimates may be computed, namely:

- The *quasi-regression* (or *projection*) estimates, which rely upon a discretization of Eq. (14).
- The *regression* estimates, which are based on the discretization of the whole expression in Eq. (13).

From now on, only the regression method is considered since it converges faster in terms of number of model evaluations, as shown empirically in [38] and theoretically in [39] (see Appendix A).

Let $\mathcal{X} = \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\}$ be a set of N realizations of the input random vector, and $\mathcal{Y} = \{y^{(1)}, \ldots, y^{(N)}\}^{\mathcal{T}}$ be the corresponding model evaluations $\{y^{(i)} = \mathcal{M}(\boldsymbol{x}^{(i)}), i = 1, \ldots, N\}$. The set $\mathcal{X}$ is called the *experimental design* (ED). The estimates of the PC coefficients are obtained by computing the empirical analogue of Eq. (13), namely:

$$\widehat{\boldsymbol{a}} = \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\psi}(\boldsymbol{x}^{(i)}) \boldsymbol{\psi}^{\mathsf{T}}(\boldsymbol{x}^{(i)}) \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\psi}(\boldsymbol{x}^{(i)}) \mathcal{M}(\boldsymbol{x}^{(i)}) \right), \tag{15}$$

which may be recast as:

$$\widehat{\boldsymbol{a}} = \left(\boldsymbol{\Psi}^{\mathsf{T}} \boldsymbol{\Psi} \right)^{-1} \boldsymbol{\Psi}^{\mathsf{T}} \mathcal{Y}, \tag{16}$$

where the *data matrix* $\boldsymbol{\Psi}$ is defined by

$$\boldsymbol{\Psi}_{ij} \equiv \psi_{\boldsymbol{\alpha}_j}(\boldsymbol{x}^{(i)}), \quad i = 1, \ldots, N, \;\; j = 0, \ldots, P-1. \tag{17}$$

In order to ensure the numerical stability of the regression problem in Eq. (16), the size N of the ED must be selected in such a way that the information matrix is well-conditioned. In practice, the design size is usually set equal to $N = kP$ with $k \in [2,3]$.

2.4. Post-processing of the PC expansion

Once the coefficients have been estimated, one gets the following *PC approximation*:

$$\widehat{\mathcal{M}}_p(\boldsymbol{X}) \equiv \sum_{|\boldsymbol{\alpha}| \leqslant p} \widehat{a}_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{X}). \tag{18}$$

The latter may used as a metamodel (i.e. an analytical surrogate of the model response $\mathcal{M}(\boldsymbol{X})$) in order to compute quantities of interest in uncertainty and sensitivity analysis at a negligible computational cost. In particular, the mean and the variance of the response may be respectively approximated by

$$\hat{\mu}_{Y,p} \equiv \hat{a}_{\boldsymbol{0}}, \tag{19}$$

$$\hat{\sigma}^2_{Y,p} \equiv \sum_{0 < |\boldsymbol{\alpha}| \leqslant p} \hat{a}^2_{\boldsymbol{\alpha}}. \tag{20}$$

It is also possible to derive inexpensively the *global sensitivity indices* of the response to the input variables [12]. In particular, the *total sensitivity index* to the variable X_i, $i = 1, \ldots, M$, which quantify the part of the total response variance σ^2_Y that is explained by X_i, is estimated by

$$S_i^{T,p} = \frac{1}{\widehat{\sigma}^2_Y} \sum_{\boldsymbol{\alpha} \in \mathcal{I}_i^+} a^2_{\boldsymbol{\alpha}}, \tag{21}$$

where $\mathcal{I}_i^+$ denotes the set of all indices with a non zero i-th component, that is

$$\mathcal{I}_i^+ \equiv \left\{ \boldsymbol{\alpha} \in \mathbb{N}^M 0 \leqslant |\boldsymbol{\alpha}| \leqslant p, \, \alpha_i \neq 0 \right\}. \tag{22}$$

2.5. Limitation of the regression approach in high dimension

As shown in Eq. (11), the PC expansion of the model response is usually truncated in such a way that only those basis polynomials with total degree not greater than p are retained. The size of this basis is $P = \binom{M+p}{p}$. Considering a number $N = kP$ ($k \in [2,3]$) of model evaluations, the computational effort blows up for large values of M or p, say $M \geqslant 10$ or $p \geqslant 10$. As the present work is aimed at minimizing the computational cost N, a reduction of the dimensionality P of the truncated PC basis is necessary. A first remedee consists in defining an alternative truncation scheme, as shown in the sequel.

3. Hyperbolic polynomial chaos expansion

3.1. Hyperbolic scheme for truncating the polynomial chaos expansion

Let us first respectively define the *degree* and the *rank* of an index $\boldsymbol{\alpha} \in \mathbb{N}^M$ by

$$|\boldsymbol{\alpha}| \equiv \|\boldsymbol{\alpha}\|_1 \equiv \alpha_1 + \cdots + \alpha_M, \tag{23}$$

and

$$\|\boldsymbol{\alpha}\|_0 \equiv \sum_{i=1}^{M} \boldsymbol{1}_{\{\alpha_i > 0\}}. \tag{24}$$

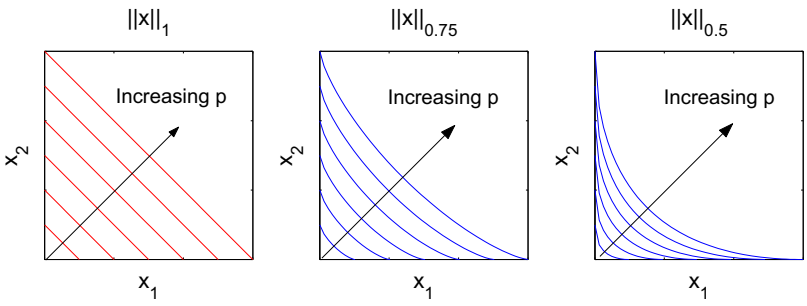

Fig. 1. Principle of the truncation strategy based on q-norms ($0 < q \leqslant 1$).

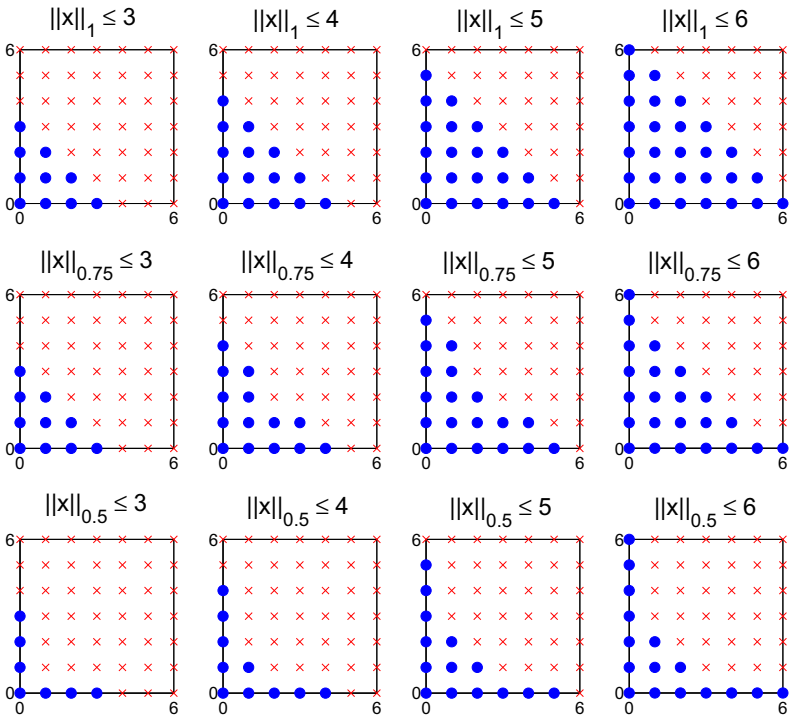

Fig. 2. Retained basis terms in the polynomial chaos expansion when varying the parameter q of the index sets $\mathcal{A}_q^{M,p}$ and the total degree $p = 3, 4, 5, 6$, for $M = 2$. The x-axes (resp. y-axes) correspond to the partial degree of the polynomials in X_1 (resp. X_2).

Any truncation strategy of the PC representation corresponds to a specific choice of a non empty finite set $\mathcal{A}$ of indices $\boldsymbol{\alpha}$. For instance the common scheme is equivalent to the choice:

$$\mathcal{A} = \mathcal{A}^{M,p} \equiv \left\{\boldsymbol{\alpha} \in \mathbb{N}^M : \|\boldsymbol{\alpha}\|_1 \leqslant p\right\}. \tag{25}$$

As an alternative, the use of the following index sets based on q-quasi-norms,[1] $0 < q < 1$ is proposed:

$$\mathcal{A}_q^{M,p} \equiv \left\{\boldsymbol{\alpha} \in \mathbb{N}^M : \|\boldsymbol{\alpha}\|_q \leqslant p\right\}, \tag{26}$$

where

$$\|\boldsymbol{\alpha}\|_q \equiv \left(\sum_{i=1}^{M} \alpha_i^q\right)^{1/q}. \tag{27}$$

Such norms penalize the high-rank indices all the more since q is low, as shown in Fig. 1 for a two-dimensional case. This leads to favor the main effects and low-order interactions, which are more likely to be significant than the high-order interactions in the governing equations of the model according to the *sparsity-of-effects principle* [40]. Note that setting q equal to 1 corresponds to the usual truncation scheme, i.e. $\mathcal{A}_1^{M,p} = \mathcal{A}^{M,p}$. When using $q < 1$, the retained basis polynomials

[1] The quantity $\|\cdot\|_q$ defined in Eq. (26) is not a norm since it does not satisfy the triangular inequality.

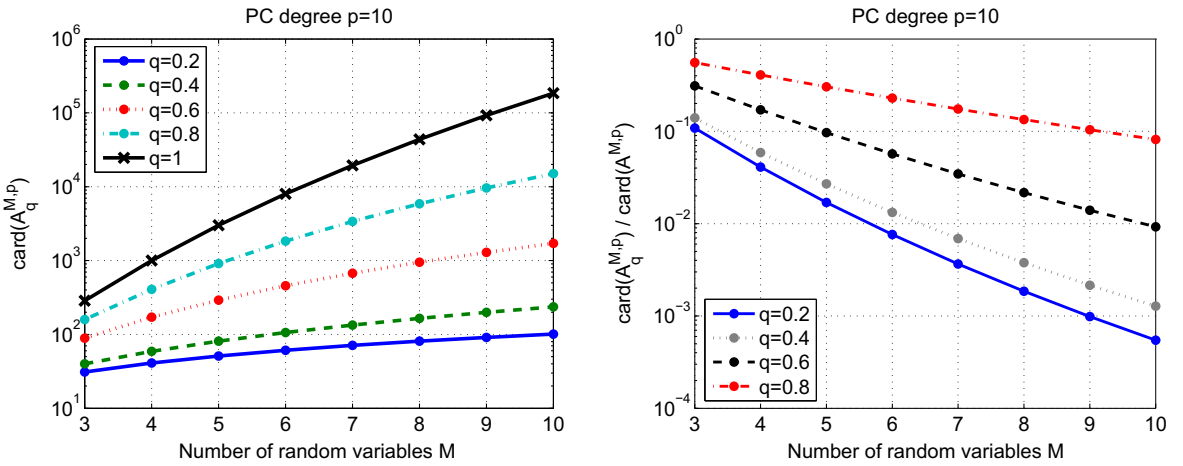

Fig. 3. Cardinality of the index sets $\mathcal{A}_q^{M,p}$ with respect to the number of input random variables M. (Left) Cardinality of $\mathcal{A}_q^{M,p}$ compared to the cardinality of $\mathcal{A}^{M,p}$. (Right) Relative cardinality of $\mathcal{A}_q^{M,p}$.

