Optimized Sparse Polynomial Chaos Expansion With Entropy Regularization

SiJie Zeng

National University of Defense Technology <https://orcid.org/0000-0002-3858-641X>

XiaoJun Duan

National University of Defense Technology

JiangTao Chen

China Aerodynamics Research and Development Center

Liang Yan (✉ yanliang@nudt.edu.cn)

National University of Defense Technology <https://orcid.org/0000-0002-2168-5470>

Research Article

Keywords: Uncertainty quantification, Sparse polynomial chaos expansion, Entropy, Regularization

Posted Date: July 27th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-729114/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Advances in Aerodynamics on January 17th, 2022. See the published version at <https://doi.org/10.1186/s42774-021-00095-6>.

RESEARCH

Optimized Sparse Polynomial Chaos Expansion with Entropy Regularization

Sijie Zeng¹, Xiaojun Duan¹, Jiangtao Chen² and Liang Yan^{1*}

*Correspondence:

yanliang@nudt.edu.cn

¹College of Liberal Arts and Sciences, National University of Defense Technology, Changsha, China

Full list of author information is available at the end of the article

Abstract

Sparse Polynomial Chaos Expansion(PCE) is widely used in various engineering fields to quantitatively analyse the influence of uncertainty, while alleviate the problem of dimensionality curse. However, current sparse PCE techniques focus on choosing features with the largest coefficients, which may ignore uncertainties propagated with high order features. Hence, this paper proposes the idea of selecting polynomial chaos basis based on information entropy, which aims to retain the advantages of existing sparse techniques while considering entropy change as output uncertainty. A novel entropy-based optimization method is proposed to update the state-of-the-art sparse PCE models. This work further develops an entropy-based synthetic sparse model, which has higher computational efficiency. Two benchmark functions and a CFD experiment are used to compare the accuracy and efficiency between the proposed method and classical methods. The results show that entropy-based methods can better capture the features of uncertainty propagation, and the problem of over-fitting in existing sparse PCE methods can be avoided.

Keywords: Uncertainty quantification; Sparse polynomial chaos expansion; Entropy; Regularization

1 Introduction

Due to the variety of uncertainties frequently involved in engineering applications, which may cause fluctuations in the performance of a system, it is necessary to comprehensively consider the impact of uncertain factors[1, 2]. In order to deal with the problem of limited experiment resources, various surrogate models[3, 4] has been proposed, which aims at constructing a mathematical model to accurately mimic the behavior of the original problem with an affordable experimental design[5]. Popular surrogate techniques includes the Kriging method[6, 7, 8, 9], artificial neural network[10, 11, 12], polynomial chaos expansion (PCE) method[13, 14, 15, 16] *etc.* In this paper, we focus on the PCE model, which has been widely used in engineering to quantify uncertainties efficiently.

The basic idea of PCE is to expand an exact solution of a random process space by polynomial expansion[17, 18, 19]. Many successful applications of uncertainty quantification based on the PCE method have been achieved[20, 21, 22]. However, the cost of constructing a PCE model increases exponentially with the dimension of input parameters, *i.e.* the curse of dimensionality, hence severely restricts practical applications of the model at the industrial level[23, 24, 25]. In order to solve this problem, lots of sparse algorithms has been developed in recent years. Some typical

examples of well-established methods are the sparse regression method[26, 27] and the compressive sampling method[28, 29].

A problem with some OLS-solution methods is the possibility of over-fitting but can be avoided by regularization, which is another practical solution to seek a *sparse* representation[30]. Regularization is a typical method of model selection, which introduces a regularization term to the empirical risk to minimize the risk. From the perspective of Bayesian estimation, the regularization term corresponds to the prior probability of the model[31]. There are many regularization methods, but it is ℓ_0 regularization that is required to obtain a real sparse model. However, solving model with a ℓ_0 regularization term is a NP-Hard problem[32, 33], so researchers usually adopt the ℓ_1 regularization when dealing with sparse problems[34], which is an optimal convex approximation of the ℓ_0 paradigm, and can easy to optimize the solution to obtain the sparse model.

Although scholars have conducted a considerable amount of research on sparse polynomial chaos expansions, most of the existing sparse methods are derived from the perspective of regression[35]. In these methods, researchers usually pay attention to one or two numerical characteristics in the distribution, which are important statistics to measure uncertainty. However, the distribution is very complicated in common situation, especially when there is a noise in the original problem. Furthermore, the aforementioned methods focus on statistics, such as the relative mean squared error (MSE) or the cross validation error, to optimize parameters, which ignores the overall uncertainties propagated from the input variables.

To depict the *full* uncertainty of the output and leverage the impact of over-fitting, we introduce information entropy as a measure of evolution of the output distribution to further optimize the sparse representation of a PCE model. Information entropy was first introduced to measure microscopic uncertainty and multiplicity by Shannon[36]. The idea was then extended to learning tasks to measure the information changes with difference statistical models. Jaynes [37] proposed the maximum entropy principle (MEP) to provide an optimization criterion. B.Grechuk *et al.*[38] have studied the relationships among error, likelihood, and entropy in regression analysis, and found that a normal distribution can be recovered from the maximum entropy principle. Wang[39] introduced a maximum entropy penalty to model and incorporate the entropy controlled framework with other conventional learning algorithms. Liang[40] proposed a sparse subspace clustering with entropy-norm by using information entropy as the regularization term of the objective function. Researchers further makes variable selection through the evaluation of entropy, and extended the use of entropy as an important criterion for model selection[41, 42], which can be embedded in existing methods, but does not form a framework.

