Elsevier Editorial System(tm) for Applied Mathematical Modelling

Manuscript Draft

Manuscript Number: APM-D-14-02414R2

Title: Incremental modelling for high-order polynomial surrogate models

Article Type: Research Paper

Keywords: Surrogate models; High-order Chebyshev polynomials; Sampling; Incremental modelling.

Corresponding Author: Dr. Zhen Luo, Ph.D.

Corresponding Author's Institution: University of Technology, Sydney

First Author: Jinglai Wu

Order of Authors: Jinglai Wu; Zhen Luo, Ph.D.; Jing Zheng; Chao Jiang, Ph.D.

Abstract: This study will develop a new high-order polynomial surrogate model (HOPSM) to overcome routines of expensive computer simulations in engineering. The proposed HOPSM is expected to keep advantages of the traditional low-order polynomial models in efficiency, transparency and simplicity, while avoid their disadvantage in accuracy. The zeros of Chebyshev polynomials having the highest allowable order will be utilized as the sampling candidates to improve stability and accuracy of the approximation. In the numerical process, a space-filling scheme is used to generate the initial set of samples, and then an incremental method based on the maximin principle is established to select more samples from all candidates. At the same time, the order of HOPSM is sequentially updated by using an order incremental scheme, to adaptively increase the polynomial order along with the increase of the sample size. After the order increment, the polynomial with the largest adjusted R-square is determined as the final HOPSM. Several typical test functions and two engineering applications are used to demonstrate the effectiveness of the proposed surrogate modelling method.

Submitted to Applied Mathematical Modelling

Original submission APM-D-14-02414, December 2014;
Revised submission APM-D-14-02414.R1, August 2015.
Revised submission APM-D-14-02414.R2, November 2015

Incremental modelling of a new high-order polynomial surrogate model

by

Jinglai Wu [1], Zhen Luo[1,*], Jing Zheng [2], Chao Jiang [2]

[1] *School of Electrical, Mechanical and Mechatronic Systems*
The University of Technology, Sydney, NSW 2007, Australia

[2] *School of Mechanical and Vehicle Engineering*
Hunan University, Changsha, Hunan Province, 410082, China

*Correspondence author in manuscript submission
(Dr. Z. Luo, Email: zhen.luo@uts.edu.au, Tel: +61-2-9514-2994; Fax +61-2-9514-2655)

This paper is submitted for possible publication in ***Applied Mathematical Modelling***. It has not been previously published, is not currently submitted for review to any other journals, and will not be submitted elsewhere during the peer review.

Highlights for Review:

1. The proposal of a new high-order polynomial surrogate model, termed as HOPSM;
2. Zeros of Chebyshev polynomials with the highest allowable order used as sampling candidates;
3. An incremental sampling scheme using the *maximin* principle to sequentially collect samples;
4. An order incremental method to adaptively improve the order of the polynomial.

Paper No.: APM-D-14-02414.R2
Authors: Jinglai Wu, Zhen Luo, Jing Zheng, Chao Jiang
Title: Incremental modelling for high-order polynomial surrogate models
Submitted to: Applied Mathematical Modelling

Response to Editor's Comments

Editor's Comment: *The contents of this paper appear to have been presented at THE ELEVENTH WORLD CONGRESS OF STRUCTURAL AND MULTIDISCIPLINARY OPTIMISATION (wcsmo-11) (7-12 june 2015, Sydney Australia). The Abstract submitted to the conference is word for word the same as that of this manuscript. A further revision is advisable to – reword the abstract to some extent, making clear any differences between the conference presentation and this article.*

Reply: Many thanks. We are terribly sorry for the direct use of the Abstract of this manuscript for the WCSMO-2015 conference. We have fixed the problem by following the Editor's comment.

It is noted that we only submitted an Abstract to the WCSMO2015 conference (without full-length paper), and furthermore we haven't used any of the contents of this manuscript for any other conferences. As a matter of fact, this first version of this manuscript was submitted to APM in November 2014, and the WCSMO conference was held in June 2015.

As a PhD supervisor, it is my reasonability and I should have carefully checked the Abstract submitted to the WCSMO2015. I have talked to the PhD student and seriously noted him that such behaviours are inappropriate and cannot be allowed in the future in any circumstances. The student has also shown me his sorry and apology for this matter.

Again, please accept our sincere apologies for this matter.

The new Abstract for this manuscript is now written as follows:

"This study will develop a new high-order polynomial surrogate model (HOPSM) to overcome routines of expensive computer simulations in engineering. The proposed HOPSM is expected to keep advantages of the traditional low-order polynomial models in efficiency, transparency and simplicity, while avoid their disadvantage in accuracy. The zeros of Chebyshev polynomials having the highest allowable order will be utilized as the sampling candidates to improve stability and accuracy of the approximation. In the numerical process, a space-filling scheme is used to generate the initial set of samples, and then an incremental method based on the *maximin* principle is established to select more samples from all candidates. At the same time, the order of HOPSM is sequentially updated by using an order incremental scheme, to adaptively increase the polynomial order along with the increase of the sample size. After the order increment, the polynomial with the largest adjusted R-square is determined as the final HOPSM. Several typical test functions and two engineering applications are used to demonstrate the effectiveness of the proposed surrogate modelling method."

The original Abstract (also the Abstract for WCSMO2015) was given as follows:

"To avoid routine tasks of expensive computer simulations in engineering, the surrogate models have been widely used. This paper will propose a high-order polynomial surrogate model (HOPSM) that includes two novel aspects compared to conventional approximation of functions in high dimension, with a view to retaining advantages of low-order polynomial models in efficiency, transparency and simplicity

while overcoming their disadvantage in accuracy. In constructing HOPSM, firstly, the zeros of Chebyshev polynomials with the highest allowable order will be used as sampling candidates to improve stability and accuracy. An incremental sampling scheme using the maximin principle is developed to collect sampling points from the set of all candidates, with a space-filling scheme generating the initial samples. Secondly, the order of HOPSM is updated through an order incremental method, which will adaptively improve the order of the polynomial sequentially with the increase of the sampling size. The final HOPSM after the order increment will be determined as the polynomial that has the largest adjusted R-square. The HOPSM is compared with the well-known Kriging and RBF surrogate models using both test functions and two engineering applications, to demonstrate accuracy and robustness of the proposed method."

Submitted to Applied Mathematical Modelling

Original submission APM-D-14-02414, December 2014;
Revised submission APM-D-14-02414.R1, August 2015.
Revised submission APM-D-14-02414.R2, November 2015

Incremental modelling of a new high-order polynomial surrogate model

by

Jinglai Wu [1], Zhen Luo[1,*], Jing Zheng [2], Chao Jiang [2]

[1] *School of Electrical, Mechanical and Mechatronic Systems*
The University of Technology, Sydney, NSW 2007, Australia

[2] *School of Mechanical and Vehicle Engineering*
Hunan University, Changsha, Hunan Province, 410082, China

*Correspondence author in manuscript submission
(Dr. Z. Luo, Email: zhen.luo@uts.edu.au, Tel: +61-2-9514-2994; Fax +61-2-9514-2655)

This paper is submitted for possible publication in ***Applied Mathematical Modelling***. It has not been previously published, is not currently submitted for review to any other journals, and will not be submitted elsewhere during the peer review.