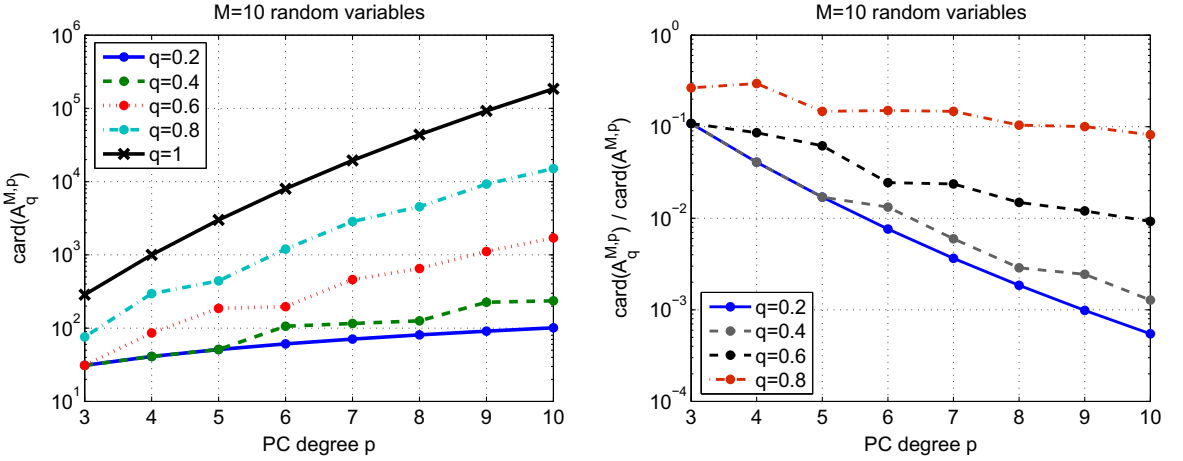

Fig. 4. Cardinality of the index sets $\mathcal{A}_q^{M,p}$ with respect to the PC degree p. (Left) Cardinality of $\mathcal{A}_q^{M,p}$ compared to the cardinality of $\mathcal{A}^{M,p}$. (Right) Relative cardinality of $\mathcal{A}_q^{M,p}$.

are located under an hyperbola-like surface, hence the name *hyperbolic index sets*. Thus any PC representation associated with such a truncation set is referred to as *hyperbolic PC expansion*.[2] For the sake of illustration, the retained basis terms in PC expansions of varying degree p are plotted in Fig. 2 for $q = 1$, $q = 0.75$ and $q = 0.5$.

The evolution of the cardinality of $\mathcal{A}_q^{M,p}$ with respect to the number of input random variables M for various values of q and $p = 10$ is depicted in Fig. 3. It is shown that decreasing q results in a significant reduction of the number of terms compared to the usual truncated PC expansion (i.e. $q = 1$). For instance if $M = p = 10$, and say $q = 0.4$, the cardinality of $\mathcal{A}_q^{M,p}$ is 3 orders of magnitude smaller than that of $\mathcal{A}^{M,p}$. On the other hand, the evolution of $\mathrm{card}(\mathcal{A}_q^{M,p})$ with respect to the PC degree p for various values of q and $M = 10$ is depicted in Fig. 4.

In order to further reduce the number of terms in the PC expansion, one develops a procedure for only selecting a small number of non zero coefficients, leading to a *sparse* representation. This is the scope of the next section.

4. Error estimates for sparse polynomial chaos representations

4.1. Sparse polynomial chaos expansion

Let $\mathcal{A}$ be any finite subset of $\mathbb{N}^M$, and let us consider the associated truncated PC expansion:

$$\mathcal{M}_{\mathcal{A}}(\boldsymbol{X}) \equiv \sum_{\boldsymbol{\alpha}\in\mathcal{A}} a_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(\boldsymbol{X}), \tag{28}$$

[2] Note that this concept of hyperbolic PC expansions is not related to the so-called *hyperbolic cross space* method which is devoted to the approximation of periodic functions [41,42].

The maximum length of the indices in $\mathcal{A}$ is denoted by p as shown below:

$$p \equiv \max_{\boldsymbol{\alpha}\in\mathcal{A}} \|\boldsymbol{\alpha}\|_1, \tag{29}$$

and is referred to as the *degree* of the truncated PC representation. This allows one to consider the corresponding index set $\mathcal{A}_q^{M,p}$ and then to define the *index of sparsity* of $\mathcal{A}$ by

$$IS(\mathcal{A}) \equiv \frac{\mathrm{card}(\mathcal{A})}{\mathrm{card}(\mathcal{A}^{M,p})}. \tag{30}$$

The index set $\mathcal{A}$ and the PC expansion $\mathcal{M}_{\mathcal{A}}(\boldsymbol{X})$ are said to be *sparse* if the index of sparsity is small compared to 1.

Of course it is not possible to determine *a priori* the non zero coefficients. An iterative procedure based on stepwise regression has been proposed in [43,21] to circumvent this difficulty. The algorithm relies upon a rigourous estimation of the approximation error. A robust and conservative estimate is proposed in the sequel.

4.2. Estimation of the approximation error

4.2.1. Generalization error vs. empirical error

Let us consider an experimental design $\mathcal{X} = \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\}^T$. Let $\mathcal{Y} = \{y^{(i)} \equiv \mathcal{M}(\boldsymbol{x}^{(i)}), i = 1, \ldots, N\}^\top$ be the vector of the corresponding model evaluations. As shown in Section 2.3, one may use this data in order to compute the following PC approximation by regression:

$$\widehat{\mathcal{M}}_{\mathcal{A}}(\boldsymbol{X}) \equiv \sum_{\boldsymbol{\alpha}\in\mathcal{A}} \hat{a}_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{X}), \tag{31}$$

where $\mathcal{A}$ is a finite non empty subset of $\mathbb{N}^M$. In this work, we focus on the estimation of the approximation error in the $\mathcal{L}^2$-norm:

$$Err \equiv \mathbb{E}\left[\left(\mathcal{M}(\boldsymbol{X}) - \widehat{\mathcal{M}}_{\mathcal{A}}(\boldsymbol{X})\right)^2\right]. \tag{32}$$

The quantity *Err* is sometimes referred to as the *generalization error* in statistical learning [44]. In practice, *Err* may be estimated by the *empirical error* (or *training error*) defined by

$$Err_{emp} \equiv \frac{1}{N}\sum_{i=1}^{N}\left(\mathcal{M}(\boldsymbol{x}^{(i)}) - \widehat{\mathcal{M}}_{\mathcal{A}}(\boldsymbol{x}^{(i)})\right)^2, \tag{33}$$

where the $\boldsymbol{x}^{(i)}$'s are the points of the experimental design $\mathcal{X}$. The *relative* training error is defined by

$$\varepsilon_{emp} \equiv \frac{Err_{emp}}{\hat{\mathbb{V}}[\mathcal{Y}]}, \tag{34}$$

where $\hat{\mathbb{V}}[\mathcal{Y}]$ denotes the empirical variance of the response sample $\mathcal{Y}$:

$$\hat{\mathbb{V}}[\mathcal{Y}] \equiv \frac{1}{N-1}\sum_{i=1}^{N}\left(y^{(i)} - \overline{\mathcal{Y}}\right)^2, \quad \overline{\mathcal{Y}} \equiv \frac{1}{N}\sum_{i=1}^{N} y^{(i)}. \tag{35}$$

Of common use is the corresponding *coefficient of determination* R^2 that reads:

$$R^2 \equiv 1 - \varepsilon_{emp}. \tag{36}$$

However it is well-known that Err_{emp} underpredicts the generalization error. The quantity Err_{emp} is even systematically reduced by increasing the complexity of the PC approximation (i.e. the cardinality of $\mathcal{A}$), whereas *Err* may increase. This is the so-called *overfitting* phenomenon. An error estimate which is known to be much less sensitive to overfitting than Err_{emp} is investigated in the next section.

4.2.2. Leave-one-out error

Let us denote by $\widehat{\mathcal{M}}_{\mathcal{A}}^{(-i)}$ the metamodel that has been built from the experimental design $\mathcal{X} \setminus \{\boldsymbol{x}^{(i)}\}$, i.e. when removing the i-th observation. The *predicted residual* is defined as the difference between the model evaluation at $\boldsymbol{x}^{(i)}$ and its prediction based on $\widehat{\mathcal{M}}_{\mathcal{A}}^{(-i)}$:

$$\Delta^{(i)} \equiv \mathcal{M}(\boldsymbol{x}^{(i)}) - \widehat{\mathcal{M}}_{\mathcal{A}}^{(-i)}(\boldsymbol{x}^{(i)}). \tag{37}$$

The expected risk is then estimated by the following *leave-one-out error*:

$$Err_{LOO} \equiv \frac{1}{N}\sum_{i=1}^{N} {\Delta^{(i)}}^2. \tag{38}$$

The quantity Err_{LOO} is sometimes referred to as *predicted residual sum of squares* (PRESS) [45] or *jacknife error* [46]. In our context of linearly parametrized regression, it is possible to calculate analytically each predicted residual as follows [47]:

$$\Delta^{(i)} = \frac{\mathcal{M}(\boldsymbol{x}^{(i)}) - \widehat{\mathcal{M}}_{\mathcal{A}}(\boldsymbol{x}^{(i)})}{1 - h_i}, \tag{39}$$

where h_i is the i-th diagonal term of the matrix $\boldsymbol{\Psi}\,(\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^{\,\mathrm{sf\ T}}$, using the notation:

$$\boldsymbol{\Psi}_{ij} \equiv \left(\psi_{\alpha_j}\left(\boldsymbol{x}^{(i)}\right)\right)_{\substack{i=1,\ldots,N \\ j=0,\ldots,\mathrm{card}(\mathcal{A})-1}}. \tag{40}$$

The leave-one-out error rewrites:

$$Err_{LOO} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathcal{M}(\boldsymbol{x}^{(i)}) - \widehat{\mathcal{M}}_{\mathcal{A}}(\boldsymbol{x}^{(i)})}{1 - h_i}\right)^2. \tag{41}$$

As for the training error one considers the following relative leave-one-out error:

$$\varepsilon_{LOO} \equiv \frac{Err_{LOO}}{\hat{\mathbb{V}}[\mathcal{Y}]}, \tag{42}$$

as well as the following counterpart of R^2 denoted by Q^2:

$$Q^2 \equiv 1 - \varepsilon_{LOO}. \tag{43}$$

A comparative study reported in [48] indicates that the leave-one-out technique generally performs very well in terms of estimation bias and mean-square error.