In this work, we aim to retain the advantages of existing sparse PCE techniques while considering entropy change as output uncertainty. Hence, we propose a novel method where entropy is introduced as an auxiliary regularization term. Firstly, a general re-optimization model which choose the features with the largest information entropy based on existing optimized model. Although such a strategy is easy to implement, it may ignore some key features in the first stage. Hence, a hybrid entropy-based synthetic method which embedded in several commonly used classic sparse criteria is proposed. The novel regularization structure takes the value

and volatility of each feature into account simultaneously. In fact, the entropy term can be regarded as a trade-off between the predictive mean value and uncertainty along with the selected features. The above two algorithms can be easily computed with existing sparse algorithms, hence improves their usage scope and availability. Experiments are performed to compare the proposed two algorithms and classical sparse methods, *i.e.* the Orthogonal Matching Pursuit(OMP), the Least Angle Regression(LARS), the Subspace Pursuit(SP) and the Bayesian Compressive Sensing(BCS). Results show that the general re-optimization method can be optimized on the basis of any sparse PCE model, so it has the benefits of strong universality and simple operation while the hybrid entropy-based synthetic method has strong pertinence, and can achieve better optimization performance in specific situations.

The remainder of the paper is organized as follows. Section 2 introduces the fundamental theory of sparse PCE models. Section 3 proposes the theories and algorithms of the two optimization methods mentioned in the previous paragraph. Section 4 shows the performance of the proposed methods, illustrated by numerical examples and comparisons. Finally, the conclusions are summarized in Section 5.

2 The fundamental theory

2.1 Sparse polynomial chaos expansion

The basics of PCE are elaborated briefly as follows. Let $\mathcal{M}$ be the original model and $Y = \mathcal{M}(\xi)$ be the output. The finite-order PCE model expanded in the full polynomial space can be expressed as:

$$Y \approx Y_{PC} = \sum_{i=0}^{P-1} c_i \Psi_i(\xi), \quad (1)$$

where ξ is d -dimensional random variables with a probability density function (PDF) of $f(\xi)$, and $\Psi_i(\xi)$ is a set of polynomial basis functions truncated at the p -th order with P denotes the degree of the PCE. Generally, P varies with different truncation scheme. For example, when the *total degree space* are chosen, P

$$P = \frac{(p+d)!}{p!d!}. \quad (2)$$

with p as the maximum degree for each dimension. Given an design of experiment(DoE) $\{\xi, Y\}$, where $\xi = (\xi_1, \xi_2, \dots, \xi_n) \in \mathbb{R}^{d \times n}$ is a specific sample and $Y(\xi) \in \mathbb{R}^n$ are corresponding responses. The main effort of constructing a PCE model is solving the following generalized linear equation system:

$$\begin{bmatrix} Y(\xi_0) \\ Y(\xi_1) \\ \vdots \\ Y(\xi_n) \end{bmatrix} = \begin{bmatrix} \Psi_0(\xi_0) & \Psi_1(\xi_0) & \cdots & \Psi_{P-1}(\xi_0) \\ \Psi_0(\xi_1) & \Psi_1(\xi_1) & \cdots & \Psi_{P-1}(\xi_1) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_0(\xi_n) & \Psi_1(\xi_n) & \cdots & \Psi_{P-1}(\xi_n) \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{P-1} \end{bmatrix} \quad (3)$$

which can be rewritten in matrix form:

$$\mathbf{Y} = \mathbf{\Psi} \mathbf{c}. \quad (4)$$

Although we are able to obtain a least squares estimate $\hat{\mathbf{c}} = (\Psi^T \Psi)^{-1} \Psi^T \mathbf{Y}$ easily, the degree P increases dramatically with p and d , hence we have to solve a under determined system with limited DoE resources, the least squares solution can be inaccurate and unstable. The sparse PCE method is proposed to solve this problem, and it restores the complete model response almost accurately by selecting a small number of basis that dominate the system output. ℓ_0 -norm is the most widely used regularization criterion, i.e. limit the degree of non-zero coefficients by solving the following optimization problem:

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c} \in \mathbb{R}^P} \|\mathbf{c}\|_0 \text{ subject to } \mathbf{Y} = \Psi \mathbf{c}. \quad (5)$$

However, eq. (5) is a non-convex optimization problem, which can be difficult to solve in practice due to its NP-hardness. A typical practical technique is to replace the ℓ_0 norm term with the ℓ_1 norm term[43], and the new objective function becomes

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c} \in \mathbb{R}^P} \|\mathbf{c}\|_1 \text{ subject to } \mathbf{Y} = \Psi \mathbf{c}. \quad (6)$$

2.2 Error estimation and model evaluation

The optimization problem defined in eq. (6) are usually solved by relaxing the constraint, *i.e.*

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c} \in \mathbb{R}^P} \|\mathbf{c}\|_1 \text{ subject to } \|\Psi \mathbf{c} - \mathbf{Y}\|_2 \leq \varepsilon, \quad (7)$$

where ε is recorded as the truncation error, determined by measurement noise, which is a more natural parameterization choice. If ε is too large, the reconstructed PC is not accurate enough; however, if ε is too small, the reconstructed PC may be over-fitting. The leave-one-out(LOO) error is used to measure the degree of over-fitting in practice, which can be expressed as follows:

$$E_{LOO} = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - y_{PC}(\xi_i)}{1 - h_i} \right)^2, \quad (8)$$

where h_i denotes the i -th diagonal element of the matrix $\Psi (\Psi^T \Psi)^{-1} \Psi^T$. The E_{LOO} can be calculated easily when the least squares solution *w.r.t* full design are available (see [44]). We choose the PCE model with the smallest LOO error during the iteration of finding the best sparse solution.