Abstract

This study will develop a new high-order polynomial surrogate model (HOPSM) to overcome routines of expensive computer simulations in engineering. The proposed HOPSM is expected to keep advantages of the traditional low-order polynomial models in efficiency, transparency and simplicity, while avoid their disadvantage in accuracy. The zeros of Chebyshev polynomials having the highest allowable order will be utilized as the sampling candidates to improve stability and accuracy of the approximation. In the numerical process, a space-filling scheme is used to generate the initial set of samples, and then an incremental method based on the *maximin* principle is established to select more samples from all candidates. At the same time, the order of HOPSM is sequentially updated by using an order incremental scheme, to adaptively increase the polynomial order along with the increase of the sample size. After the order increment, the polynomial with the largest adjusted R-square is determined as the final HOPSM. Several typical test functions and two engineering applications are used to demonstrate the effectiveness of the proposed surrogate modelling method.

Keywords: Surrogate models; High-order Chebyshev polynomials; Sampling; Incremental modelling.

1. Introduction

Numerical simulations using accurate models for many real world problems in engineering often become unaffordable, as they involve routine evaluations of a large number of cost-prohibitive computations. A surrogate model is an engineering method for approximation of practical design problems, to avoid computation extensive simulations. For example, the finite element analysis of vehicle crashworthiness will usually take tens of hours to run one simulation, while the whole crashworthiness will take hundreds and even thousands of iterations to complete. Furthermore, the simulation model is in general a black box, with little or no additional information available for its inner mechanism except for the output it generates [1]. It is hard to explore, optimize or gain insight into the system. Hence, the surrogate models have been widely used as inexpensive approximation for computationally expensive models [2]. The surrogate model, also termed as meta-model, response surface or emulator, refers to any relatively simple relationship between parameters and response often based on limited data [3]. There are two main steps involved in the construction of a surrogate model: (1) the sampling or design of experiment (DOE) and (2) metamodeling via interpolation or regression algorithms after the sampling.

There have been several types of surrogate models, including the traditional response surface (low order polynomials) [4], radial basic function (RBF) [5], Kriging [6, 7], multivariate adaptive regression splines (MARS) [8], support vector regression (SVR) [9, 10], high dimensional model representation (HDMR) [11, 12], or the combination of these surrogate models [13-15]. For example, Jin et al. [16] studied several surrogate models based on the multiple performance criteria, including accuracy, robustness, efficiency, transparency, and conceptual simplicity. Their results showed that the polynomial surrogate models may have advantages in efficiency, transparency, and conceptual simplicity over other models. They also noted that the performance of a surrogate model is influenced by sampling. Simpson et al. [17] showed that the approximations of Kriging and RBF models for high-order nonlinear problems are more accurate, while quadratic polynomials are better for low-order nonlinear functions. At the same time, Rijpkema et al. [18] shown that the Kriging model was less stable than the polynomials regression model in some cases. Hence, there is no one surrogate model suitable for all problems over different sampling schemes and different sample sizes. Further studies about the comparison of different kinds of surrogate models can be found in references [19-22].

As aforementioned, the low-order polynomials have been applied to many engineering problems due to their advantages, especially efficiency and transparency. However, due to low accuracy in fitting high-

order nonlinear functions, the low-order polynomials have difficulty for problems with high nonlinearity. To keep the merits of low-order polynomials while overcome their weakness for building surrogate models, the high-order polynomials [23] can be applied to establish surrogate models, such as, the Bernstain polynomials [24, 25], Chebyshev polynomials [26, 27] and Gegenbauer functions [28]. However, how to build surrogate models using high-order polynomials is seldom studied in engineering, mainly because of two reasons: the first is the numerical instabilities, e.g. the Runge phenomena, and the second is the large number of samples for estimation of the unknown coefficients, particularly for high-dimensional problems [29]. In fact, the first can be avoided by selecting new samples to improve stability and accuracy, e.g. the zeros of the first kind Chebyshev polynomials [30-34]. The second can be improved by using a suitable expression of polynomials (e.g. the simplex), which may reduce the number of high-order coefficients to be estimated in the model. It is noted that the required number of samples is more influenced by the extent of complexity and dimension of the function rather than the type of surrogate models. Thus, the required sampling size will increase with the increase of complexity and dimension, no matter which type of surrogate models is used.

The approximation accuracy of a surrogate model is not only determined by the type of surrogate models but also by sampling information. The accuracy of a surrogate model will be improved when more data points are sampled. However, it is impossible to choose too many sample points due to the computational cost. How to evaluate the unknown information only in terms of a limited number of sampling points, to maintain a well trade-off between computational cost and accuracy is an important issue for sampling (or DOE). Traditionally, the DOE can be categorised as Factor Design (FD) [35], Central Composite Design (CCD) [35], Pseudo-Monte Carlo Sampling [36] (PMCS, e.g. the Latin Hypercube sampling [37], and Orthogonal Sampling [38]) and Quasi-Monte Carlo Sampling (QMCS) [3, 36, 39]. The FD and CCD, belonging to the classical DOE [3], are usually employed for laboratory experiments where the random errors are assumed to exist, while the modern DOE (PMCS and QMCS) are used in deterministic computer simulations without random errors [36].

The above sampling methods can be classified as “one-shot” sampling schemes, as the samples are chosen once and fixed in the fitting process [1]. These methods can be easily implemented and provide a good coverage of the design space without incorporating any prior knowledge of the system. However, the “one-shot” DOE may suffer from its inflexibility to learn the special characteristics of the shape of the response surface [40], and the number of sampling points is easily over or under estimated.

To improve flexibility and efficiency of sampling, the sequential sampling strategy (e.g. the adaptive sampling [41], and incremental sampling [42]) has been developed. Sequential sampling analyses the data from samples and surrogate models in order to select new samples from the regions that are difficult to approximate, resulting in a more efficient distribution of samples in the entire design space compared to the traditional one-shot sampling scheme. That is, in the sequential sampling, a surrogate model is first built using an initial set of samples and then sequentially updated by adding new sample points. There are two schemes used in the sequential sampling: the first is the global exploration, which scatters samples in regions containing no sampling points, and the second is the local exploitation [43] which adds more samples to regions identified to be interesting. The exploration selects sampling points to fill the entire design space, which is mainly used to build global surrogate models, while the exploitation is mainly used in the surrogate model-based optimization. Furthermore, some studies combined exploration with exploitation to build global surrogate models, e.g. [1, 40, 43].

The exploration aims to place samples in the entire design space uniformly, which is the same as some modern DOEs. However, most QMCS and PMCS are non-incremental sampling methods, as augmenting the number of samples implies a completely different sampling of the parameter space with all new point-locations [44], which will be too expensive to be used. Romero et al. [42] used the Progressive Lattice Sampling (PLS) incremental sampling designs to construct the progressive response surface. However, the PLS allows only a quantized increment M of samples to be added to an existing PLS level (point set) to achieve to a new level. This quantized incremental cost M accelerates quickly with the increase of the PLS level and dimension of the parameter space. To make the sequential sampling more flexible, Romero et al. [44] suggested to use the Halton points to build the progressive response surface, because Halton sampling does not suffer from the cost-scaling problems that the PLS does. Halton [45] is a lower discrepancy (degree of the nonuniformity) sequence method and has a hierarchical structure. The Halton points will be used to compare with the sampling scheme proposed in this paper.

This study will focus on the proposal of a new global surrogate model using high-order polynomials, which can remain the merits of traditional low-order polynomial models while improve approximation accuracy. The sampling points will be sequentially and incrementally selected from a candidate set which is comprised of the zeros of first kind Chebyshev polynomials, to make the surrogate model more stable. In building the surrogate model, only some of the candidate samples are chosen as the required sampling points by using a sequential sampling scheme based on the *maximin* principle [46]. Since the initial samples, to be distributed uniformly, have obvious influence on the sequential sampling, a new efficient initialization algorithm will be proposed to select the initial samples from the candidate set. Furthermore,

the simplex format of polynomials is used to express HOPSM, to reduce the number of coefficients of the polynomials to be estimated. To achieve the best approximation accuracy, an order incremental strategy will also be used to update the order of HOPSM. Several mathematical test examples and two engineering applications are used to show the accuracy and robustness of the proposed method, in comparison with the two well-known Kriging and RBF surrogate models.