4.2.3. Corrected error estimates

Various penalty-based methods have been proposed in the literature in order to reduce the sensitivity of error estimates to overfitting. Considering for instance the empirical error Err_{emp}, one gets estimates under the form:

$$Err^{*}_{emp} = Err_{emp}T(P,N), \tag{44}$$

where P denotes the number of terms in the PC approximation and $T(P,N)$ is a correcting factor. In particular the so-called *adjusted empirical error* corresponds to:

$$T(P,N) \equiv \frac{N-1}{N-P-1}. \tag{45}$$

The quantity Err^{*}_{emp} is all the larger since the complexity P increases, i.e. as extra terms are included in the metamodel. In this section we focus on a correcting factor which has been derived in [49] for regression using a small experimental design. The factor is defined by

$$T(P,N) \equiv \frac{N}{N-P}\left(1 + \frac{\mathrm{tr}\left(\boldsymbol{C}^{-1}_{emp}\right)}{N}\right), \tag{46}$$

where

$$\boldsymbol{C}_{emp} \equiv \frac{1}{N}\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi}. \tag{47}$$

In the following one suggests the use of an heuristic *corrected leave-one-out error* Err^{*}_{LOO}. Indeed, it has been shown in [39] that such an error estimate remains robust and conservative. The scaled counterpart of Err^{*}_{LOO} is denoted by ε^{*}_{LOO}.

5. Least angle regression (LAR)

Least angle regression (LAR) [23] is an efficient procedure for variable selection. In other words, it is aimed at selecting those predictors (i.e. the basis polynomials $\psi_{\boldsymbol{\alpha}}$ in our case) that have the greatest impact on the model response $Y \equiv \mathcal{M}(\boldsymbol{X})$, among a possibly large set of candidates. LAR eventually provides a *sparse* PC approximation, i.e. which only contains a small number of terms compared to a classical full representation.

More exactly, LAR provides not only a single PC metamodel but a *collection* of PC representations in such a way that the first metamodel includes a single predictor, the second includes two predictors, and so on (Section 5.1). Then a criterion for selecting the "best" metamodel is given (Section 5.2). Precisely, every metamodel provided by LAR is given an accuracy estimate by a *cross validation* procedure. Eventually the metamodel associated with the greatest estimate is retained. In practice its sparsity level (i.e. the number of active basis functions) is much lower than the cardinality of the whole candidate basis. Lastly, adaptive procedures that rely upon iterations on the LAR procedure are detailed (Section 5.3).

5.1. Description of the LAR algorithm

The LAR algorithm is detailed below:

(i) Initialize the coefficients to $a_{\boldsymbol{\alpha}_0}, \ldots, a_{\boldsymbol{\alpha}_{P-1}} = 0$. Set the initial residual equal to the vector of observations $\mathcal{Y}$.
(ii) Find the vector $\boldsymbol{\psi}_{\boldsymbol{\alpha}_j}$ which is most correlated with the current residual.
(iii) Move $a_{\boldsymbol{\alpha}_j}$ from 0 toward the least-square coefficient of the current residual on $\boldsymbol{\psi}_{\boldsymbol{\alpha}_j}$, until some other predictor $\boldsymbol{\psi}_{\boldsymbol{\alpha}_k}$ has as much correlation with the current residual as does $\boldsymbol{\psi}_{\boldsymbol{\alpha}_j}$.
(iv) Move jointly $\{a_{\boldsymbol{\alpha}_j}, a_{\boldsymbol{\alpha}_k}\}^{\mathrm{T}}$ in the direction defined by their joint least-square coefficient of the current residual on $\{\boldsymbol{\psi}_{\boldsymbol{\alpha}_j}, \boldsymbol{\psi}_{\boldsymbol{\alpha}_k}\}$, until some other predictor $\boldsymbol{\psi}_{\boldsymbol{\alpha}_l}$ has as much correlation with the current residual.
(v) Continue this way until $m \equiv \min(P, N-1)$ predictors have been entered.

Steps (ii) and (iii) mention a "move" of the *active* coefficients toward their least-square value. It corresponds to an updating of the form $\hat{\boldsymbol{a}}^{(k+1)} = \hat{\boldsymbol{a}}^{(k)} + \gamma^{(k)} \tilde{\boldsymbol{w}}^{(k)}$. Vector $\tilde{\boldsymbol{w}}^{(k)}$ and coefficient $\gamma^{(k)}$ are referred to as the LAR *descent direction* and *step*, respectively. Both quantities may be derived algebraically as shown in [23].

Note that if $N \geqslant P$, then the last step of LAR provides the ordinary least-square solution. It is shown in [23] that LAR is noticeably efficient since it only requires $\mathcal{O}(NP^2 + P^3)$ computations (i.e. the computational cost of ordinary least-square regression on P predictors) for producing a set of m metamodels. The optimal number of predictors in the metamodel (i.e. the optimal number of LAR steps) may be determined using a suitable criterion. This point is discussed in the next section.

The so-called *hybrid* LAR procedure is a variant of the original LAR [23]. Let us assume that after k steps the LAR algorithm has included k predictors in a index set $\mathcal{A}^{(k)}$. Instead of the associated LAR-based coefficients $\hat{\boldsymbol{a}}^{(k)}$ one may prefer the *least-square estimates* coefficients (denoted here by $\hat{\boldsymbol{a}}_{LS}^{(k)}$). In this setup, LAR is only used in order to select a set of predictors, but not to estimate the coefficients. It is shown in [23] that hybrid LAR always increase the usual empirical measure of fit R^2 compared to the original LAR.

5.2. A modified cross validation criterion for selecting the optimal LAR metamodel

In the current section, the classical cross-validation scheme for assessing the LAR results is first described. Then a more computationally efficient procedure is proposed.

The genuine cross-validation selection of the optimal LARS solution is as follows [50]:

(i) For $i = 1, \ldots, v$:
  (a) Run the LAR procedure from the reduced data set $\mathcal{Z} \setminus \mathcal{Z}_i$ in order to build up a sparse PC approximation. Denoting the number of iterations by $m \equiv \min N, P$, one obtains a set of solution coefficients $\{\hat{\boldsymbol{a}}_i^{(1)}, \ldots, \hat{\boldsymbol{a}}_i^{(m)}\}$ with increasing $\mathcal{L}^1$-norm.
  (b) Compute the residual sums of squares $\{Err_i^{(1)}, \ldots, Err_i^{(m)}\}$ of the various coefficients estimates on the validation sets $\{\mathcal{Z}_1, \ldots, \mathcal{Z}_v\}$, respectively.
(ii) For $j = 1, \ldots, m$: compute the mean residual sum of squares $\overline{Err}^{(j)} \equiv 1/v \sum_{i=1}^{v} Err_i^{(j)}$.
(iii) Find the optimal LARS step $j^* = \arg\min_j \overline{Err}_j$.
(iv) Run the LARS procedure from the whole data set $\mathcal{Z}$. One gets a set of solution coefficients $\{\hat{\boldsymbol{a}}^{(1)}, \ldots, \hat{\boldsymbol{a}}^{(m)}\}$.
(v) Eventually return the set of coefficients estimates $\hat{\boldsymbol{a}}^{(j^*)}$.

This cross-validation scheme dedicated to the selection of the optimal LAR solution will be denoted by CV from now on. The procedure may be time-consuming though since it requires $(v+1)$ calls to the LAR procedure. This may lead to a significant computational cost when applying an iterative strategy, which is the scope of the current work.

To overcome this difficulty, a new cross-validation scheme is proposed that only requires a single call to the *hybrid* LAR procedure. This method is referred to as *modified cross-validation* (MCV) in the following. It is recalled that hybrid LAR provides a set of metamodels $\{\widehat{\mathcal{M}}_{\mathcal{A}^{(1)}}, \ldots, \widehat{\mathcal{M}}_{\mathcal{A}^{(m)}}\}$ in two steps:

- Perform variable selection using the original LAR procedure.
- Compute the coefficients associated with the retained predictors by ordinary least-square regression.

The MCV strategy is designed as follows:

- Run the LAR procedure once and for all.
- Recompute the coefficients of each produced sparse metamodel by least-square regression.
- Assess each metamodel using a cross-validation procedure.
- Eventually retain the metamodel associated with the lowest error estimate.

It may be noticed that in contrast to the CV approach, MCV provides directly an estimate of the approximation error of the LAR-based PC approximations. As suggested in Section 4.2.3, the corrected *leave-one-out* error estimate is employed for assessing the various sparse metamodels obtained by LAR. The MCV method has been validated in [39].

5.3. Basis-adaptive LAR algorithm

5.3.1. Fixed experimental design

A limitation of LAR lies in the requirement of an *a priori* truncation set $\mathcal{A}$. To circumvent this difficulty, one proposes a procedure for enriching iteratively the index set of the PC approximation, i.e. the set of active basis functions. In this section, a given set of model evaluations (i.e. a fixed experimental design) $\mathcal{X} \equiv \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\}^{\mathrm{T}}$ is assumed. Without loss of generality, one considers PC approximations of the model response based on *hyperbolic index sets* $\mathcal{A}_q^{M,p}$, that is

$$Y \approx \mathcal{M}_{\mathcal{A}_q^{M,p}}(\boldsymbol{X}) \equiv \sum_{\boldsymbol{\alpha} \in \mathcal{A}_q^{M,p}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{X}). \tag{48}$$

The computational flowchart of the proposed basis adaptive LAR procedure is sketched in Fig. 5. LAR is first applied to a first-order PC approximation corresponding to $\mathcal{A}_q^{M,1}$. The selection of the best LAR metamodel is performed using the modified cross-validation (MCV) scheme outlined in Section 5.2. The associated optimal subset of indices (resp. error estimate) is stored in the variable $\mathcal{A}^{(1)}$ (resp. $\varepsilon^{(1)}$). Note that $\varepsilon^{(1)}$ has been already computed in order to select the optimal LAR-based PC approximation ($\varepsilon^{(1)}$ is the *corrected leave-one-out error estimate* of the metamodel $\mathcal{M}_{\mathcal{A}^{(1)}}(\boldsymbol{X})$). One sets $\varepsilon^* \equiv \varepsilon^{(1)}$ and one denotes by $\mathcal{A}^*$ the corresponding set of indices. If ε^* is less than a prescribed target error ε_{tgt}, one stops the algorithm. Otherwise one iterates with the second-order PC approximation related to $\mathcal{A}_q^{M,2}$ and sets $\varepsilon^* \equiv \min(\varepsilon^*, \varepsilon^{(2)})$, and so on. Thus the sparse PC approximation is sought among the best LAR metamodels for each degree $p = 1, \ldots, p_{\max}$.

It is possible that the approximation error $\varepsilon^{(p)}$ increases from a given value of the degree p. This might be due to an *overfitting* situation, in which the number of accepted predictors gets too important with respect to the size N of the experimental design $\mathcal{X}$. In order to avoid this phenomenon, the following heuristic criterion is introduced: if the approximation error $\varepsilon^{(p)}$ increases twice in a row (say $\varepsilon^{(p)} \geqslant \varepsilon^{(p-1)} \geqslant \varepsilon^{(p-2)}$), then the algorithm is stopped, returning a warning message of possible overfitting.

The optimal subset $\mathcal{A}^*$ is eventually retained. The coefficients of the related sparse PC approximation $\mathcal{M}_{\mathcal{A}^*}(\boldsymbol{X})$ are computed by ordinary least-square regression.