When comparing the performance of different surrogate models, our interest lies in the precision and uncertainty. The relative mean squared error (MSE) is a widely used measure to quantify the precision of a surrogate, which is defined by the following equation:

$$\text{Relative MSE} = \frac{\sum_{i=1}^n (y_i - y_{PC}(\xi_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (9)$$

where $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. As for the uncertainty, we pay special attention to the probability density function (PDF) or cumulative distribution function (CDF).

2.3 The concept of information entropy

The type of entropy addressed in this work is information entropy, which first proposed by Shannon in 1948[36]. The most general explanation of entropy is the measure of uncertainty. Uncertainty refers to the information contained in the system[45]. The more information the system contains, the smaller the uncertainty and corresponding entropy.

For any random variable $X = \{x_1, \dots, x_n\}$, the definition of information entropy is expressed in terms of a discrete set of probabilities $P(x_i)$, as follows:

$$H(X) = - \sum_{i=1}^n P(x_i) \ln(P(x_i)). \quad (10)$$

Similarly, for a continuous random variable, differential entropy is analogous to entropy, which is an extension of Shannon entropy and can be obtained as follows:

$$H(x) = - \int p(x) \ln p(x) dx. \quad (11)$$

Based on information entropy, when predicting a probability model of random variables, the main idea of the MEP is that among all candidate distributions, the distribution that maximizes H should be selected. Specifically, the best probability distribution has maximal entropy, which comes up with an optimization problem:

$$\text{maximize } H = - \int p(y) \ln p(y) dy \text{ subject to } \|\Psi \mathbf{c} - \mathbf{Y}\|_2 \leq \varepsilon, \quad (12)$$

where $p(y)$ is the probability density function of element y in $\mathbf{Y}$.

Therefore, by combining MEP and the PCE model, entropy can be used as a sparse optimization criterion for the sparse PCE model. Thus, it can make the model sparse while retaining the largest amount of uncertainty.

3 Methodology

In this section, two adaptive methods are proposed. The first is the re-optimization of results achieved with the alternative model based on entropy, which can be thought of as a simple external plugin so has the characteristics of simple operation and strong versatility. The second is hybrid optimization using entropy as a regularization term, which is the main method of this paper and it is faster, more targeted and efficient.

3.1 Entropy-based re-optimization sparse PCE

As mentioned at the end of the 2.3 section, our goal is to sparse the coefficients and preserve the maximum uncertainty of the results. Therefore, we can directly perform entropy-based secondary optimization from Eq.12 on the basis of existing sparse algorithms (i.e., entropy-based re-optimization sparse PCE). Typically, solving problems with the sparse PCE surrogate model can be converted to:

$$\mathbf{Y} = \sum_{i \in \mathcal{I}} \Psi_i \mathbf{c}, \quad (13)$$

where $\mathcal{I}$ is the sparse index set.

After solving by the classic sparse PCE method, we get the model response $\mathbf{Y}_S$. Additional model responses are produced through the repeated sampling process, which can then be re-optimized based on entropy sorting. According to the MEP, the best probability distribution is the one with maximal entropy. Therefore, our goal is to select W items with the largest entropy values. The objective function during re-optimization is defined as follows:

$$\hat{\mathbf{c}} = \arg \min_{\mathcal{J} \subset \mathcal{I}} \|\Psi_{\mathcal{J}} \mathbf{c} - \mathbf{Y}_S\|_2, \quad (14)$$

where $\mathbf{Y}_S$ is the model response obtained by the classic method, and index set $\mathcal{J}$ is satisfied:

$$\mathcal{J} = \arg \max_{\mathcal{J} \subset \mathcal{I}} H(\mathbf{Y}_S) \text{ subject to } \text{card}(\mathcal{J}) \leq W, \quad (15)$$

with $H(\mathbf{Y}_S) \triangleq \sum_{j=1}^W H(\mathbf{y}_j)$, where $\mathbf{y}_j = \Psi_j \mathbf{c}_j$.

Since calculating the differential entropy of $\mathbf{Y}_S$ is more complicated, it can be approximated by the Shannon entropy of Monte Carlo. For example, if we repeat sampling m times as the verification set, the Shannon information entropy of $\mathbf{y}_j$ can be obtained:

$$H(\mathbf{y}_j) = - \sum_{i=1}^m \mathbf{y}_{ij}^2 \ln \mathbf{y}_{ij}^2. \quad (16)$$

The pseudo code of the algorithm for entropy based re-optimization sparse PCE is as follows:

Algorithm 1 Entropy based re-optimization sparse PCE(Ent-PCE)

Input: Ψ : a set of candidate basis functions; $\mathbf{y}$: observation vector. W : desired sparsity of Ent-PCE;
 1: Establish a surrogate model based on a classic sparse PCE algorithm.
 2: Repeat sampling m times as the verification set and get the response $\mathbf{Y}_S$.
 3: **for** each $j = 1, \dots, p$ **do**
 4: $\mathbf{y}_j = \Psi_j \mathbf{c}_j$
 5: Calculate Shannon entropy of $\mathbf{y}_j$: $H_j = - \sum_{i=1}^m \mathbf{y}_{ij}^2 \ln \mathbf{y}_{ij}^2$
 6: **end for**
 7: Sort the indicators according to $\mathbf{H}$; $\mathcal{J}$ is W indices corresponding to the largest value of entropy.
 8: Least-squares solution based on index set $\mathcal{J}$: $\hat{\mathbf{c}} = \arg \min_{\mathcal{J} \subset \mathcal{I}} \|\Psi_{\mathcal{J}} \mathbf{c} - \mathbf{Y}_S\|_2$