2. Polynomial approximation using Chebyshev series

2.1 The approximation theory of Chebyshev polynomials

In this paper, it is assumed that the simulation model is continuous, and can be approximated by using polynomials. To simplify the problem, we consider a continuous function $f(x)$ on $x\in[-1,1]$ (any interval $[a, b]$ can be normalized to [-1,1] via a linear transformation). $f(x)$ can be approximated by the truncated Chebyshev series $p_n(x)$ as follows:

$$f(x) \approx p_n(x) = \frac{1}{2} f_0 + \sum_{i=1}^{n} f_i C_i(x) \tag{1}$$

where f_i are the Chebyshev coefficients that are constant, and $C_i(x)$ denotes the Chebyshev polynomials

$$C_i(x) = \cos i\theta,\ \ \theta = \arccos(x) \in [0, \pi] \tag{2}$$

The coefficients can be obtained via the following equation as

$$f_i = \frac{2}{\pi}\int_{-1}^{1} \frac{f(x)C_i(x)}{\sqrt{1-x^2}} dx \approx \frac{2}{\pi}\frac{\pi}{m}\sum_{j=1}^{m} f(x^{(j)})C_i(x^{(j)}) = \frac{2}{m}\sum_{j=1}^{m} f(\cos\theta^{(j)})\cos i\theta^{(j)} \tag{3}$$

where m is the order of the integral formula, $x^{(j)}$ or $\theta^{(j)}$, denoting the interpolation points of the integral formula in the **x** space or $\boldsymbol{\theta}$ space, are the zeros of the Chebyshev polynomials of degree m, given by

$$x^{(j)} = \cos\theta^{(j)},\ \text{where}\ \theta^{(j)} = \frac{2j-1}{m}\frac{\pi}{2},\ j = 1, 2, ..., m \tag{4}$$

To guarantee the accuracy of the Gauss-Chebyshev quadrature in computing the highest order coefficient f_n, the order m in Eq. (3) should be larger than n, and it is usually set as n+1 to save the computational cost. More details about the Chebyshev polynomial can be found in References [32-34].

The coefficients can also be computed by using the least square method. To unify the symbol of coefficients, let $\beta_0=f_0/2$ and $\beta_i=f_i$ for $i>0$, then truncated Chebyshev series can be expressed by

$$f(x) \approx \frac{1}{2} f_0 + \sum_{i=1}^{n} f_i C_i(x) = \sum_{i=0}^{n} \beta_i C_i(x) \tag{5}$$

where the coefficients β_i can also be calculated by using the least square method directly as

$$\boldsymbol{\beta} = \begin{bmatrix}\beta_0 & \beta_1 & \cdots & \beta_n\end{bmatrix}^{\mathrm{T}} = \left(\mathbf{C}^{\mathrm{T}}\mathbf{C}\right)^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{f} \tag{6}$$

where **C** is the transform matrix with size $m\times(n+1)$ at the interpolation points, and **f** is the column vector including function values over all the interpolation points $\mathbf{x}=[x^{(1)} \quad \cdots \quad x^{(m)}]^{\mathrm{T}}$:

$$\mathbf{C}=\left[C_0(\mathbf{x}) \quad C_1(\mathbf{x}) \quad \cdots \quad C_n(\mathbf{x})\right], \text{ and } \mathbf{f}=f(\mathbf{x}) \tag{7}$$

Consider Eqs. (5)-(7), it can be found that the procedure of constructing the approximation polynomial is just the same process of constructing a response surface model, in which the Chebyshev polynomials are used as the basis. Therefore, when the sampling points are selected as the zeros (**x** vector) of the Chebyshev polynomials $C_i(x)$, we can get the approximated Chebyshev series $p_n(x)$, which is a polynomial with order n. As aforementioned, it is termed as high-order polynomial surrogate model (HOPSM). The Eq. (5) is the expression of 1-dimensional HOPSM, the multi-dimensional HOPSM will be developed in the following subsection.

2.2 The expression of high-order polynomial surrogate model

Extending the 1-dimensional space to k-dimensional space, the input variables are expressed as $\mathbf{x}\in[-1,1]^k$. The HOPSM of the k-dimensional continuous function $f(\mathbf{x})$ with order n can be expressed as the tensor product of the HOPSM, with respect to each dimensional variable x_i [34]

$$\hat{f}_n(\mathbf{x})=\sum_{0\le i_1,\ldots,i_k\le n}\beta_{i_1\ldots i_k}C_{i_1\ldots i_k}(\mathbf{x}), \text{ and } i_1,\ldots,i_k=0,1,\ldots,n \tag{8}$$

where $\beta_{i_1\ldots i_k}$ denote the coefficients, $C_{i_1\ldots i_k}(\mathbf{x})=\cos(i_1\theta_1)\ldots\cos(i_k\theta_k)$ represent the k-dimensional Chebyshev polynomials, and $\boldsymbol{\theta}=\arccos(\mathbf{x})\in[0,\pi]^k$. The subscript of coefficients $\beta_{i_1\ldots i_k}$ forms a "hypercube" (from 0 to n), so Eq. (8) is called the hypercube format of HOPSM. It should be noted that the number of the coefficients need to be determined in Eq. (8) is $(n+1)^k$, which will be extremely large when k and n are large, so it may not be suitable for high dimensional problems.

Besides the hypercube format, the HOPSM can also be expressed as a "simplex" format, which only contains the terms when the order without exceeding n

$$\hat{f}_n(\mathbf{x})=\sum_{0\le i_1+\ldots+i_k\le n}\beta_{i_1\ldots i_k}C_{i_1\ldots i_k}(\mathbf{x}), \text{ and } i_1,\ldots,i_k=0,1,\ldots,n \tag{9}$$

It should be noted that Eq. (9) is different from Eq. (8), where the subscript in the former forms a simplex while the latter forms a hypercube. As a result, the number of coefficients to be estimated in Eq. (9) is much less than that in Eq. (8) for high dimensional problems, which is given by

$$N_C(n,k)=\frac{(k+n)!}{k!n!} \tag{10}$$

To reduce the number of coefficients to be estimated, this paper uses the simplex format for HOPSM. All the coefficients can be calculated by the least square method (LSM) after the two problems are solved: the selection of sampling points and the order determination of HOPSM. Both will be addressed in Section 3.

2.3 The evaluation index of surrogate models

To evaluate the accuracy denoting the fitting goodness of the surrogate model, some evaluation indexes should be introduced. The root-mean-square error (RMSE) and average absolute error (AAE) are widely employed to describe global accuracy, while the maximum absolute error (MAE) is usually used to indicate the local accuracy. In order to avoid ambiguity or to enable comparisons of surrogate models across disciplines, then relative error averages are sometimes used in the literature, such as AAE relative to the standard deviation, R-square [16], or AAE and RMSE measures relative to the average response f [47]. This paper will use R-square, AAE, and MAE to describe the accuracy, but some normalized operation will be used to make these indexes more clearly.