5.3.2. Sequential experimental design

The algorithm proposed in the last subsection allows one to detect automatically the significant terms in the PC expansion. However it is based on a given experimental design whose size N is arbitrary. This problem is tackled by proposing a version of the procedure in which the design is automatically enriched, so that the target error may be reached. In this purpose, *sequential experimental designs* may be used, such as Monte Carlo sampling, Nested Latin Hypercube sampling (NLHS) [51] or quasi-Monte Carlo sampling (QMC) [52,53].

Using a sequential sampling scheme, a modification of the basis adaptive LAR procedure outlined in the previous section is devised. As soon as overfitting is detected in the adaptive LAR iterations (i.e. when the error estimate increases twice in a row), the experimental design is enriched. Note that the number of additional points is fixed *a priori*. Then the procedure is restarted from the PC degree p that yielded the best metamodel prior to complementing the design. The computational flowchart of the algorithm is sketched in Fig. 6.

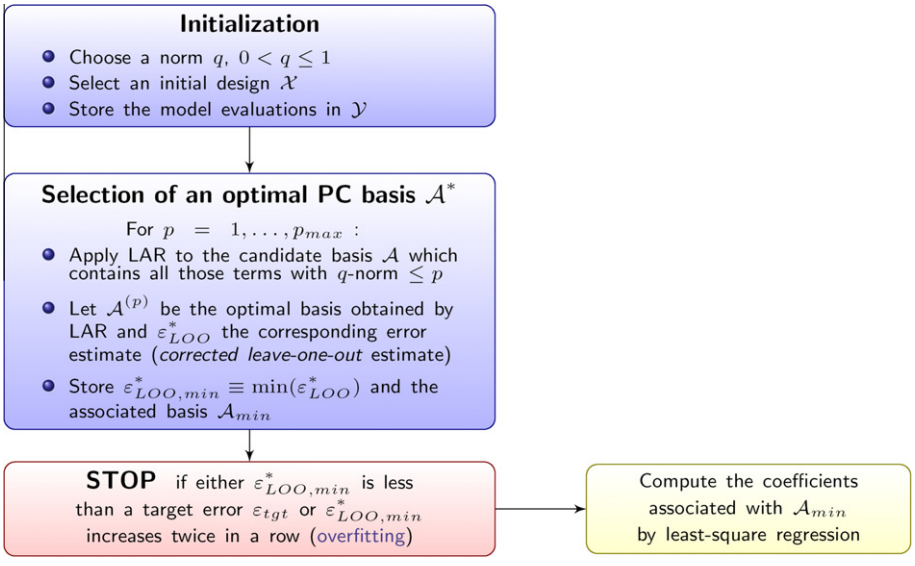

Fig. 5. Computational flowchart of the basis-adaptive LAR procedure for building up an adaptive sparse polynomial chaos expansion.

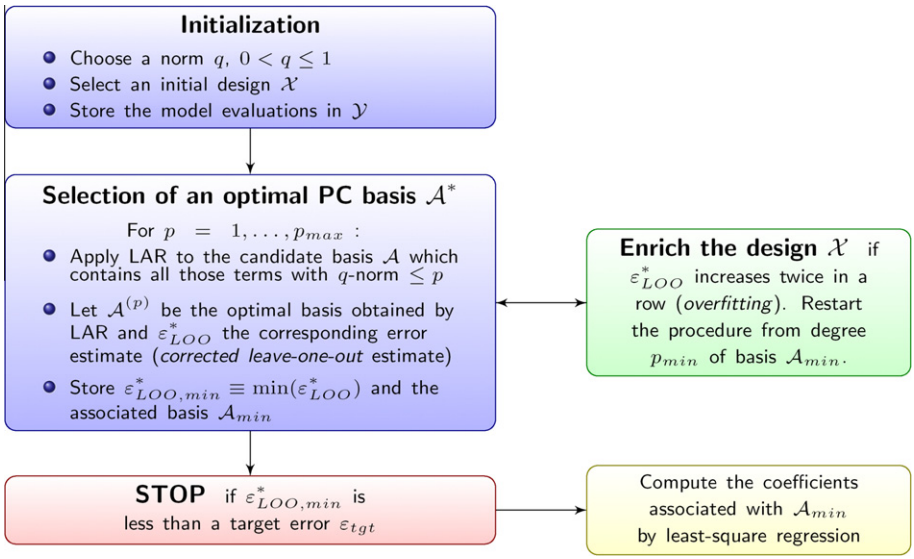

Fig. 6. Computational flowchart of the basis and experimental design adaptive LAR procedure for building up an adaptive sparse polynomial chaos expansion.

6. Application examples

6.1. Methodology

The following sections are dedicated to the uncertainty, sensitivity and reliability analysis of several application examples. Estimates of the related quantities of interest, namely the statistical moments, the Sobol' indices and the probabilities of failure are obtained by post-processing sparse polynomial chaos (PC) expansions of the reponse of the model being studied. The proposed LAR methods is employed in order to build up these sparse metamodels. Several choices have to be made in this purpose.

First, we will only consider *hyperbolic* PC expansions of the form:

$$Y \approx \widehat{\mathcal{M}}(\boldsymbol{X}) \equiv \sum_{\boldsymbol{\alpha} \in \mathcal{A}_q^{M,p}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{X}), \tag{49}$$

where

$$\mathcal{A}_q^{M,p} \equiv \left\{ \boldsymbol{\alpha} \in \mathbb{N}^M : \|\boldsymbol{\alpha}\|_q \equiv \left(\sum_{i=1}^{M} \alpha_i^q \right)^{1/q} \leqslant p \right\}, \tag{50}$$

The parameter q will be set equal to 0.4. Indeed, such a choice has often led to good results as shown in [39] and from the authors' experience. Note that selecting a too small or too high q-value for approximating a given model will only affect the convergence rate (i.e. the trade-off between the accuracy and the number of model evaluations) but will *not* prevent the sparse PC expansion from converging. Indeed, whatever the value of q, all the terms of the (infinite) PC expansion are likely to be added to the approximation basis when increasing the degree p. This is due to the fact that $\lim_{p\to\infty}(\mathcal{A}_q^{M,p})_p = \mathbb{N}^M$.

The sparsity of the obtained sparse PC expansions will be quantified by means of *indices of sparsity* which are defined by

$$IS_1 \equiv \frac{\text{card}(\mathcal{A}_q^{M,p})}{\text{card}(\mathcal{A}^{M,p})}, \tag{51}$$

$$IS_2 \equiv \frac{\text{card}(\mathcal{A})}{\text{card}(\mathcal{A}_q^{M,p})}, \tag{52}$$

where $\mathcal{A}$ is the final index set that has been eventually returned by the algorithms. These quantities respectively correspond to the sparsity due to the choice of the hyperbolic index set and the sparsity due to the adaptive selection of the terms in the PC decomposition.

Second, whatever the approach for building a PC representation (may it be sparse or full), the PC coefficients will be systematically computed using an experimental design made of Sobol' quasi-random numbers. When applying one of the adaptive approaches (i.e. stepwise regression or LAR), one will employ a sequential design strategy. The initial size of the design (resp. the number of additional sample when detecting overfitting) will be set equal by default to $N_{ini} = 100$ (resp. $N_{add} = 100$), unless alternative choices are specified.

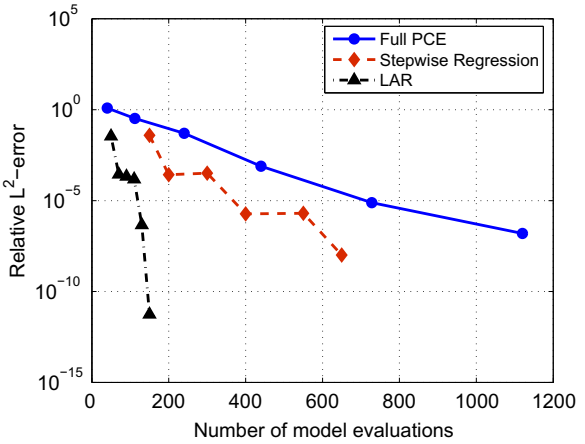

Fig. 7. Ishigami function – convergence rates of full and sparse polynomial chaos expansions (quasi-random experimental designs are used).

6.2. Ishigami function

Let us consider the so-called Ishigami function which is widely used for benchmarking in global sensitivity analysis [54,55]:

$$Y = \sin X_1 + 7 \sin^2 X_2 + 0.1 X_3^4 \sin X_1, \tag{53}$$

where the X_i's ($i = 1,\ldots,3$) are independent random variables that are uniformly distributed over $[-\pi,\pi]$. The convergence of the two following methods is investigated:

- Classical full PC expansions with degree varying from 3 to 13 ($N = 2P$ model evaluations are performed, where P denotes the PC size).
- Sparse PC expansions based on the stepwise regression procedure detailed in [21,22].
- Sparse PC expansions based on the adaptive LAR procedure.

Whatever the computational scheme, the PC coefficients are computed from a quasi-random experimental design. As the final number of model evaluations is expected to be relatively low, the number N_{add} of additional points in the sequential design is set equal to 50 rather than 100. The convergence rate of the metamodels is reported in Fig. 7.

As expected, the full PC expansion has the slowest rate of convergence. In contrast, the LAR-based approach reveals particularly efficient, outperforming both the full PC and the stepwise regression schemes. It yields a relative error less than 10^{-10} using about 150 model evaluations. This shows how the LAR procedure can take advantage of the genuine sparsity of the model response, especially in the case of a smooth function.

6.3. Maximum deflection of a truss structure

6.3.1. Problem statement

Let us consider the truss structure sketched in Fig. 8. The structure comprises 23 members, namely 11 horizontal bars and 12 oblical bars. The upper portion of the truss is subjected to vertical loads. A finite element model made of 23 bar elements

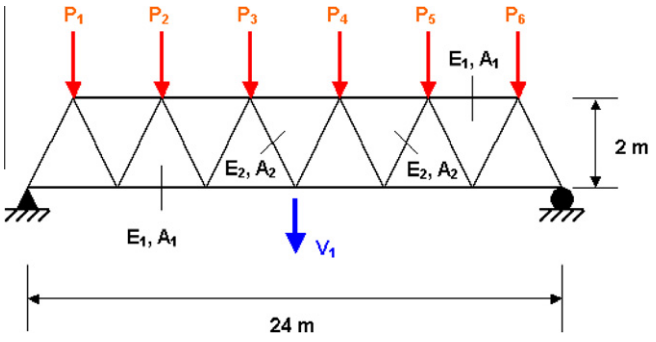

Fig. 8. Example #2: truss structure comprising 23 members.

Table 1
Example #2: truss example – input random variables.

Variable	Distribution	Mean	Standard deviation
E_1, E_2 (Pa)	Lognormal	2.10×10^{11}	2.10×10^{10}
A_1 (m^2)	Lognormal	2.0×10^{-3}	2.0×10^{-4}
A_2 (m^2)	Lognormal	1.0×10^{-3}	1.0×10^{-4}
P_1–P_6 (N)	Gumbel	5.0×10^{4}	7.5×10^{3}

Table 2
Example #2: truss structure – estimates of the total Sobol' indices.