The advantages of this method are that it is simple and easy to implement, as well as natural from a theoretical viewpoint, and the selection of parameters is very reliable. From an experimental viewpoint, this method can be adapted to almost all existing sparse algorithms without modifying their internal codes (i.e., it is more universal and maneuverable). However, due to the need to perform secondary calculations, and the OLS algorithm is used in the re-optimization process, longer calculation times may be required when dealing with high-dimension situations.

3.2 Hybrid entropy-based comprehensive sparse PCE

In section 3.1, a universal entropy-based re-optimization sparse PCE method is proposed. However, due to its poor pertinence, the optimization effect is also limited.

In fact, there are several commonly used sparse algorithms in the literature, some of which can be found in open-source toolboxes. In this section, for these specific sparse PCE algorithms, a method involving hybrid entropy-based comprehensive sparse PCE will be proposed.

In the PCE model, the unknown coefficients correspond to polynomial chaos basis. In the process of sparsing the chaos polynomial, we reduce the dimension of the coefficient, which can be seen as a model selection for the polynomial chaos basis. From the perspective of regression, we usually choose the feature basis that has a greater impact on the residual, such as the OMP method. However, from the MEP perspective, the feature basis with larger entropy contains more uncertainty information and is more important to the model response. Thus, the polynomial chaos basis stores the probabilistic information of the predicted response to some extent.

To maximize the input uncertainty distribution in the response, we introduce the entropy of the basis function to regularize the classic model. From this perspective, the objective function is re-defined as follows:

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c} \in \mathbb{R}^P} \|\Psi \mathbf{c} - \mathbf{Y}\|_2 + \lambda (\|\mathbf{c}\|_1 + \gamma H(\Psi_{\mathcal{I}})), \quad (17)$$

An example of a hybrid entropy-based comprehensive sparse PCE algorithm based on the classic OMP method is Algorithm 2.

Algorithm 2 Hybrid entropy-based orthogonal matching pursuit(HEnt-OMP)

Input: Ψ : a set of candidate basis functions; $\mathbf{y}$: observation vector.
1: Initialize: feature index set: $\mathcal{I} = \emptyset$; residual vector: $\mathbf{r} = \mathbf{y}$
2: **for** each $i = 1, \dots, n$ **do**
3: Find $\Psi_i \in \Psi$ which is closely related to the residual and has a larger entropy value by solving:
4: $\Psi_i = \arg \max_{\Psi_i \in \Psi \setminus \Psi_{\mathcal{I}}} \left(\frac{|\langle \Psi_i, \mathbf{r} \rangle|}{\|\Psi_i\|_2} - \gamma \sum_{j=1}^P \Psi_{ij}^2 \ln(\Psi_{ij}^2) \right)$
5: Update the index set: $\mathcal{I} = \mathcal{I} \cup \{i\}$
6: Compute the coefficients $\mathbf{c}^*$ using only the active indices by least-squares:
7: $\mathbf{c}^* = \arg \min_{\mathbf{c} \in \mathbb{R}} \|\mathbf{y} - \Psi_{\mathcal{I}} \mathbf{c}^*\|_2$
8: Update the residual: $\mathbf{r} = \mathbf{y} - \Psi_{\mathcal{I}} \mathbf{c}^*$
9: **end for**

In each iteration, OMP selects the regressor that is most closely correlated with the current residual, while HEnt-OMP introduces entropy as regularization on this basis. Therefore, the regressor must not only be related to the residual but also ensure that its entropy value is large.

In the sparse process, traditional OMP methods often suffer from over-fitting and unreliable LOO error estimates, which can be detrimental in basis-adaptive settings. By introducing entropy as regularization, the uncertainty in the results can be preserved, and the problem of OMP can also be solved effectively. Furthermore, compared with Algorithm 1, this method is more targeted and has faster computing speeds. However, due to the different classic methods, the selection of regularization parameters in different methods is also worthy of attention.

3.3 Parameter optimization and selection

Before conducting the experiment, we need to set some parameters for these algorithms. For example, in Algorithm 1, we are concerned about the effectiveness of

changing parameter W on the effect of the final sparse result. Similarly, when using hybrid entropy-based algorithms, γ is a parameter worthy of attention. For each of the various methods used in entropy-based algorithms, the choice of parameters will be different.

The choice of parameter W is performed automatically: we can directly set W so that it satisfies the desired active basis of size. Or given a prior interval to conduct a preliminary experiment, and then select the best W .

Regarding the selection of γ , consider using a small number of sample points in preliminary experiments before conducting the formal experiment, so that γ can get an ideal value. The algorithm is designed as Algorithm 3, the idea of which comes from dichotomy.