The R-square, denoted as R^2, is a statistical characteristic, which is expressed as follows:

$$R^2 = 1 - \frac{SSE}{SST} \tag{11}$$

where SSE and SST denote the residual sum of square and the total sum of square, respectively

$$SSE = \sum_{i=1}^{N} \left(\hat{f}_n(\mathbf{x}_i) - f(\mathbf{x}_i) \right)^2 \text{ and } SST = \sum_{i=1}^{N} \left(f(\mathbf{x}_i) - \bar{f} \right)^2 \tag{12}$$

where $\bar{f}$ is the mean value of observed data (real value at the test points), $\mathbf{x}_i$ is the test point, and N is the size of test points. The R^2 can also be used to pre-estimate the accuracy by using the sampling points as the test points in the construction of HOPSM. Generally, the larger value of R^2 means higher fitness of the regression model. However, if the sampling points are used as the test points, the R^2 will improve when the number of estimated coefficients increases, so we need to adjust the R-square to pre-estimate the fitting degree of regression model. The adjusted R-square $\bar{R}^2$ has been widely used, defined as

$$\bar{R}^2 = 1 - \frac{SSE/(s_0 - N_c(n,k) - 1)}{SST/(s_0 - 1)} \tag{13}$$

where s_0 denotes the number of samples, $N_c(n,k)$ is the number of estimated coefficients given by Eq. (10). In this case, even when the number of coefficients increases, the adjusted R-square $\bar{R}^2$ may still decrease. $\bar{R}^2$ will be used to pre-estimate the fitness of HOPSM in the model constructing period, while R^2 will be used to indicate its actual fitness in the validation period.

The definition of AAE and MAE are given as

$$\text{AAE} = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{f}_n(\mathbf{x}_i) - f(\mathbf{x}_i)\right| \text{ and } \text{MAE} = \max_i\left|\hat{f}_n(\mathbf{x}_i) - f(\mathbf{x}_i)\right| \quad (14)$$

In solving different problems, the AAE and MAE may change in large range. To make the two indexes more clearly, we transform them into the normalized average absolute error and the normalized maximum absolute error, respectively, noted as NAAE and NMAE.

$$\text{NAAE} = \text{AAE}\Big/\left(\frac{1}{N}\sum_{i=1}^{N}\left|f(\mathbf{x}_i)\right|\right) \text{ and } \text{NMAE} = \text{MAE}\Big/\left(\max_i f(\mathbf{x}_i) - \min_i f(\mathbf{x}_i)\right) \quad (15)$$

The smaller values of NAAE and NMAE are, the better the surrogate model is. Both NAAE and NMAE will be used in constructing the HOPSM, to pre-estimate the fitness by using the sampling points as the test points. More detailed content will be given in next section.

3. Incremental modelling of HOPSM

3.1 The sampling candidate set

As described in Section 2.1, the sampling points of one-dimensional HOPSM are the zeros of Chebyshev polynomials, and the order of Chebyshev polynomials should be higher than the order of HOPSM. For the multi-dimensional case, the sampling points can be produced by the tensor product of the zeros of Chebyshev polynomials. However, the operation of the tensor product will produce too many sampling points to be used directly. Therefore, we only use the tensor product operation to produce a sampling candidate set, and then the sampling points will be chosen from the set.

The candidate set can be expressed in $\mathbf{x}$ space and $\boldsymbol{\theta}$ space, respectively

$$\boldsymbol{X} = \mathbf{x}_1 \otimes \ldots \otimes \mathbf{x}_k, \quad \boldsymbol{\theta} = \boldsymbol{\theta}_1 \otimes \ldots \otimes \boldsymbol{\theta}_k \quad (16)$$

where $\mathbf{x}_i$ and $\boldsymbol{\theta}_i$ denote the candidates (the zeros of the Chebyshev polynomial) of the ith variable in $\mathbf{x}$ space and $\boldsymbol{\theta}$ space, respectively, and $\otimes$ denotes the tensor product operation. Based on Eq. (4), the candidates in dimension i are expressed by

$$\mathbf{x}_i = \cos\boldsymbol{\theta}_i, \boldsymbol{\theta}_i = \left\{\theta^{(j)} = \frac{2j-1}{m}\frac{\pi}{2}, j = 1, 2, \ldots, m\right\} \quad (17)$$

How to choose m which is the order of sampling candidate set is important. A very small m may reduce the accuracy of the surrogate model, and a very large m will usually waste the sampling information.

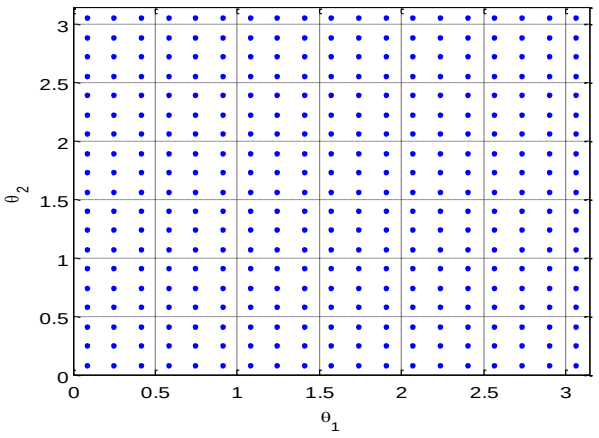


Figure.1 (a) Candidates for k=2 in $\boldsymbol{\theta}$ space

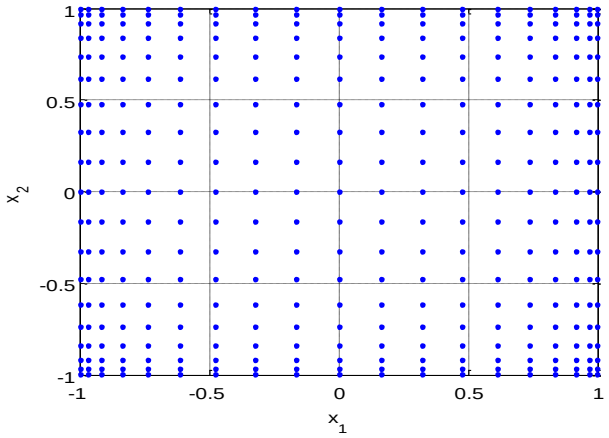


Figure.1 (b) Candidates for k=2 in $\mathbf{x}$ space

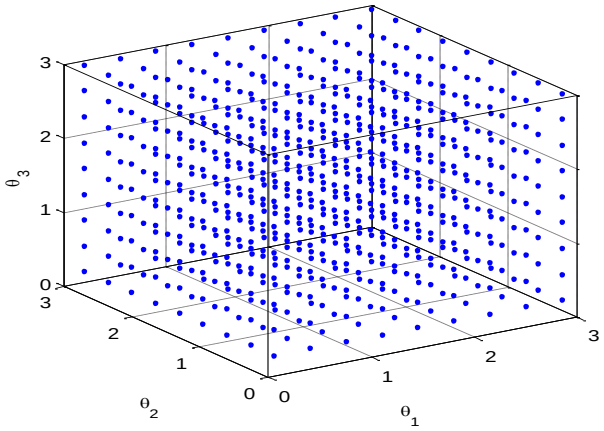


Figure.2 (a) Candidates for k=3 in $\boldsymbol{\theta}$ space

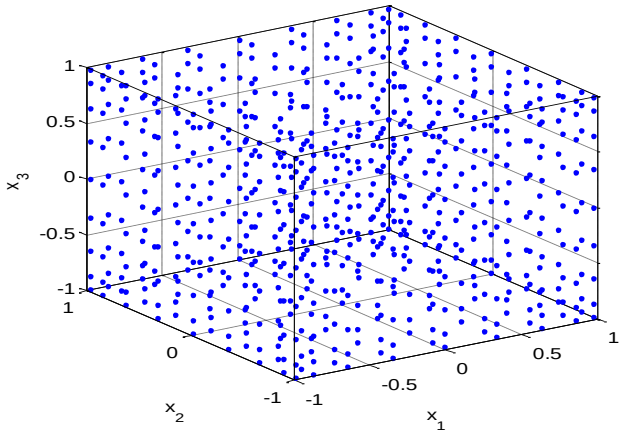


Figure.2 (b) Candidates for k=3 in $\mathbf{x}$ space

We have two approaches to determine the order of the candidate set. The first is to use the lowest allowable order to produce the candidate set, i.e. m=n+1, where n is the order of HOPSM. m will be gradually increased with the increase of n (n will increase sequentially in the process of constructing HOPSM). The drawback of this method is that there will be more redundant sampling information with the increase of order m, so the sampling efficiency is low. The second is to determine m based on the highest allowable order of HOPSM, to keep the candidate set unchanged. All the sampling points will be chosen from the unique candidate set for any n, which makes the sampling more efficiently. The highest allowable order of HOPSM ($n_{\max}$) can be determined by the most allowable sampling size $N_{\max}$. The number of coefficients must be smaller than the sampling size, so $n_{\max}$ will satisfy