Variables	Total Sobol' indices		
	Reference	Stepwise	LAR
A_1	0.388	0.374	0.374
E_1	0.367	0.378	0.373
P_3	0.075	0.073	0.073
P_4	0.079	0.069	0.079
P_5	0.035	0.036	0.034
P_2	0.031	0.037	0.037
A_2	0.014	0.013	0.013
E_2	0.010	0.014	0.012
P_6	0.005	0.005	0.004
P_1	0.004	0.005	0.005
Number of FE runs 5,500,000		60	70
Relative $\mathcal{L}^2$-error		6×10^{-3}	5×10^{-3}
PC degree		3	3
Index of sparsity 1		$76/286 \approx 27\%$	$76/286 \approx 27\%$
Index of sparsity 2		$21/76 \approx 28\%$	$32/76 \approx 42\%$

is used. Ten parameters are assumed to be random and are modelled by independent input random variables, namely the Young's moduli and the cross-section areas of the horizontal and the oblical bars (respectively denoted by E_1, A_1 and E_2, A_2) and the applied loads (denoted by P_i, $i = 1,\ldots,6$) [56], whose mean and standard deviation are reported in Table 1. Thus the input random vector is defined by

$$\boldsymbol{Z} = \{E_1, E_2, A_1, A_2, P_1, \ldots, P_6\}^{\mathrm{T}}. \tag{54}$$

The model random response Y is the deflection at midspan denoted by V_1. It is approximated by a truncated PC expansion made of normalized Hermite polynomials. In this respect, the random vector $\boldsymbol{Z}$ is recast as a standard Gaussian random vector $\boldsymbol{X}$ by transforming the random variables Z_i as follows:

$$X_i = \Phi^{-1}\left(F_{Z_i}(Z_i)\right) \quad , i = 1, \ldots, 10, \tag{55}$$

where Φ denotes the standard normal cumulative distribution function (CDF) and F_{Z_i} denotes the CDF of Z_i. This leads to the following metamodel:

$$Y \equiv V_1(\boldsymbol{X}) \simeq \sum_{\boldsymbol{\alpha} \in \mathcal{A}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{X}). \tag{56}$$

6.3.2. Sensitivity analysis

Of interest are the total Sobol' indices of the maximum deflection of the truss structure. Estimates are computed by post-processing PC approximations of the model response. In this respect, sparse metamodels are built up using stepwise regression and LAR with a target accuracy $\varepsilon = 0.01$. Reference results are obtained using crude Monte Carlo simulation (5,500,000 finite element runs are performed as a whole). In preliminary calculations, one had set the size of the initial experimental design equal to its default value 100. However both sparse metamodels could converge without adding extra points in the design. In order to check if one could reach the target accuracy with less calls to the finite element model, the initial size N_{ini} of the experimental design (resp. the number N_{add} of added points in case of overfitting) is now set equal to 50 (resp. 10). The results are reported in Table 2.

It can be concluded from the Sobol' indices that the variability of the deflection v is much more sensitive to the variables E_1 and A_1, than E_2 and A_2. This makes sense from a physical point of view since the properties of the horizontal bars are more influential on the displacement at midspan than the oblique ones. It can be also observed that the Sobol' indices associated with E_1 and A_1 (resp. E_2 and A_2) are similar. This is due to the fact that these variables have the same type of PDF and coefficient of variation, and that the displacement v only depends on them through the products E_1A_1 and E_2A_2. Finally, the Sobol' indices reflect the symmetry of the problem, giving similar importances to the loads that are symmetrically applied (e.g. P_3

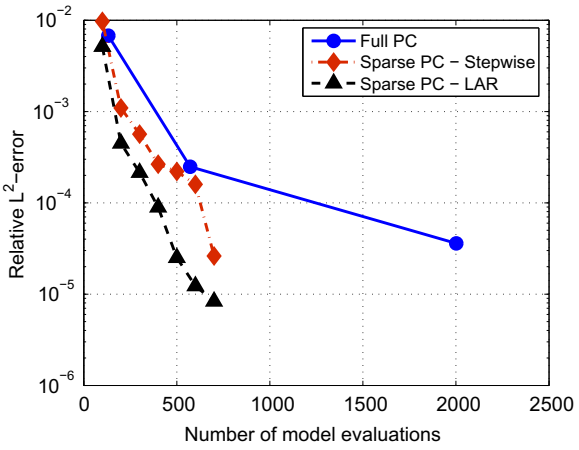

Fig. 9. Truss example – convergence rates of full and sparse polynomial chaos expansions (quasi-random experimental designs are used).

and P_4). Greater sensitivity indices are logically attributed to the forces that are close to the midspan than those located at the ends.

Such physically meaningful indices are obtained from the sparse PC expansions at a very low computational cost, say $60-70$ model evaluations. Hence the sparse PC approaches provide a huge computational gain factor with respect to Monte Carlo. Stepwise regression and LAR have the same efficiency in this example. Note that a full second-order PC approximation, which would be believed to provide accurate estimates of the sensitivity indices, would require about $2P$ model evaluations, where $P \equiv \binom{10+2}{2} = 66$, hence a computational cost twice as large.

6.3.3. Convergence analysis

Adaptive LAR is now compared to the full PC expansions and the stepwise regression method (the degree p of the full representations is successively set equal to 2, 3 and 4). The coefficients are computed by regression using $N = 2P$ points in the experimental design (quasi-random numbers from the Sobol' sequence), where P denotes the PC size. On the other hand, the target accuracy ε_{tgt} of the sparse approximations is set equal to 10^{-5}. The convergence rate of both methods is plotted in Fig. 9.

As noticed in the case of the Ishigami function in Section 6.2, adaptive LAR appears to be the most efficient approach. In particular, it yields a relative error less than 10^{-5} using only $N = 700$ model evaluations, whereas $N = 2{,}000$ simulations are not sufficient for a full fourth-degree PC expansion to attain such an accuracy.

6.4. Settlement of a foundation

6.4.1. Problem statement

Let us study the problem of the settlement of a foundation on an elastic soil layer showing spatial variability in its material properties, already addressed in [32,57]. A structure to be founded on this soil mass is idealized as a uniform pressure P applied over a length $2B$ of the free surface (see Fig. 10). The soil is modelled as an elastic linear isotropic material. A plane strain analysis is carried out.

The finite element model displayed in Fig. 11(a) was chosen. The foundation width is equal to 20 m and the soil mesh width is equal to 120 m. The soil layer thickness is equal to 30 m and its Poisson's ratio to 0.3. The finite element mesh is made of 448 Q4-elements and 495 nodes.

6.4.2. Probabilistic model

The Young's modulus of the soil is considered to vary both in the vertical and the horizontal directions. It is modelled by a two-dimensional homogeneous lognormal random field. Its mean value is set equal to $\mu_E = 50$ GPa and its coefficient of variation is $\delta_E = \sigma_E/\mu_E = 0.3$. The autocorrelation coefficient function of the underlying Gaussian field $N(\boldsymbol{x},\omega)$ is

$$\rho_N(\boldsymbol{x},\boldsymbol{x}') = \exp\left[-\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|^2}{\ell^2}\right], \tag{57}$$

where $\ell = 15$ m. The Gaussian field $N(\boldsymbol{x},\omega)$ is discretized using the Karhunen–Loève (KL) decomposition [31]. The problem is then recast as a function of M independent standard Gaussian random variables $\boldsymbol{\xi} = \{\xi_1,\ldots,\xi_M\}^{\mathrm{T}}$. A relative variance error less than 1% is obtained by selecting $M = 38$ random variables. The discretization error over the structure is illustrated by Fig. 12.

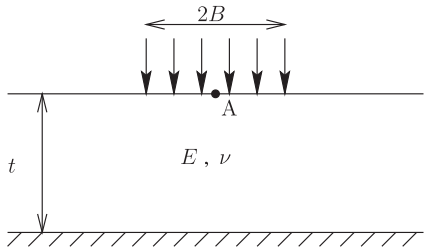

Fig. 10. Example #3: settlement of a foundation – problem definition.

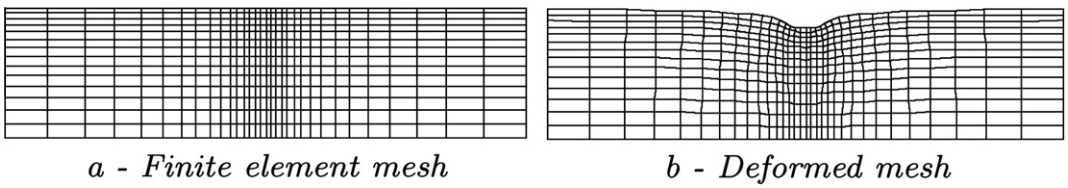

Fig. 11. Example #3: settlement of a foundation – finite element mesh of the soil layer.

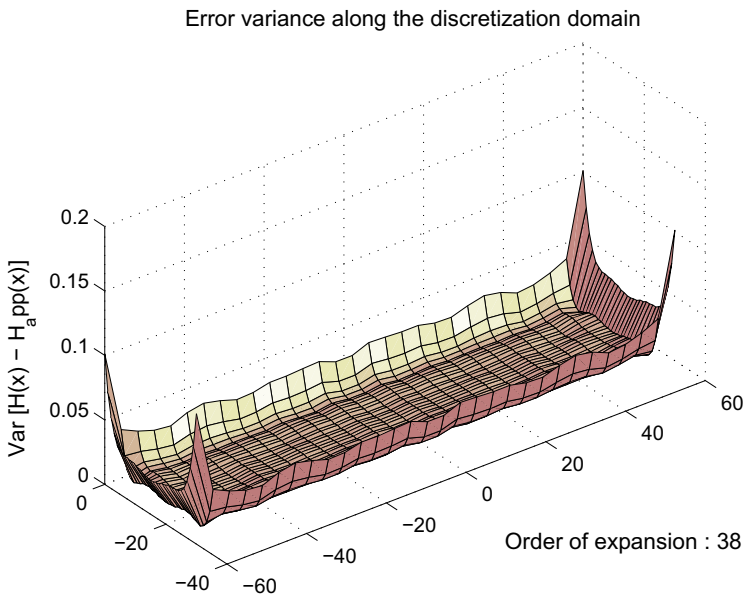

Fig. 12. Example #3: settlement of a foundation – discretization error over the structure of the Young's modulus random field using a *38*-term Karhunen–Loève expansion.

6.4.3. Average vertical settlement under the foundation

The average vertical displacement $\bar{u}$ under the foundation is of interest. It may be regarded as a random variable denoted by $Y = \mathcal{M}(\boldsymbol{\xi})$ (where $\boldsymbol{\xi} \equiv \{\xi_1, \ldots, \xi_M\}$ is a set of independent standard normal variables) due to the probabilistic assumptions presented in the previous section.

The sensitivity of the average vertical displacement to each eigenmode ξ_i in the Karhunen–Loève expansion of the Young's modulus is investigated. In this purpose, the total Sobol' indices of the response are derived by post-processing sparse PC approximations obtained by stepwise regression and LAR.

It appears that both methods provide total Sobol' indices whose sum is approximately equal to 1. This means that the interaction effects between the eigenmodes are negligible and allows one to treat the total indices as the first-order ones, i.e. as ratios between partial variances and the total variance. The sensitivity indices associated with the 15 first eigenmodes are represented in Fig. 13.