Algorithm 3 Parameter selection

Input: $\mathcal{S}$: surrogate model with unknown γ ; γ_9 : the upper bound of γ

```

1: while The variance of Relative MSE is less than  $\varepsilon$  do
2:   for each  $i = 0, \dots, 9$  do
3:      $l = \frac{\gamma_9 - \gamma_0}{10}, \gamma_i = \gamma_0 + i \times l$ 
4:     Run the model  $\mathcal{S}$  with  $\gamma_i$  and calculate Relative MSE:  $e_i$ 
5:   end for
6:   Find the index  $i$  that minimizes the Relative MSE:  $i = \arg \min_i e_i$ 
7:   Update the boundary:  $\gamma_0 = \gamma_i - \frac{1}{2}l, \gamma_9 = \gamma_i + \frac{1}{2}l$ 
8: end while

```

4 Experiment Analysis

In the experiments, we analyze and compare the effectiveness of entropy-based sparse PCE and classic sparse PCE algorithms: OMP[46], LARS[44], SP[47], and BCS[48].

To implement the benchmark, we consider an Ishigami function($d = 3$) and an OAKLEY & O'HAGAN function($d = 15$) to show the performance of the method in low-dimensional and high-dimensional cases, respectively. Both benchmark functions have been inserted small random interference items to better fit the actual. The LHS sequence is used to generate sample points in experiment. To further prove the effective performance of the proposed method, the approach is applied to a more realistic CFD application: RAE2822 airfoil($d = 5$). For all experiments, we use the general-purpose uncertainty quantification software UQLab[35].

4.1 Toy model

Before starting the formal experiment, we first consider a toy model to explain the problem of the paper. The traditional sparse method only considers the value of the coefficient, rather than the impact of feature changes on uncertainty. So we constructed a function that is a combination of smooth function and high-order features.

$$f(x) = 6\sin(x) + 4x(\cos(x) + 1) + 5x^6 + 10^{-10}9^x \cdot x^{22} \quad (18)$$

If classic methods such as OMP are used directly, these high-level features will be deleted directly. It can be seen from the results that the MSE is good, but there is a

certain difference between the final PDF and the real PDF (KL divergence). Using the two methods proposed in this paper can reduce the KL divergence, because some high-order features are also reserved.

4.2 Case 1: The Ishigami function

The Ishigami model is a non-linear, non-monotonic smooth three-dimensional function:

$$Y(x_1, x_2, x_3) = \sin x_1 + a \sin^2 x_2 + bx_3^4 \sin x_1, \quad (19)$$

where x_1 , x_2 , and x_3 are three independent input random variables uniformly distributed on $[-\pi, \pi]$, and in typical $a = 7$, $b = 0.1$.

In the case of obtaining better parameters, γ and W , we add the entropy based re-optimization method and hybrid entropy-based comprehensive method to other sparse methods for optimization, and the results are as follows.

It can be seen from Table 2 that both optimization methods can be further sparse on the existing classic sparse PCE methods in low dimensions. It is worth noting that Ent-PCE shows its advantage in that it can control the sparsity level of the coefficient. Further results are shown below.

In the Fig.2, both optimization methods can achieve a good degree of optimization in low dimensions, and the optimization effect of Ent-PCE is better. While the level of sparsity is reduced, only fewer points need to be used to achieve a smaller relative MSE. The conclusion is, when the same sampling point is used, the entropy-based methods can obtain higher accuracy. Furthermore, in the PDF image, the optimized PDF is closer to the sample point, which is consistent with the previous theory.

4.3 Case 2: The OAKLEY & O'HAGAN function

The OAKLEY & O'HAGAN function was proposed in 2010[49]. It is a high-dimensional function and commonly used in uncertainty quantification and sensitivity analysis, which is expressed as follows:

$$f(\mathbf{x}) = \mathbf{a}_1^T \mathbf{x} + \mathbf{a}_2^T \sin(\mathbf{x}) + \mathbf{a}_3^T \cos(\mathbf{x}) + \mathbf{x}^T \mathbf{M} \mathbf{x}. \quad (20)$$

The independent distributions of the input random variables are $x_i \sim \mathcal{N}(0, 1), i = 1, \dots, 15$. $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ are 1×15 vectors, and $\mathbf{M}$ is a square matrix, and the data are taken directly from [49].

Comparing Table 2 and 3, because of the principle of BCS is related to uncertainty, so the optimization effect of HEnt-PCE is more obvious, similar to SP and Ent-PCE. In addition, when the dimensionality increases, the calculation speed of Ent-PCE is significantly slower, while HEnt-PCE remains a faster calculation speed. As a result, if there are no specific requirements for sparsity, it is not recommended to use Ent-PCE in high-dimensional situations.

From the results of Relative MSE shown in Fig.3, in the case of high dimensionality, both optimization methods can achieve optimization. However, Ent-PCE can only be slightly optimized on the basis of the existing results, while HEnt-PCE can achieve better optimization results, especially when the sample points are insufficient. This may be due to the fact that in the case of insufficient sample points,

the uncertainty contained is greater, and important features can be better proposed when considering entropy. With an increased number of sample points, the dependence of the result on uncertainty factors decreases, so the effect will gradually converge.

4.4 Case 3: RAE2822 airfoil

The final application involves the RAE2822 airfoil, a supercritical airfoil, which is a challenging case used by various researchers [50, 51] to test for quantification of uncertainty. The reference value of uncertainty factors in the test are shown in Table 4. We specify that all random variables are subject to a uniform distribution, with the upper and lower boundaries at a reference value of $1 \pm 5\%$.