$$N_C\left(n_{\max},k\right)=\frac{\left(k+n_{\max}\right)!}{k!n_{\max}!}<N_{\max} \tag{18}$$

The order m of the candidate set should be larger than $n_{\max}$, so it will be set to m=$n_{\max}$+1 to minimize the size of the candidate set. We should note that the number of sampling candidates N_S shown in Eq. (19) is much larger than the number of the coefficients N_C, especially for a larger dimension k.

$$N_S(m,k) = (n_{\max} + 1)^k \tag{19}$$

For example, if the $N_{\max}$=200, and k=2 or 3, then $n_{\max}$=18 or 8, and m=19 or 9, respectively. The plot of candidate sets in **x** space and **θ** spaces for the two cases are shown in Figs. 1 and 2. It is noted that the candidate set is a symmetrical grid in **θ** space but an unsymmetrical grid in **x** space. The samples distribute denser in the regions closing the boundaries of the design space, which is different from the traditional equidistant full factor design.

3.2 The initial sampling algorithm

The initial samples have large influence on the performance of the final sampling. Generally, the initial samples are expected to distribute uniformly in the entire space. One of the most widely used measures to evaluate the uniformity of a sampling set is the *maximin* metric introduced by Johnson et al [46]. The following scalar-valued criterion function [48] is mainly used to rank competing sampling set as

$$\Phi_q(\boldsymbol{\Theta}) = \left(\sum_{i=1}^{s_0} \sum_{j=i+1}^{s_0} d\left(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}^{(j)}\right)^{-q} \right)^{1/q} \tag{20}$$

where the q is a large positive integer and set as 100 in this paper, **Θ** is the sampling set, s_0 denotes the number of samples, and the distance $d\left(\boldsymbol{\theta}^{(j_1)}, \boldsymbol{\theta}^{(j_2)}\right)$ is measured by the *p-norm* in the **θ** space:

$$d\left(\boldsymbol{\theta}^{(j_1)}, \boldsymbol{\theta}^{(j_2)}\right) = \left(\sum_{i=1}^{k} \left| \theta_i^{(j_1)} - \theta_i^{(j_2)} \right|^p \right)^{1/p} \tag{21}$$

Here p is set as 2, which yields the Euclidean norm.

In the above equation, a smaller Φ_q indicates more uniformity of the sampling set. However, minimizing Φ_q is an NP-complete problem. In this study, a new sequential algorithm will be developed to produce a uniform distribution of the initial samples. To seek the uniformity, the sampling points should be located in all levels in each dimension, and they should be equivalently located in each level, although it is difficult to achieve.

The number of initial samples is set to N_0=mk. To simplify the notation, we use the levels (from 1 to m) to denote the location of samples. The samples can be expressed by a matrix with a size of k×mk, where the column denotes the sequence of variables and row denotes the sequence of sampling points. The first m elements of the first row are set as 1, 2 … m sequentially, and the level of the first column for the ith variables is set as i. If the order m is smaller than i, the level of the ith variable will be set as the remainder of i/m. The first row is called as the reference row, marked by "*". For example, considering

the case m=5 and k=6, the matrix is given in Table 1, and the levels are denoted by the bold numbers, while the reference row includes the numbers marked with blue colour.

Table 1 has given the first column and the first 5 elements of the first row, so we need to determine the remaining elements to uniformly distribute the samples. Choose new sampling points in sequentially, through minimizing the *maximin* metric Φ_q. Denote the sampled set as $\boldsymbol{\Theta}$ and the remaining candidate set as $\boldsymbol{\Psi}$. When a new point from $\boldsymbol{\Psi}$ is added into the sampling set $\boldsymbol{\Theta}$, this point should minimize Φ_q of the new sampled set comprised by $\boldsymbol{\Theta}$ and the new sampling point $\boldsymbol{\theta}_1^{(j)}$, expressed as

$$\Phi_q\left(\boldsymbol{\Theta},\boldsymbol{\theta}_1^{(j)}\right)=\left(\sum_{i_1=1}^{s_0}\sum_{i_2=i_1+1}^{s_0} d\left(\boldsymbol{\theta}_0^{(i_1)},\boldsymbol{\theta}_0^{(i_2)}\right)^{-q}+\sum_{i_1=1}^{s_0} d\left(\boldsymbol{\theta}_0^{(i_1)},\boldsymbol{\theta}_1^{(j)}\right)^{-q}\right)^{1/q}=\left(\Phi_q\left(\boldsymbol{\Theta}\right)^q+\sum_{i_1=1}^{s_0} d\left(\boldsymbol{\theta}_0^{(i_1)},\boldsymbol{\theta}_1^{(j)}\right)^{-q}\right)^{1/q} \quad (22)$$

where $\boldsymbol{\theta}_0^{(i)}\in\boldsymbol{\Theta}$ are the sampled points, $\boldsymbol{\theta}_1^{(j)}\in\boldsymbol{\Psi}$ is the new sampling point, and s_0 is the number of points in the sampled set. Since the first term in the right side keeps unchanged, the minimizing operation will be applied to the second term, denoted as

$$\phi_q\left(\boldsymbol{\theta}_1^{(j)},s_0\right)=\left(\sum_{i=1}^{s_0} d\left(\boldsymbol{\theta}_0^{(i)},\boldsymbol{\theta}_1^{(j)}\right)^{-q}\right)^{1/q} \quad (23)$$

Therefore, the new sampling points can be chosen by minimizing ϕ_q. If all the candidates are directly used to calculate ϕ_q, the computational cost will still be expensive as O(m^k). Producing the level of each dimensional variable sequentially, the number of calculation will be reduced to O($m\times k$).

To describe the procedure more clearly, we still use the data shown in Table 1. The first step is to calculate the second element of the second column (the level of the 2nd sampling point in the 2nd variable). In this case, the distance defined in Eq. (21) is calculated in a 2-dimensional space constructed by the first and second design variable. After the evaluation, when the level of the second variable is 5, the ϕ_q get the minimum value. Go to the third row of the same column, and determine the level of the third variable by minimizing ϕ_q. However, the distance in Eq. (23) is calculated in a 3-dimensional space of the first three design variables. Repeat this procedure until the level of the last variable is determined.

After the second column is determined, the sampled set $\boldsymbol{\Theta}$ should be updated, and then we can calculate the elements in the third column based on the same approach. Repeat the process until the mth column is determined, and the results of the first 5 sampling points are shown in Table 2.