LAR appears to be efficient since it makes use of only $N = 200$ runs of the finite element model. In comparison, using a full second-order PC expansion would require performing more than $\binom{M+p}{p} = \binom{38+2}{2} = 780$ model evaluations.

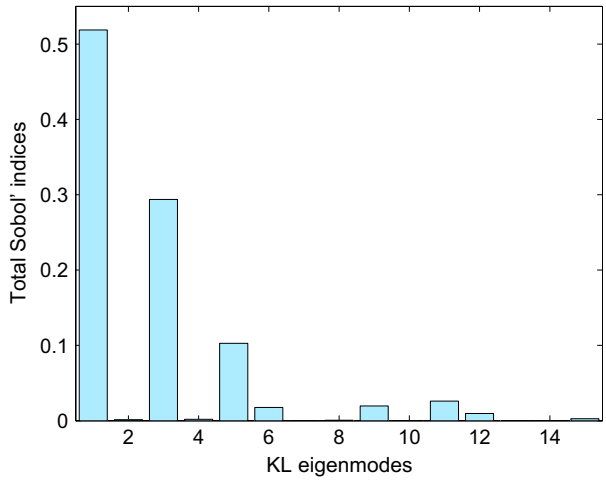

Fig. 13. Example #3: settlement of a foundation – total Sobol' indices of the 15 first eigenmodes of the Karhunen–Loève expansion (which explain more than 99% of the total response variance).

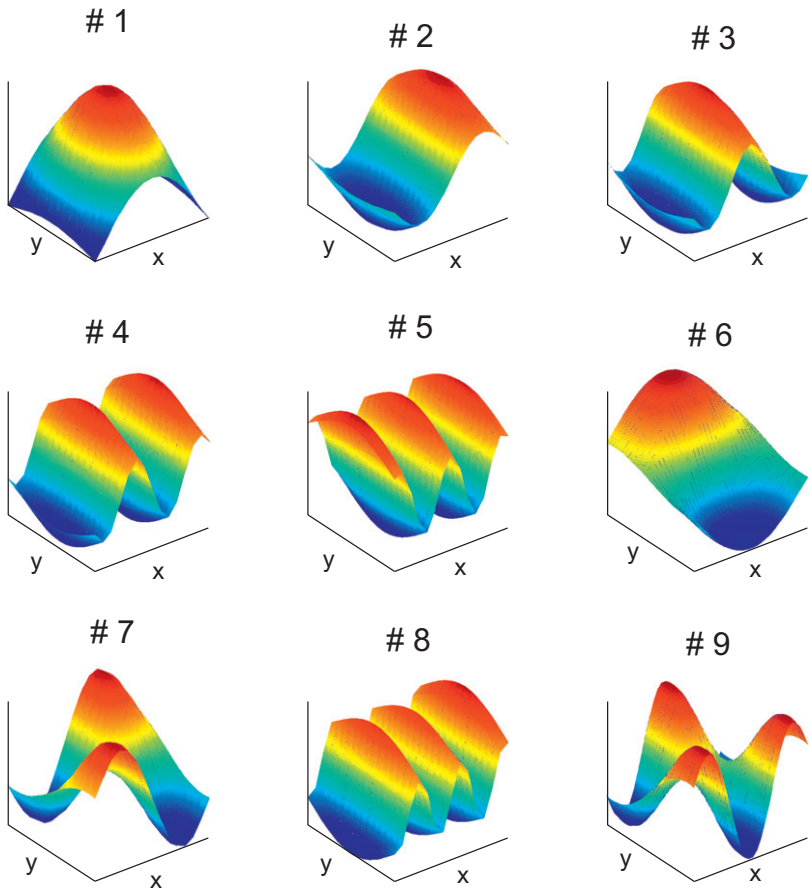

Fig. 14. Example #3: settlement of a foundation – representation of the 9 first eigenmodes of the Karhunen–Loève expansion of the Young's modulus random field.

A very fast decay of the importance of the eigenmodes is noticed, with the 5 first eigenmodes explaining more than 90% of the response total variance. This was expected since the model response of interest is an averaged quantity over the domain of application of the load, which is therefore quite insensitive to small-scale fluctuations of the spatially variable random Young's modulus.

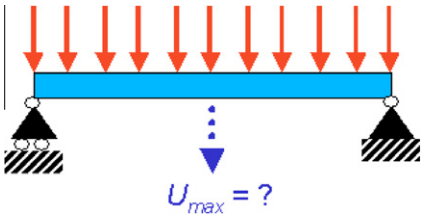

Fig. 15. Example #4 – problem of an elastic beam bending.

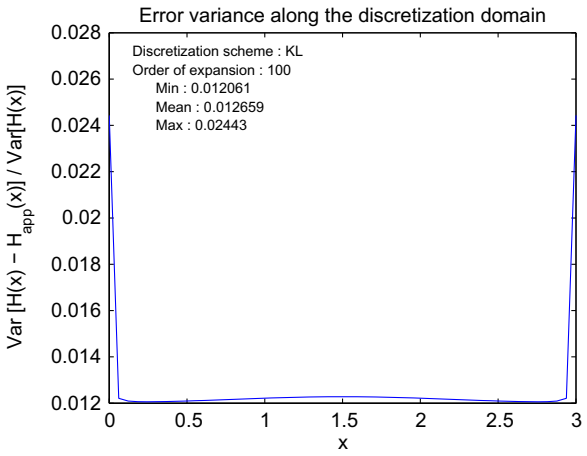

Fig. 16. Example #4: beam bending – discretization error along the beam length of the Young's modulus random field using a 100-term Karhunen–Loève expansion.

In addition, it appears that only the sensitivity indices corresponding to those eigenfunctions that are symmetric with respect to the vertical axis are non zero, reflecting the genuine symmetry of the problem under consideration. This is made clear by associating the results in Fig. 13 with the plot of the 9 first eigenmodes in Fig. 14. Accordingly, many coefficients should vanish in the true PC expansion of the model response, hence a very sparse structure. The antisymmetric modes #2, 4, 7 and 8 with respect to the vertical axis have a zero sensitivity index as expected. This shows the small *effective dimension* of the response in spite of a large nominal dimension $M = 38$.

6.5. Example #4 – bending of a simply supported beam

6.5.1. Problem statement

We now investigate a problem of very low effective dimension in spite of a particularly large number of random input variables. In this purpose, let us consider the elastic beam bending problem depicted in Fig. 15. The beam is simply supported and subjected to an uniformly distributed load. In this study, the beam length is set equal to 3 m and its moment of inertia to 0.8×10^{-5} m^4. The applied pressure is equal to 13 MPa. The finite element mesh is made of 100 beam elements. The maximum deflection $U_{\max}$ of the structure is the output of interest.

The Young's modulus of the beam is modelled by an homogeneous lognormal random field. Its mean value is set equal to $\mu_E = 210$ MPa and its coefficient of variation is $\delta_E = \sigma_E/\mu_E = 20\%$. The autocorrelation coefficient function of the underlying Gaussian field $N(\boldsymbol{x},\omega)$ is

$$\rho_N(\boldsymbol{x},\boldsymbol{x}') = \exp\left[-\frac{|x-x'|}{\ell}\right], \tag{58}$$

where $\ell = 0.5$ m. The Gaussian field $N(\boldsymbol{x},\omega)$ is discretized using the Karhunen–Loève decomposition, which allows one to recast the problem as a function of M independent standard Gaussian random variables $\boldsymbol{\xi} = \{\xi_1,\ldots,\xi_M\}^{\mathrm{T}}$. A relative variance error of about 1% is obtained by selecting $M = 100$ random variables. The discretization error along the beam length is illustrated in Fig. 16. The maximum deflection may thus be regarded as a random variable denoted by $Y = \mathcal{M}(\boldsymbol{\xi})$.

6.5.2. Statistical moments of the maximum deflection

The statistical moments of the response are now considered. Reference results are obtained using crude Monte Carlo simulation of the problem with 10,000 samples. In addition, 95%-confidence intervals of the skewness and kurtosis coefficients

Table 3
Example #4: beam bending – estimation of the four first statistical moments.

	Reference	LAR – $\varepsilon_{tgt} = 0.01$	LAR – $\varepsilon_{tgt} = 0.001$
Mean (mm)	2.83	2.81	2.85
Standard deviation (mm)	0.37	0.36	0.36
Skewness	[0.38; 0.52]†	0.30	0.39
Kurtosis	[3.1; 3.7]†	3.09	3.21
Number of FE runs	10,000	200	700
Error estimate		10^{-3}	10^{-3}
PC degree		3	4
Number of PC terms		10	46
Index of sparsity IS_1		2×10^{-3}	10^{-3}
Index of sparsity IS_2		7%	1%

† 95%-Confidence intervals obtained by bootstrap.

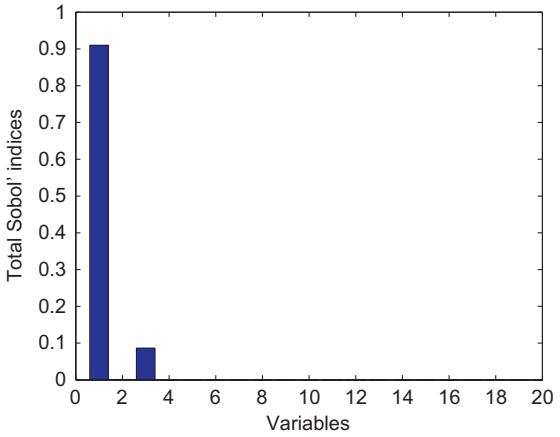

Fig. 17. Example #4: beam bending – estimates of the total Sobol' indices of the 20 first eigenmodes of the Young's modulus Karhunen–Loève decomposition.

have been computed by bootstrap using 1000 replicates. On the other hand, two sparse PC expansions based on LAR are post-processed, namely the one used for sensitivity analysis in the previous section ($\varepsilon_{tgt} = 0.01$) and a more accurate approximation with ε_{tgt} set equal to 0.001. The results are gathered in Table 3.

The sparse PC approximation associated with $\varepsilon_{tgt} = 0.01$ provides accurate estimates of the second moments of the response at a low computational cost. However the estimates of the third and fourth moments are outside the 95%-confidence intervals. In contrast, the metamodel associated with $\varepsilon_{tgt} = 0.001$ yields estimates which lie inside these intervals, at the computational cost of 700 model evaluations. In comparison, a full second-order expansion would contain $P\binom{100+2}{2} = 5151$ terms and would thus require about $2P \approx 10{,}000$ finite element runs in order to get accurate results, hence a computational cost multiplied by 14 compared to the LAR approach. Note that a simple full first-order expansion could not provide satisfactory estimates of the skewness and kurtosis coefficients γ_Y and κ_Y, since it corresponds to a Gaussian approximation which would yield the estimates $\widehat{\gamma}_Y = 0$ and $\widehat{\kappa}_Y = 3$.

6.5.3. Sensitivity analysis of the maximum deflection

The sensitivity of the maximal deflection to each eigenmode ξ_i in the Karhunen–Loève expansion of the Young's modulus is investigated. In this purpose, the total Sobol' indices of the response are derived by post-processing a sparse PC approximation obtained by LAR. Indeed, this method revealed efficient in the previous example which also involved a Karhunen–Loève (KL) expansion of a Young's modulus random field. The target accuracy ε_{tgt} is set equal to 0.01.