For deterministic CFD solutions, the Spalart-Allmaras one-equation turbulence model is used for turbulence modeling, and SU2 simulation software is used as a black box solver. The model of 5rd order, using different sparse PCE method is constructed with 100 samples. The results (pressure coefficient) obtained with four classic sparse PCE methods and two improved methods proposed above. The mean, standard deviation, the CDF image of one random point in the RAE2822 airfoil and relative error are shown in Fig.4(OMP), Fig.5(LARS), Fig.6(SP), Fig.7(BCS).

For Fig.4-7, it can be observed that the results of optimized models are in acceptable agreement with the results of the classical models, and are consistent with the mean of the sampling points (show in (a)). The main difference between classical models and corresponding optimized models is the standard deviation (show in (b)). Each position will form a standard deviation with respect to the parameters, because of the sensitivity of the RAE2822 airfoil model. The standard deviation of the classical models are usually greater than the actual results, while optimized models can reduce the standard deviation and achieved better results by adjusting the parameters. However, the least squares used in Ent-PCE will make the standard deviation extremely small. Moreover, although the dimensionality is not high, one REA2822 case needs to build hundreds of PCE models (each position of the airfoil needs a model, the number is related to the density of grid nodes), which takes longer calculation time and cost for Ent-PCE. Thus, HEnt-PCE is a better choice in such situation. Since the mean and standard deviation of each model are available, the distribution of result at each position can also be obtained. CDF image were evaluated for the distribution results at a random point on the airfoil (shown in (c)). It can be seen that optimized models greatly improves the distribution of the model on the basis of the classic models. Although the relative error of the mean hardly changes, the optimized models reduce the relative error of the standard deviation(show in (d) and (e)), so the result will be closer to the original distribution.

5 Conclusion

In this article, an effective framework of sparse PCE is realized and verified. The main contribution of this paper is the development of two adaptive regression methods for optimizing sparse polynomial chaos expansion, Ent-PCE and HEnt-PCE. The former can be regarded as an external plugin and can be optimized on the basis of any sparse PCE model, so it has the benefits of strong universality and simple

operation; the latter has strong pertinence, embeds the regular term into a known sparse algorithm, and can select ideal parameters in different situations to achieve better optimization performance.

The advantages of the proposed method are twofold. First, while retaining the accuracy of the model, the problem of over-fitting in the classic algorithm is solved. Secondly, considering the uncertainty propagation in the model, the distribution rules of model input are passed, so the output retains uncertainty to the greatest extent.

Furthermore, due to inconsistencies in the original standards for model selection, the regularization parameters also change. To achieve the best optimization effect, a set of parameter selection rules was formed. For verification, several applications were considered. Two benchmark analysis functions were initially studied. The results show that the two proposed algorithms can achieve a certain optimization effect. Among them, Ent-PCE is simpler and more universal, as well as more practical, in low-dimensional situations and can form a more sparse model. In the case of high dimensions, because HEnt-PCE is more targeted, the results are more accurate, and the calculation speed is faster. Additionally, the algorithm effect was analyzed and compared through the wind tunnel application. The experimental results show that Ent-PCE has greater limitations, but it can be used independently, and it can be easily optimized for small-scale models. HEnt-PCE is more efficient and can achieve higher accuracy, so is an effective method that worth to be promoted at the industrial level to establish a high-quality sparse PCE model.

Acknowledgements

The authors are thankful to the reviewers for their valuable comments to improve the quality of the manuscript.

Funding

This work was funded by the National Numerical Wind tunnel project (NNW2019ZT7-B23) and National Natural Science Foundation of China (1771450). The authors gratefully acknowledge this funding.

Availability of data and materials

The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors read and approved the final manuscript.

Author details

¹College of Liberal Arts and Sciences, National University of Defense Technology, Changsha, China. ²China Aerodynamics Research and Development Center, Mianyang, China.

References

1. Tang, T., Zhou, T.: Recent developments in high order numerical methods for uncertainty quantification. *entia Sinica* **58**(7), 891 (2015)
2. Hu, X., Parks, G.T., Chen, X., Seshadri, P.: Discovering a one-dimensional active subspace to quantify multidisciplinary uncertainty in satellite system design. *Advances in Space Research* **57**(5), 1268–1279 (2016)
3. Ky, Khac, Vu, Claudia, D'Ambrosio, Youssef, Hamadi, Leo, Liberti: Surrogate-based methods for black-box optimization. *International Transactions in Operational Research* (2017)
4. Dutta, S., Gandomi, A.H.: Design of experiments for uncertainty quantification based on polynomial chaos expansion metamodels, pp. 369–381 (2020)
5. Lüthen, N., Marelli, S., Sudret, B.: Sparse polynomial chaos expansions: Literature survey and benchmark (2020)
6. Gaspar, B., Teixeira, A.P., Soares, C.G.: Assessment of the efficiency of kriging surrogate models for structural reliability analysis. *Probabilistic Engineering Mechanics* **37**(jul.), 24–34 (2014)
7. Zhang, L., Lu, Z., Pan, W.: Efficient structural reliability analysis method based on advanced kriging model. *Applied Mathematical Modelling* **39**(2), 781–793 (2015)