Table 1. The initialization design matrix for k=6, m=5

No. of variable \ No. of samples	1	2	3	4	5	6	…	30
1*	**1**	**2**	**3**	**4**	**5**			
2	**2**							
3	**3**							
4	**4**							
5	**5**							
6	**1**							

Table 2. The design matrix for k=6 and m=5

No. of variable \ No. of samples	1	2	3	4	5	6	…	30
1*	**1**	**2**	**3**	**4**	**5**			
2	**2**	**5**	**1**	**4**	**2**			
3	**3**	**1**	**5**	**4**	**1**			
4	**4**	**1**	**1**	**5**	**2**			
5	**5**	**1**	**1**	**1**	**4**			
6	**1**	**5**	**3**	**1**	**4**			

More columns need to be determined. It can be implemented by using the same approach. However, we can find that the levels of Table 2 are not distributed uniformly except for the first variable (reference row) which is pre-fixed uniformly. The levels of other design variables should also be distributed uniformly, so we can pre-fix the levels of other design variables by moving the first row to the last, and other rows forward in sequence. After that, the reference row will be the second variable. Pre-fix the elements from the column 6 to 10 of the reference row as 1 to 5, shown in Table 3, and then determine the levels of other rows of columns 6 to 10. After per m new samples are determined, there will be rows moving operation. The operation of moving rows will guarantee every level occurring in each variable at least once.

Table 3. The design matrix for k=6 and m=5

No. of variable \ No. of samples	1	2	3	4	5	6	7	8	9	10	…	30
2*	**2**	**5**	**1**	**4**	**2**	**1**	**2**	**3**	**4**	**5**		
3	**3**	**1**	**5**	**4**	**1**							
4	**4**	**1**	**1**	**5**	**2**							
5	**5**	**1**	**1**	**1**	**4**							
6	**1**	**5**	**3**	**1**	**4**							
1	**1**	**2**	**3**	**4**	**5**							

The plot of the initial sampling points is shown in Fig. 3, expressed through the level. The number of each level located in each dimensional variable, termed as occurrence number, is summarized in Table 4. It can be found that the occurrence number for each level of each dimensional variable is close to 6 (changing

from 4 to 9), with the samples uniformly distributed. Fig. 4 shows the plot of producing two dimensional sampling points with m=9. The number in the figure denotes the sequence of sampling. The points denoted by "□" are the samples determined by using the first variable as the reference row, while "❍" denotes the sampling points selected by considering the second variable as reference row.

Table 4. The occurrence number of each dimensional variable

No. of variable \ No. of levels	1	2	3	4	5
1	**8**	**5**	**6**	**5**	**6**
2	**8**	**5**	**5**	**5**	**7**
3	**9**	**4**	**5**	**4**	**8**
4	**9**	**4**	**7**	**6**	**4**
5	**8**	**5**	**7**	**4**	**6**
6	**7**	**5**	**6**	**6**	**6**

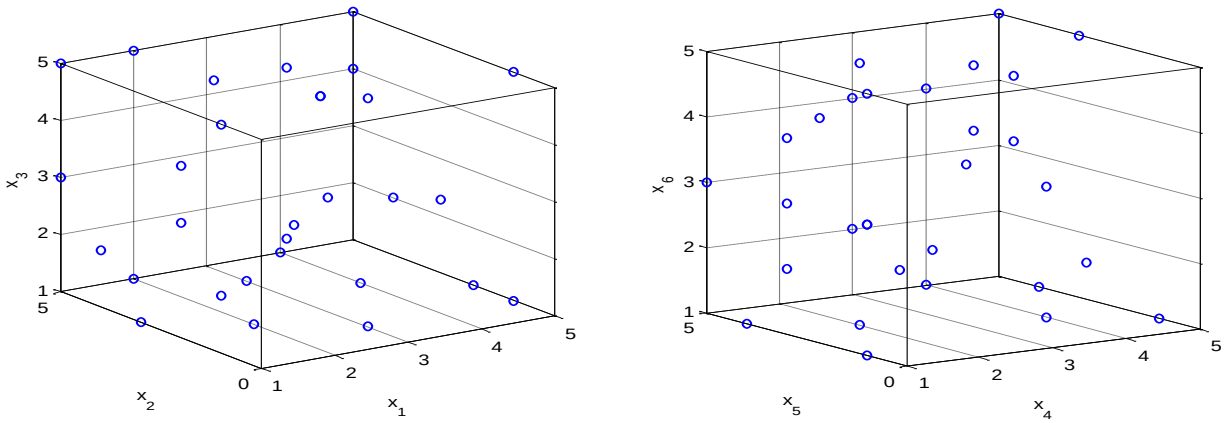


Figure.3 The distribution of initial sampling points

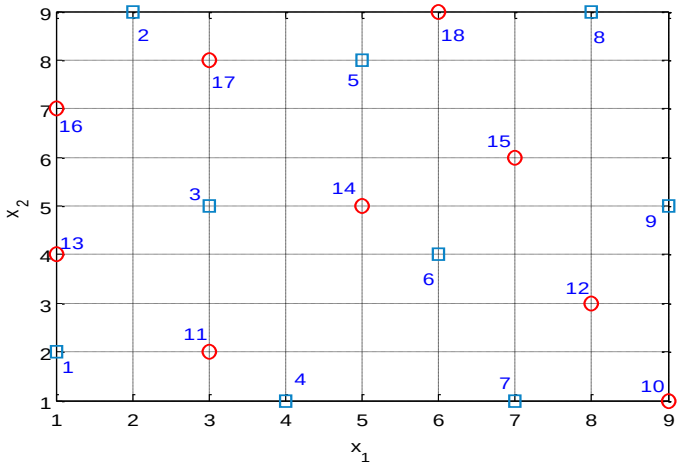


Figure.4 The sequence of initial sampling

3.3 The sequential sampling scheme

The exploration strategy, to fill up the domain as evenly as possible, will be used to choose more sampling points. The new samples can still be chosen by minimizing ϕ_q in Eq. (23), to locate the new sample to the region without containing samples. This can be described by the following optimization

$$\begin{aligned} &\text{Find:} \quad \boldsymbol{\theta}_1^{(j)}, \quad j = 1, ..., s_1 \\ &\min: \quad \phi_q\left(\boldsymbol{\theta}_1^{(j)}, s_0\right) \end{aligned} \tag{24}$$

where s_1 denotes the number of points in the remaining candidate set $\boldsymbol{\Psi}$.

The sampling point can be chosen by performing the above optimization one by one, until the termination condition is satisfied. However, in each exploration, it should be noted that the optimization model requires to compute the distance d in $O(s_0 \times s_1)$ time. This may require expensive computational cost, especially when k and m are relatively large, since s_1 is close to m^k. However, we do not need to compute the *maximin* criterion for all candidates except in the first exploration, as ϕ_q always keeps increasing after each exploration, which can be used to reduce the computational cost.

For the first exploration, the *maximin* criterion ϕ_q of each candidate point will be calculated firstly. Then we can sort the candidates in ascending order of ϕ_q, such that $\phi_q(\boldsymbol{\theta}_1^{(1)}, s_0) \le \cdots \le \phi_q(\boldsymbol{\theta}_1^{(s_1)}, s_0)$. The candidate that has the minimum $\phi_q(\boldsymbol{\theta}_1^{(j)}, s_0)$ will be chosen as the new sampling point, and so the first candidate $\boldsymbol{\theta}_1^{(1)}$ will be the new sampling point. After the new data point is picked, we can update the amount of samples to $s_0 + 1$. In this way, a new sample ($\boldsymbol{\theta}_0^{(s_0+1)} = \boldsymbol{\theta}_1^{(1)}$) will be added into the sampled set $\boldsymbol{\Theta}$, while the candidate $\boldsymbol{\theta}_1^{(1)}$ is also deleted from the candidate set $\boldsymbol{\Psi}$. Subtracting 1 for the sequential number of each candidate, the remaining candidates still keep the ascending sequence as $\phi_q(\boldsymbol{\theta}_1^{(1)}, s_0) \le \cdots \le \phi_q(\boldsymbol{\theta}_1^{(s_1-1)}, s_0)$.