The sparse PC approximation could be determined using only 200 finite element runs, which is especially low since a full second-order PC representation contains 5151 terms. The estimates of the total Sobol' indices of the 20 first KL eigenmodes are depicted in Fig. 17.

As expected, a dramatically fast decay of the sensitivity indices is observed. In addition, exactly as for the previous example, only the symmetric modes are non zero. For the sake of clarity, the four first KL eigenmodes are plotted in Fig. 18. Thus it appears that more that 99.7% of the variance of the maximum deflection is explained by the eigenmodes #1 and #3.

Therefore, one considers now a crude KL expansion of the Young's modulus random field that contains only 3 terms. The corresponding average variance error of discretization is quite large, say about 25%. Thus the maximal deflection $U_{\max}$ is

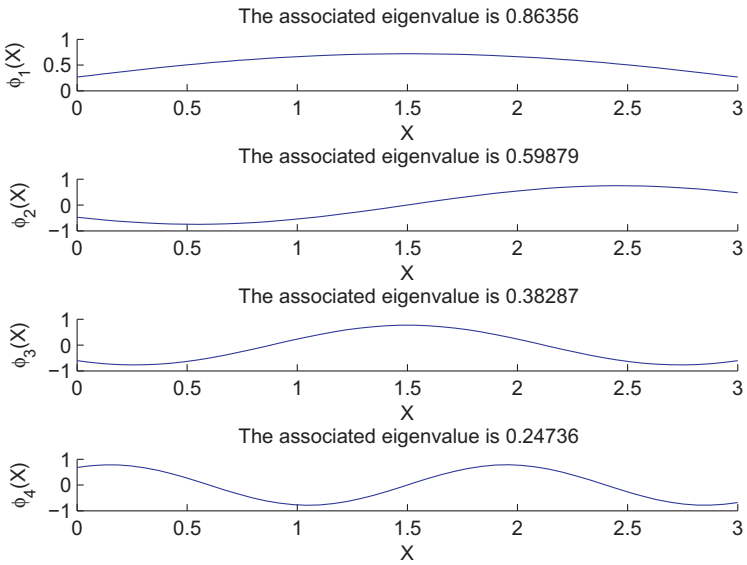

Fig. 18. Example #4: beam bending – four first eigenmodes of the Karhunen–Loève decomposition of the spatially variable Young's modulus.

Table 4
Example #4: beam bending – parametric study varying the number M of input parameters.

	$M = 30$ ($\ell = 2.2$ m)		$M = 50$ ($\ell = 1.3$ m)		$M = 70$ ($\ell = 1.0$ m)		$M = 100$ ($\ell = 0.5$ m)
	Full PC	Sparse PC	Full PC	Sparse PC	Full PC	Sparse PC	Sparse PC
Error estimate	$3 \cdot 10^{-3}$	$9 \cdot 10^{-4}$	$8 \cdot 10^{-3}$	$9 \cdot 10^{-4}$	$3 \cdot 10^{-2}$	$7 \cdot 10^{-4}$	10^{-3}
# FE runs	992	300	2652	450	5112	500	700
# PC terms	496	5	1326	12	2556	20	46
PC degree	2	4	2	4	2	4	4
IS_1	–	$1.2 \cdot 10^{-2}$	–	$4.5 \cdot 10^{-3}$	–	$2.3 \cdot 10^{-3}$	$1.2 \cdot 10^{-3}$
IS_2	–	$9.0 \cdot 10^{-3}$	–	$2.3 \cdot 10^{-2}$	–	$3.6 \cdot 10^{-2}$	$8.6 \cdot 10^{-2}$

approximated by a LAR-based sparse PC expansion featuring 2 standard normal random variables. The target accuracy is $\varepsilon_{tgt} = 0.001$. As the required number of model evaluations is expected to be small, one uses a sequential experimental design of initial size $N_{ini} = 10$, and the number of additional points when detecting overfitting is $N_{add} = 10$.

The adaptive LAR procedure converges at the cost of only $N = 40$ runs of the model. The obtained sparse PC approximation contains $P = 18$ terms, and its degree is $p = 5$. The estimate of the approximation error (*corrected leave-one-out error*) is equal to 2×10^{-4}. The second moments of the maximal deflection are estimated again from this reduced model. The estimates of the mean and the standard deviation are respectively 2.81 and 0.36 mm. These results are as accurate (comparing with the reference values) as those obtained from the "fine" KL discretization ($M = 100$ terms) using $N = 200$ model evaluations, hence a computational gain factor greater than 5.

6.5.4. Parametric study varying the dimensionality of the problem

The influence of the number M of eigenmodes in the Karhunen–Loève decomposition (i.e. the number of input random variables) on the computational costs required by the full and sparse PC approaches is now investigated. In this purpose, the correlation scale ℓin Eq. (58) is successively set equal to 2.2, 1.3 and 1.0 m. In order to obtain a discretization error less than 1%, the associated numbers of retained eigenmodes M are respectively equal to 30, 50 and 70. In each case, both the full and the LAR-based PC methods are employed to build up an approximation of the model response. Precisely, second-order full PC expansions are used whereas the LAR procedure is run with a target error ε_{tgt} equal to 10^{-3}. The results of the parametric study are reported in Table 4 together with those obtained by LAR when $M = 100$ (Section 6.5.3). Note that in the latter case, a second-order PC expansion could not be computed due to too important memory requirements (indeed the number P of PC terms is equal to $\binom{100}{100+2} = 4,598,126$).

The LAR-based approach appears to provide more accurate PC approximations than the "full" approach (the error estimates have orders of magnitude 10^{-4} instead of 10^{-3}) while making use of a much lower number of model evaluations. Actually the computational gain factor noticeably increases with the dimension M, varying from 3.3 to 10.2 when M varies from 30 to 70. Besides, the index of sparsity IS_1 of the sparse PC representations decreases when increasing M due to the use of the

Table 5
Example #4: Beam bending - High-dimensional case ($M = 500$).

	Reference	Sparse PC
Mean (mm)	2.86	2.88
Standard deviation (mm)	0.21	0.22
Skewness	[0.2; 0.3]†	0.2
Kurtosis	[3.0; 3.2]†	3.1
Number of FE runs	10,000	1000

† 95%-Confidence intervals obtained by bootstrap.

hyperbolic truncation scheme. Indeed, as shown in Fig. 3, the latter leads to a number of PC terms which increases slowly with M compared to the classical truncation scheme. In contrast, the index of sparsity IS_2 increases from 0.9% to 8.6%, indicating a decrease of the sparsity of the model response when reducing the correlation scale of the random field.

6.5.5. High dimensional case

The correlation scale $\ell$of the input random field is now set equal to 0.14 m. Accordingly, $M = 500$ eigenmodes are retained in the Karhunen–Loève decomposition. The model response is approximated by a sparse PC representation using the LAR approach. The target error ε is set equal to 10^{-2}.

A PC approximation of degree $p = 2$ is obtained, containing $P = 92$ non zero terms. The indices of sparsity IS_1 and IS_2 are respectively equal to $8 \cdot 10^{-3}$ and $9.2 \cdot 10^{-2}$. The approximation error of the PC metamodel is equal to $9.4 \cdot 10^{-3}$. The PC-based four first statistical moments of the model response are reported in Table 5 together with reference values obtained by direct Monte Carlo simulation (*10,000* samples are used). Accurate estimates of the response mean and standard deviation are provided, with relative errors of 0.6% and 5% with respect to the reference values, respectively. In addition, the estimates of the skewness and kurtosis coefficients belong to the reference 95%-confidence intervals.

7. Conclusion

In this paper, a method is proposed to build *iteratively* a polynomial chaos (PC) expansion of the random response of a model with random input parameters. It is based on an efficient algorithm for variable selection in statistics, namely *least angle regression* (LAR). The procedure provides a set of less and less sparse solutions (i.e. sets of coefficients) to the approximation problem. The experimental design (ED) employed for this purpose is systematically enriched so that the overfitting phenomenon is avoided. Quasi-random designs are employed to this end. The optimal LAR solution is then obtained using a simplified cross validation scheme.

The iterative procedure is first tested on an analytical function of three random variables. It is shown that the sparse PC expansions which are produced by the algorithm can reach any prescribed accuracy, illustrating the heuristic convergence of the proposed scheme. Furthermore, the proposed procedure noticeably overperforms the classical strategy based on full PC expansions in terms of accuracy and efficiency. Then the method is tested on three finite element structural models. The first problem involves 10 random variables, whereas the other ones feature respectively 38 and $30 - 500$ random variables arising from the discretization of an input random field. These examples reveal the sparsity (i.e. the small effective dimension) of the model responses under consideration, hence the significant computational gain provided by the proposed scheme compared to full PC expansions. Moreover, adaptive LAR outperformed the stepwise regression algorithm proposed in [21,22], and can thus be regarded as an alternative to the latter. Note that the LAR-based procedure should be challenged with problems of large effective dimension in further studies.

The proposed method has been mainly applied so far to problems governed by elliptic stochastic partial differential equations. Other types of problem (e.g. hyperbolic, dynamic) should be considered in future investigations. In addition, the proposed method may be optimized by automatically updating the parameter q of the truncating norm when running the algorithm. Moreover, the computational cost could be dramatically reduced by integrating an *iterative* discretization of the random fields into the proposed algorithms. The investigation of a suitable algorithm for adding progressively the random variables into the metamodel is in progress.

Appendix A. Discussion of non intrusive techniques to compute the PC coefficients

Various non intrusive schemes have been described for estimating the PC coefficients. Stochastic collocation [15,14] and quadrature [58] are attractive methods since they rely upon well-established mathematical results in order to select optimal points in the experimental design. In the authors' opinion, quadrature should be preferred to stochastic collocation in practice since:

- The computational manipulation of multivariate Lagrange polynomials is cumbersome [59].
- Quadrature is based on an explicit representation in a PC basis, which allows straightforward post-processing (e.g. second moments and sensitivity indices).

The original tensor-product quadrature approach generally suffers the curse of dimensionality since the required number of model evaluations is given by $N = n^M$. To bypass this issue, one rather uses Smolyak quadrature [60] which leads to the following computational cost in high dimensions:

$$N \sim \frac{2^p}{p!} M^p, \quad M \to \infty, \tag{A.1}$$

On the other hand, an asymptotic equivalent of the number of terms in a PC expansion of degree p is obtained by

$$P = \binom{M+p}{p} \sim \frac{1}{p!} M^p, \quad M \to \infty. \tag{A.2}$$

Hence the ratio N/P tends to the factor 2^p for large M. The Smolyak construction has been labelled *optimal* in [61] insofar as this quantity does not depend on M. However the computational cost may be important if a great accuracy of the PC expansion (i.e. a large p) is required.