8. Sen, O., Gaul, N.J., Choi, K.K., Jacobs, G., Udaykumar, H.S.: Evaluation of kriging based surrogate models constructed from mesoscale computations of shock interaction with particles. *Journal of Computational Physics* **336**, 235–260 (2017)
9. Song, K., Zhang, Y., Zhuang, X., Yu, X., Song, B.: An adaptive failure boundary approximation method for reliability analysis and its applications. *Engineering with Computers*, 1–16 (2020)
10. Dai, H., Hao, Z., Rasmussen, K., Wei, W.: Wavelet density-based adaptive importance sampling method. *Structural Safety* **52**, 161–169 (2015)
11. Li, S., Yang, B., Qi, F.: Accelerate global sensitivity analysis using artificial neural network algorithm: Case studies for combustion kinetic model. *Combustion & Flame* **168**, 53–64 (2016)
12. Tripathy, R.K., Ilias, B.: Deep uq: Learning deep neural network surrogate models for high dimensional uncertainty quantification. *Journal of Computational Physics* **375**, 565–588 (2018)
13. Bhosekar, A., Ierapetritou, M.: Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Computers & Chemical Engineering* **108**(jan.4), 250–267 (2018)
14. Xu, J., Kong, F.: A cubature collocation based sparse polynomial chaos expansion for efficient structural reliability analysis. *Structural Safety* **74**, 24–31 (2018)
15. Mohammadi, A., Raisee, M.: Efficient uncertainty quantification of stochastic heat transfer problems by combination of proper orthogonal decomposition and sparse polynomial chaos expansion. *International Journal of Heat and Mass Transfer* **128**(JAN.), 581–600 (2019)
16. Tarakanov, A., Elsheikh, A.H.: Regression-based sparse polynomial chaos for uncertainty quantification of subsurface flow models. *Journal of Computational Physics* **399**, 108909 (2019)
17. Ghanem, R.G., Spanos, P.D.: *Stochastic Finite Elements: A Spectral Approach*. Springer, ??? (1992)
18. Xiu, D., Karniadakis, G.E.: The wiener–askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing* **24**(2), 619–644 (2002)
19. Xiu, D., Karniadakis, G.E.: Modeling uncertainty in flow simulations via generalized polynomial chaos. *Journal of Computational Physics* **187**(1), 137–167 (2003)
20. Schaefer, J., Hosder, S., Iv, T.K.W., Rumsey, C., Carlson, J.-R., Kleb, W.: Uncertainty quantification of turbulence model closure coefficients for transonic wall-bounded flows. *AIAA Journal* **55**(1), 195–213 (2017)
21. Zhang, W., Wang, X., Yu, J., Yan, C.: Uncertainty quantification analysis in hypersonic aerothermodynamics due to freestream. *Beijing Hangkong Hangtian Daxue Xuebao/journal of Beijing University of Aeronautics & Astronautics* **44**(5), 1102–1109 (2018)
22. Avdonin, A., Jaensch, S., Silva, C.F., Češnovar, M., Polifke, W.: Uncertainty quantification and sensitivity analysis of thermoacoustic stability with non-intrusive polynomial chaos expansion. *Combustion and Flame* **189**, 300–310 (2018)
23. Eldred, M.S., Burkardt, J.: Comparison of non-intrusive polynomial chaos and stochastic collocation methods for uncertainty quantification. In: 47th AIAA Aerospace Sciences Meeting Including The New Horizons Forum and Aerospace Exposition (2009)
24. Kumar, D., Raisee, M., Lacor, C.: An efficient non-intrusive reduced basis model for high dimensional stochastic problems in cfd. *Computers & Fluids* **138**, 67–82 (2016)
25. Huan, Z., Gao, Z., Xu, F., Zhang, Y., Huang, J.: An efficient adaptive forward-backward selection method for sparse polynomial chaos expansion. *Computer Methods in Applied Mechanics and Engineering* **355**(2019), 456–491 (2019)
26. Blatman, G., Sudret, B.: Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach. *Comptes Rendus Mécanique* **336**(6), 518–523 (2008)
27. Blatman, G., Sudret, B.: Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics* **230**(6), 2345–2367 (2011)
28. Hampton, J., Doostan, A.: Compressive sampling of polynomial chaos expansions: Convergence analysis and sampling strategies. *Journal of Computational Physics* **280**, 363–386 (2015)
29. Salehi, S., Raisee, M., Cervantes, M.J., Nourbakhsh, A.: Efficient uncertainty quantification of stochastic cfd problems using sparse polynomial chaos and compressed sensing. *Computers & Fluids* **154**, 296–321 (2017)
30. Liu, Z., Lesselier, D., Sudret, B., Wiart, J.: Surrogate modeling based on resampled polynomial chaos expansions (2018)
31. Vermet, F.: *Statistical Learning Methods*. Big Data for Insurance Companies, ??? (2018)
32. Louizos, C., Welling, M., Kingma, D.P.: Learning sparse neural networks through l_0 regularization (2017)
33. Liu, Z., Sun, F., McGovern, D.P.: Sparse generalized linear model with l_0 approximation for feature selection and prediction with big omics data. *Biodata Mining* **10**(1), 39 (2017)
34. Jakeman, J.D., Eldred, M.S., Sargsyan, K.: Enhancing l_1 -minimization estimates of polynomial chaos expansions using basis selection. *Journal of Computational Physics* **289**, 18–34 (2015)
35. Marelli, S., Sudret, B.: Uqlab: a framework for uncertainty quantification in matlab. In: *SIAM Conference on Uncertainty Quantification* (2015)
36. Shannon, C.E.: A mathematical theory of communication. *Bell System Technical Journal* **27**(3), 379–423 (1948)
37. Jaynes, E.T.: Information theory and statistical mechanics. *Physical Review* **108**(2), 171–190 (1957)
38. Grechuk, B., Zabaranin, M.: Regression analysis: likelihood, error and entropy. *Mathematical Programming* (2018)
39. Wang, X., Tao, D., Zhu, L.: Entropy controlled laplacian regularization for least square regression. *Signal Processing* **90**(6), 2043–2049 (2010)
40. Bai, L., Liang, J.: Sparse subspace clustering with entropy-norm. In: *International Conference on Machine Learning (ICML 2020)* (2020)
41. Obuchi, T., Nakanishi-Ohno, Y., Oka Da, M., Kabashima, Y.: Statistical mechanical analysis of sparse linear regression as a variable selection problem. *Journal of Statistical Mechanics Theory & Experiment* **2018**(10) (2018)
42. Murari, A., Peluso, E., Cianfrani, F., Gaudio, P., Lungaroni, M.: On the use of entropy to improve model