For the second exploration, based on the Eq. (23), ϕ_q can be calculated as follows:

$$\phi_q\left(\boldsymbol{\theta}_1^{(j)}, s_0 + 1\right) = \left(\phi_q\left(\boldsymbol{\theta}_1^{(j)}, s_0\right)^q + d\left(\boldsymbol{\theta}_0^{(s_0+1)}, \boldsymbol{\theta}_1^{(j)}\right)^{-q}\right)^{1/q} \tag{25}$$

Eq. (25) will save computational cost related to Eq. (23), as we have known $\phi_q\left(\boldsymbol{\theta}_1^{(j)}, s_0\right)$. Calculating $\phi_q\left(\boldsymbol{\theta}_1^{(j)}, s_0 + 1\right)$ in ascending order until a candidate is found to make $\phi_q\left(\boldsymbol{\theta}_1^{(J)}, s_0 + 1\right) \le \phi_q\left(\boldsymbol{\theta}_1^{(J+1)}, s_0\right)$, which means the J-th candidate is better than the candidates with larger sequential number than J, as the *maximin* criterion always increases after each exploration ($\phi_q\left(\boldsymbol{\theta}_1^{(j)}, s_0 + 1\right) > \phi_q(\boldsymbol{\theta}_1^{(j)}, s_0)$). The J is termed as criterion sequential number. This operation would also save more computational cost, because most of the candidates which are close to the sampled points can never be selected in the exploration. The ϕ_q of the first J candidates have been updated, denoted as $\phi_q\left(\boldsymbol{\theta}_1^{(j)}, s_0 + 1\right)$ for $j \le J$, while other candidates have

not updated yet, expressed by $\phi_q(\boldsymbol{\theta}_1^{(j)}, s_0)$ for $j > J$. To distinguish them, we introduce another notation M_j, which denotes the number of points in the sampled set used to calculate the ϕ_q of $\boldsymbol{\theta}_1^{(j)}$.

Therefore, ϕ_q of $\boldsymbol{\theta}_1^{(j)}$ is re-noted as $\phi_q(\boldsymbol{\theta}_1^{(j)}, M_j)$. After the second exploration, $M_j = s_0 + 1 \left(j \le J \right)$, $M_j = s_0 \left(j > J \right)$. Sorting the remaining candidates in ascending order once again, the first candidate $\boldsymbol{\theta}_1^{(1)}$ will be added into the sampled set $\boldsymbol{\Theta}$ and deleted from the candidate set $\boldsymbol{\Psi}$. After the second exploration, in a more general case, the ϕ_q can be calculated by

$$\phi_q\left(\boldsymbol{\theta}_1^{(j)}, s_0'\right) = \left(\left(\phi_q\left(\boldsymbol{\theta}_1^{(j)}\right), M_j\right)^q + \sum_{i=1}^{s_0' - M_j} d\left(\boldsymbol{\theta}_0^{\left(s_0' - i + 1\right)}, \boldsymbol{\theta}_1^{(j)}\right)^{-q} \right)^{1/q} \tag{26}$$

where s_0' is the size of the sampled set before the exploration.

Repeat the same process for the second exploration, another new sample can be selected. After the evaluation of ϕ_q , the M_j should be updated, i.e. $M_j = s_0' \left(j < J \right)$. The procedure of the proposed exploration can be described by the following **Algorithm 1**. When the number of sampling points exceeds the required samples $N_c(n,k)$, the least square method can be used to construct the HOPSM.

Algorithm 1

Input: $k, q, \boldsymbol{\Theta}, \boldsymbol{\Psi}, s_0, s_1$

Initialize: $s_0' = s_0, M_j = 0,\ \phi_q\left(\boldsymbol{\theta}_1^{(j)}, M_j\right) = 0,\ \left(j = 1, ..., s_1\right)$

Compute: Do

for j=1:s_1

$\phi_q\left(\boldsymbol{\theta}_1^{(j)}, s_0'\right); M_j = s_0';$

if $\phi_q\left(\boldsymbol{\theta}_1^{(j)}, M_j\right) \le \phi_q\left(\boldsymbol{\theta}_1^{(j+1)}, M_{j+1}\right)$

Break

end if

end for

Sort: $\phi_q\left(\boldsymbol{\theta}_1^{(j)}, M_j\right),\ \left(j = 1, ..., s_1\right);$

$s_0' = s_0' + 1; s_1 = s_1 - 1;$

$\boldsymbol{\theta}_0^{(s_0')} = \boldsymbol{\theta}_1^{(1)}$ *(Add* $\boldsymbol{\theta}_1^{(1)}$ *into* $\boldsymbol{\Theta}$*);*

$\phi_q\left(\boldsymbol{\theta}_1^{(j)}, M_j\right) = \phi_q\left(\boldsymbol{\theta}_1^{(j+1)}, M_{j+1}\right),\ \boldsymbol{\theta}_1^{(j)} = \boldsymbol{\theta}_1^{(j+1)}, M_j = M_{j+1},\ j = 1, ..., s_1$ *(Delete* $\boldsymbol{\theta}_1^{(1)}$ *from* $\boldsymbol{\Psi}$*);*

End do

Output: $\boldsymbol{\Theta}, \boldsymbol{\Psi}, s_0', s_1$

3.4 The incremental modelling of HOPSM

The order of HOPSM will largely influence the approximation accuracy. A problem to be solved is that which order is mostly suitable for balancing accuracy and efficiency, since the higher order polynomial may have higher approximation accuracy but requires more samples. This problem can be solved by the order increment strategy, which changes the polynomials from low-order to high-order gradually. The criterions of order increase and termination are based on the evaluation indexes given in Section 2.

The order of HOPSM starts from 2, and then it will increase successively until the termination conditions are satisfied. When the order of HOPSM keeps unchanged, the approximation accuracy of the surrogate model will improve gradually if the data points are added using the sequential sampling scheme. However, the improvement of the approximation accuracy will become very small when the sampling size has got a certain large number, known as the sampling saturation. When the sampling saturated, the increase the order of HOPSM to 3 may further improve the approximation accuracy. After enhancing the order, the sampling saturation may become unsaturation, as the number of coefficients to be estimated has increased. Therefore, the sequential sampling procedure will be implemented again until the sampling saturation is reached. Then the implementation of the order incremental operation was followed by the sequential sampling process repeatedly. Three evaluation indexes are used to judge when the order incremental operation should be implemented.

In the sampling saturation period, the three evaluation indexes have little variation, which can be used as the variation of these evaluation indexes as the criterion of incremental operation. The variation of the three evaluation indexes can be defined as follows:

$$\Delta\bar{R}^2_{s_0}=\left|\bar{R}^2_{s_0}\text{-}\bar{R}^2_{s_0-k}\right|,\ \ \Delta\mathrm{NAAE}_{s_0}=\frac{\left|\mathrm{NAAE}_{s_0}\text{-}\mathrm{NAAE}_{s_0-k}\right|}{\mathrm{NAAE}_{s_0}}\ \text{and}\ \Delta\mathrm{NMAE}_{s_0}=\frac{\left|\mathrm{NMAE}_{s_0}\text{-}\mathrm{NMAE}_{s_0-k}\right|}{\mathrm{NMAE}_{s_0}}\quad(27)$$

where the subscript denotes the number of sampling points, and k is the dimensional size of design variables. The incremental criterions are given by

$$\Delta\bar{R}^2_{s_0}<\varepsilon_1,\ \Delta\mathrm{NAAE}_{s_0}<\varepsilon_2,\ \text{and}\ \Delta\mathrm{NMAE}_{s_0}<\varepsilon_3\quad(28)$$

where $\varepsilon_1,\varepsilon_2,\varepsilon_3$ are some small positive percentage, e.g. 0.01%, 0.1%, and 1%, respectively.