Regression appears to be a relevant approach in order to reduce the number of model evaluations. Indeed, many studies show that a number of model evaluations given by $N = kP$ with $k = 2, 3$ often provides satisfactory results. In particular, good empirical results have been obtained in [11,38] in the context of non intrusive stochastic finite elements. A limitation of the method lies in the problem of selecting the points in the experimental design though. It is worth mentioning that an algorithm has been devised in [12] to select a minimum number of roots of orthogonal polynomials in the design. Random designs may be also employed, which allows the derivation of statistical properties of the PC coefficients estimators [62]. Thus it is shown that *regression should outperform simulation* provided that a sufficiently accurate PC expansion (i.e. a large enough degree p) has been chosen.

Indeed, even if both the regression and simulation estimates converge at the (slow) rate $N^{-1/2}$, their associated constants are respectively given by $\mathbb{E}\left[\psi_{\boldsymbol{\alpha}}^2(\boldsymbol{X})\varepsilon^2(\boldsymbol{X})\right]$ and $\mathbb{E}\left[\psi_{\boldsymbol{\alpha}}^2(\boldsymbol{X})\mathcal{M}^2(\boldsymbol{X})\right]$, where $\varepsilon(\boldsymbol{X})$ denotes the remainder of the PC series [62]. Thus the regression estimates have typically a much smaller error than their simulation counterparts provided that the remainder $\varepsilon(\boldsymbol{X})$ is "small". Moreover, just as for simulation, the convergence rate of the regression estimates should be noticeably improved by using a more efficient sampling scheme than Monte Carlo, such as quasi-random sequences [52].

References

[1] R. Ghanem, P. Spanos, Stochastic Finite Elements: A Spectral Approach, Courier Dover Publications, 2003.
[2] C. Soize, R. Ghanem, Physical systems with random uncertainties: chaos representations with arbitrary probability measure, SIAM J. Sci. Comput. 26 (2) (2004) 395–410.
[3] B. Sudret, M. Berveiller, M. Lemaire, Eléments finis stochastiques en élasticité linéaire, C. R. Mécanique 332 (2004) 531–537.
[4] O.P. Le Maître, O.M. Knio, H.N. Najm, R.G. Ghanem, Uncertainty propagation using Wiener–Haar expansions, J. Comput. Phys. 197 (2004) 28–57.
[5] H.G. Matthies, A. Keese, Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations, Comput. Methods Appl. Mech. Eng. 194 (2005) 1295–1331.
[6] A. Nouy, A generalized spectral decomposition technique to solve stochastic partial differential equations, Comput. Methods Appl. Mech. Eng. 196 (45-48) (2007) 4521–4537.
[7] D.M. Ghiocel, R.G. Ghanem, Stochastic finite element analysis of seismic soil-structure interaction, J. Eng. Mech. 128 (2002) 66–77.
[8] O.P. Le Maître, M. Reagan, H.N. Najm, R.G. Ghanem, O.M. Knio, A stochastic projection method for fluid flow – II. Random process, J. Comput. Phys. 181 (2002) 9–44.
[9] S.K. Choi, R. V Grandhi, R.A. Canfield, C.L. Pettit, Polynomial chaos expansion with Latin Hypercube sampling for estimating response variability, AIAA J. 45 (2004) 1191–1198.
[10] M. Berveiller, B. Sudret, M. Lemaire, Non linear non intrusive stochastic finite element method – application to a fracture mechanics problem, in: G. Augusti, G.I. Schuëller, M. Ciampoli (Eds.), Proceedings of the 9th International Conference on Structure, Safety and Reliability (ICOSSAR'2005), Roma, Italy, Mill Press, Rotterdam, 2005.
[11] M. Berveiller, B. Sudret, M. Lemaire, Stochastic finite elements: a non intrusive approach by regression, Eur. J. Comput. Mech. 15 (1–3) (2006) 81–92.
[12] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, Reliab. Eng. Syst. Safe. 93 (2008) 964–979.
[13] M.S. Eldred, Recent advances in non-intrusive polynomial chaos and stochastic collocation methods for uncertainty analysis and design, in: 50th AIAA/ASME/ASCE/AHS/ASSC Structures, Structural Dynamics, and Material Conference, Palm Springs, California, 2009.
[14] D. Xiu, J.S. Hesthaven, High-order collocation methods for differential equations with random inputs, SIAM J. Sci. Comput. 27 (3) (2005) 1118–1139.
[15] I. Babuška, 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.
[16] T. Gerstner, M. Griebel, Dimension-adaptive tensor-product quadrature, Computing 71 (2003) 65–87.
[17] B. Ganapathysubramanian, N. Zabaras, Sparse grid collocation schemes for stochastic natural convection problems, J. Comput. Phys. 225 (2007) 652–685.
[18] P.B. Nair, A.J. Keane, Stochastic reduced basis methods, AIAA J. 40 (2002) 1653–1664.
[19] S.K. Sachdeva, P.B. Nair, A.J. Keane, Hybridization of stochastic reduced basis methods with polynomial chaos expansions, Probabilist. Eng. Mech. 21 (2) (2006) 182–192.
[20] A. Nouy, O.P. Le Maître, Generalized spectral decomposition for stochastic nonlinear problems, J. Comput. Phys. 228 (2009) 202–235.
[21] G. Blatman, B. Sudret, An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis, Probabilist. Eng. Mech. 25 (2010) 183–197.
[22] G. Blatman, B. Sudret, Efficient computation of Sobol' sensitivity indices using sparse polynomial chaos expansions, Reliab. Eng. Sys. Safe. 95 (11) (2010) 1216–1229.
[23] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, Ann. Stat. 32 (2004) 407–499.
[24] N. Wiener, The homogeneous chaos, Am. J. Math. 60 (1938) 897–936.
[25] D. Xiu, G.E. Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, SIAM J. Sci. Comput. 24 (2) (2002) 619–644.
[26] D. Xiu, D. Lucor, C.-H. Su, G.E. Karniadakis, Stochastic modeling of flow-structure interactions using generalized polynomial chaos, J. Fluid Eng. 124 (1) (2002) 51–59.
[27] D. Lucor, G.E. Karniadakis, Adaptive generalized polynomial chaos for nonlinear random oscillators, SIAM J. Sci. Comput. 26 (2) (2004) 720–735.

[28] W. Schoutens, Stochastic Processes and Orthogonal Polynomials, Springer-Verlag, New York, 2000.
[29] X. Wan, G. Karniadakis, Beyond Wiener–Askey expansions: handling arbitrary PDFs, J. Sci. Comput. 27 (2006) 455–464.
[30] J.A.S. Witteveen, S. Sarkar, H. Bijl, Modeling physical uncertainties in dynamic stall induced fluid–structure interaction of turbine blades using arbitrary polynomial chaos, Computers & Structures 85 (2007) 866–878.
[31] M. Loève, Probability Theory, Springer-Verlag, New York, 1977.
[32] B. Sudret, A. Der Kiureghian, Stochastic finite elements and reliability: a state-of-the-art report, Technical report UCB/SEMM-2000/08, University of California, Berkeley, 2000, p. 173.
[33] J. Zhang, B. Ellingwood, Orthogonal series expansion of random fields in reliability analysis, J. Eng. Mech. 120 (12) (1994) 2660–2677.
[34] K.K. Phoon, S.P. Huang, S.T. Quek, Implementation of Karhunen–Loève expansion for simulation using a wavelet-Galerkin scheme, Probabilist. Eng. Mech. 17 (3) (2002) 293–303.
[35] M. Grigoriu, Simulation of non-Gaussian translation processes, J. Eng. Mech. 124 (2) (1998) 121–126.
[36] N.D. Lagaros, G. Stefanou, M. Papadrakakis, An enhanced hybrid method for the simulation of highly skewed non-Gaussian stochastic fields, Comput. Methods Appl. Mech. Eng. 194 (2005) 4824–4844.
[37] M. Berveiller, Stochastic finite elements: intrusive and non intrusive methods for reliability analysis, Ph.D. Thesis, Université Blaise Pascal, Clermont-Ferrand, 2005.
[38] M. Berveiller, Eléments finis stochastiques: approches intrusive et non intrusive pour des analyses de fiabilité, Ph.D. Thesis, Université Blaise Pascal, Clermont-Ferrand, 2005.
[39] G. Blatman, Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis, Ph.D. Thesis, Université Blaise Pascal, Clermont-Ferrand, 2009.
[40] D.C. Montgomery, Design and Analysis of Experiments, John Wiley and Sons, New York, 2004.
[41] K.I. Babenko, Approximation by trigonometric polynomials in a certain class of periodic functions of several variables, Soviet Math. Dokl. 1 (1960) 672–675.
[42] H.J. Bungartz, M. Griebel, Sparse grids, Acta Numer. 13 (2004) 147–269.
[43] G. Blatman, B. Sudret, Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach, C. R. Mécanique 336 (6) (2008) 518–523.
[44] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.
[45] D. Allen, The prediction sum of squares as a criterion for selecting prediction variables, Technical report 23, Department of Statistics, University of Kentucky, 1971.
[46] R.G. Miller, The Jacknife – a review, Biometrika 61 (1974) 1–15.
[47] G. Saporta, Probabilités, analyse des données et statistique, second ed., Editions Technip, 2006.
[48] A.M. Molinaro, R. Simon, R.M. Pfeiffer, Prediction error estimation: a comparison of resampling methods, Bioinformatics 21 (2005) 3301–3307.
[49] O. Chapelle, V. Vapnik, Y. Bengio, Model selection for small sample regression, Mach. Learn. 48 (1) (2002) 9–23.
[50] D. Madigan, G. Ridgeway, Discussion of "least angle regression" by Efron et al., Ann. Stat. 32 (2004) 465–469.
[51] M.D. McKay, R.J. Beckman, W.J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics 2 (1979) 239–245.
[52] H. Niederreiter, Random Number Generation and Quasi-Monte Carlo Methods, SIAM, Philadelphia, PA, USA, 1992.
[53] W.J. Morokoff, R.E. Caflisch, Quasi-Monte Carlo integration, J. Comput. Phys. 122 (1995) 218–230.
[54] T. Ishigami, T. Homma, An importance quantification technique in uncertainty analysis for computer models, in: Proceedings of the ISUMA'1990, First International Symposium on Uncertainty Modelling Analysis, University of Maryland, 1990, pp. 398–403.
[55] A. Saltelli, K. Chan, E.M. Scott (Eds.), Sensitivity Analysis, Jhon Wiley and Sons, 2000.
[56] S.H. Lee, B.M. Kwak, Response surface augmented moment method for efficient reliability analysis, Struct. Safe. 28 (2006) 261–272.
[57] B. Sudret, A. Der Kiureghian, Comparison of finite element reliability methods, Probabilist. Eng. Mech. 17 (2002) 337–348.
[58] A. Keese, H.G. Matthies, Sparse quadrature as an alternative to Monte Carlo for stochastic finte element techniques, Proc. Appl. Math. Mech. 3 (2003) 493–494.
[59] D. Xiu, Fast numerical methods for stochastic computations: a review, Commun. Comput. Phys. 5 (2-4) (2009) 242–272.
[60] S.A. Smolyak, Quadrature and interpolation formulas for tensor products of certain classes of functions, Soviet. Math. Dokl. 4 (1963) 240–243.
[61] E. Novak, K. Ritter, Simple cubature formulas with high polynomial exactness, Constr. Approx 15 (1999) 499–522.
[62] A.B. Owen, Detecting near linearity in high dimensions, Technical report, Stanford University, Department of Statistics, 1998.