- selection criteria. *Entropy* **21**(4), 394 (2019)
43. Bruckstein, A.M., Donoho, D.L., Elad, M.: From sparse solutions of systems of equations to sparse modeling of signals and images. *Siam Review* **51**(1), 34–81 (2009)
 44. Blatman, G., Sudret, B.: Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics* **230**(6), 2345–2367 (2011)
 45. Tribus, M., M.E.C.: Energy and information. *Scientific American* **224**, 178–184 (1971)
 46. Rezaifar, Y.C.P.R., Krishnaprasad, P.S.: Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In: *Conference on Signals, Systems & Computers* (2002)
 47. Wei, D., Milenkovic, O.: Subspace pursuit for compressive sensing signal reconstruction (2009)
 48. Sargsyan, K., Safta, C., Najm, H.N., Debusschere, B.J., Ricciuto, D., Thornton, P.: Dimensionality reduction for complex models via bayesian compressive sensing. *International Journal for Uncertainty Quantifications* **4**(1), 63–93 (2014)
 49. Oakley, J.E., O'Hagan, A.: Probabilistic sensitivity analysis of complex models: a bayesian approach. *Journal of the Royal Statistical Society* **66**(3), 751–769 (2010)
 50. Witteveen, J., Doostan, A., Pecnik, R., Iaccarino, D.: Uncertainty quantification of the transonic flow around the rae 2822 airfoil. *Annual Research Briefs 2009* (2010)
 51. Kumar, D., Raisee, M., Lacor, C.: An efficient non-intrusive reduced basis model for high dimensional stochastic problems in cfd. *Computers & Fluids* **138**, 67–82 (2016)

Figures

Figure 1 A one-dimensional toy model consisting of a combination of simple functions followed by a high-dimensional function with small coefficients.

Figure 2 Comparison of results for the Ishigami model: the Relative MSE of the classic methods and the two optimization methods based on them. The subgraph is part of the PDF of sampling and several methods.

Figure 3 Comparison of results for the OAKLEY & O'HAGAN function: the Relative MSE of the classic methods and the two optimization methods based on them. The subgraph is the PDF of sampling and the several methods.

Figure 4 Comparison of results for the RAE2822 airfoil: using OMP and the two optimization methods based on it.

Figure 5 Comparison of results for the RAE2822 airfoil: using LARS and the two optimization methods based on it.

Figure 6 Comparison of results for the RAE2822 airfoil: using SP and the two optimization methods based on it.

Figure 7 Comparison of results for the RAE2822 airfoil: using BCS and the two optimization methods based on it.

Tables

Table 1 Comparison of relative MSE and KL divergence for the toy model

Results	OMP	Ent-OMP	HEnt-OMP
Relative MSE	0.323	0.167	0.192
KL Divergence	1175.7	381.3	620.8

Table 2 Comparison of level of sparsity and average calculation time for the Ishigami model ($d=3$, $p=8$)

Level of sparsity	OMP	LARS	SP	BCS	Time
Classic methods	22	18	17	41	0.6915
Ent-PCE	13	13	13	13	0.8821
HEnt-PCE	16	16	26	13	0.6707
Full PCE		165			0.7107

Table 3 Comparison of level of sparsity and average calculation time for the OAKLEY & O'HAGAN function ($d=15$, $p=3$)

Level of sparsity	OMP	LARS	SP	BCS	Time
Classic methods	220	227	240	146	2.8447
Ent-PCE	46	46	46	46	5.7239
HEnt-PCE	89	92	132	2	2.3035
Full PCE		816			2.3801

Table 4 Reference value of Uncertainty Factors in Experiments

Uncertainty Factors	Reference Value
Mach number	0.729
Angle of attack	2.31
Free-stream temperature	288.15
Reynolds number	6.5×10^6
Reynolds length	1.0

Figures

Figure 1

A one-dimensional toy model consisting of a combination of simple functions followed by a high-dimensional function with small coefficients.

Figure 2

Comparison of results for the Ishigami model: the Relative MSE of the classic methods and the two optimization methods based on them. The subgraph is part of the PDF of sampling and several methods.

Figure 3

Comparison of results for the OAKLEY & O'HAGAN function: the Relative MSE of the classic methods and the two optimization methods based on them. The subgraph is the PDF of sampling and the several methods.

Figure 4

Comparison of results for the RAE2822 airfoil: using OMP and the two optimization methods based on it.

Figure 5

Comparison of results for the RAE2822 airfoil: using LARS and the two optimization methods based on it.

Figure 6

Comparison of results for the RAE2822 airfoil: using SP and the two optimization methods based on it.

Figure 7

Comparison of results for the RAE2822 airfoil: using BCS and the two optimization methods based on it.