The order will increase when all the three inequalities are satisfied. In some cases the rate of convergence of these indexes is very slow, so we add another incremental criterion, given by

$$s_0\geq 2N_C\left(n,k\right)\quad(29)$$

This incremental criterion may avoid the phenomenon that the order of HOPSM is trapped at low orders. When one of the criterions in (28) and (29) is satisfied, the order increment operation will be performed.

The following termination criterions are also based on the three indexes. The first one is the same as incremental criterion given in inequality (28), while the second is expressed as follows:

$$\bar{R}^2_{s_0} \geq \mu_1, \text{NAAE}_{s_0} \leq \mu_2, \text{ and } \text{NMAE}_{s_0} \leq \mu_3 \tag{30}$$

where μ_1 is a number close to but smaller than 1, e.g. 0.99, and μ_2 and μ_3 are two small positive percentages (e.g. 2% and 10%), respectively. The third termination criterion is the number of sampling points, which should not exceed the allowed maximum sampling size.

$$s_0 \geq N_{\max} \tag{31}$$

The procedure of the sequential sampling will be stopped when both the inequalities in (28) and (30) are satisfied simultaneously, or the inequality (31) is satisfied.

When the sampling process is finished, the highest order of the HOPSM may not be the best order in demand, and the best order will be chosen by using the adjusted R-square. Therefore, after the sequential sampling is terminated, we will fit the HOPSM with respect to all the allowable orders, the coefficients to be estimated are less than the sampling points. The order of the HOPSM which has the largest value of adjusted R-square will be selected as the final order. Use $\bar{R}^2_{n,s_0}$ to denote the adjusted R-square of HOPSM with the order n and sampling size s_0. The best order of HOPSM will be determined by

$$n_{opt} = \left\{ n \mid \max_n \bar{R}^2_{n,s_0}, N_C(n,k) < s_0 \right\} \tag{32}$$

The final order of HOPSM will be n_{opt}, and the best HOPSM is given by $\hat{f}_{n_{opt}}(\mathbf{x})$ produced by s_0 samples.

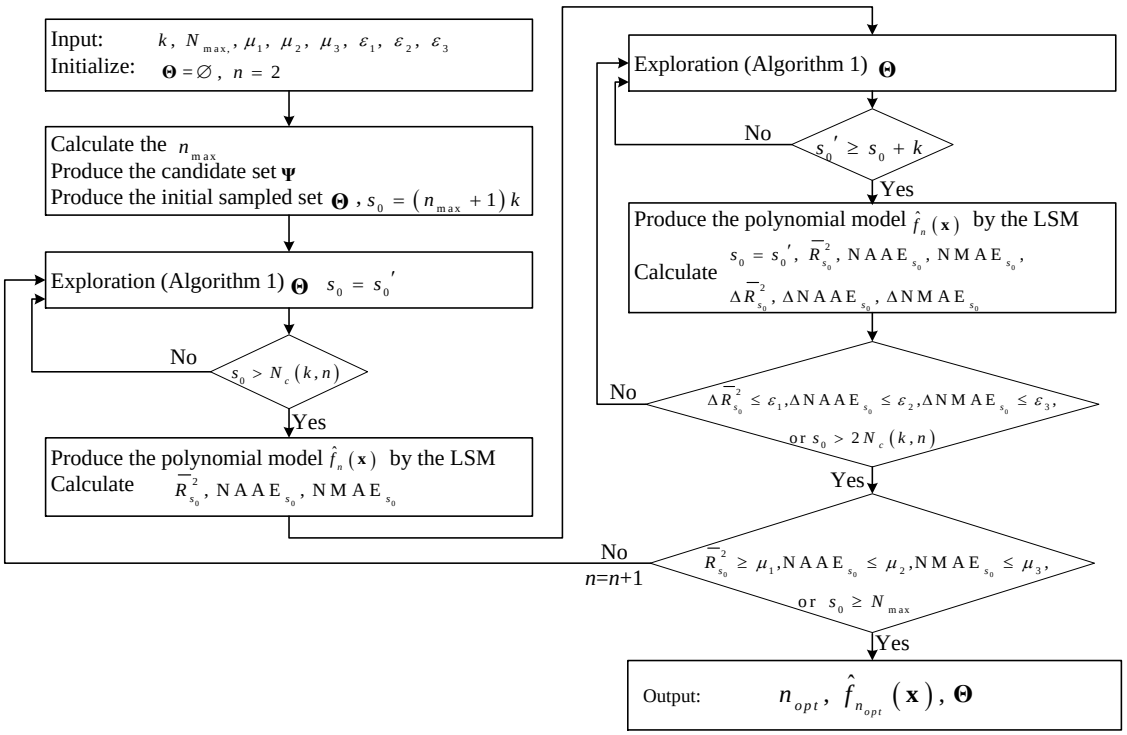


Figure.5 The flow chart of constructing HOPSM

The flow chart of constructing HOPSM is shown in Fig. 5, which mainly includes five steps: candidates generation, initial sampling, exploration sampling, order increment, and order determination.

The candidates $\boldsymbol{\Psi}$ are the tensor product of the zeros of Chebyshev polynomials, shown in Section 3.1. The initial sampling method has been given in Section 3.2, and it will produce a small uniform sampling set $\boldsymbol{\Theta}$. The algorithm of the exploration sampling is given in Section 3.3, which selects more samples from the candidates and keeps the data points distribute uniformly. The order incremental plan is used to update the order of HOPSM, and improve the approximation accuracy gradually. The exploration sampling and order incremental operation will be repeated until the termination criterions are satisfied. Lastly, the order determined by Eq. (32) may make HOPSM best.

4. Numerical examples

4.1 Mathematical test examples

Here we consider several benchmark mathematical testing examples, their expressions are given in Table 5. These mathematical examples contain strong nonlinear characteristics and different dimensional size, which are suitable for demonstrate the accuracy and robustness of the proposed surrogate model.

Table 5 Mathematical test functions

Functions	Expression	Domain	Dimension (k)
Michalewicz	$f(\mathbf{x})=\sum_{i=1}^{k}\sin(x_i)\left[\sin\left(\frac{ix_i^2}{\pi}\right)\right]^{20}$	$0\le x_i\le\pi$	2
Ackley	$f(\mathbf{x})=-20\exp\left(-0.2\sqrt{\frac{1}{k}\sum_{i=1}^{k}x_i^2}\right)-\exp\left(\frac{1}{k}\sum_{i=1}^{k}\cos 2\pi x_i\right)+20+\exp(1)$	$-2\le x_i\le 2$	3
Deceptive	$f(\mathbf{x})=-\left(\frac{1}{k}\sum_{i=1}^{k}g_i(x_i)\right)^{0.5}$, $g_i(x_i)=\begin{cases}-x_i/\alpha_i+0.8 & 0\le x_i\le 0.8\alpha_i\\ 5x_i/\alpha_i-4 & 0.8\alpha_i\le x_i\le\alpha_i\\ 5(x_i-\alpha_i)/(\alpha_i-1)+1 & \alpha_i<x_i\le 0.2+0.8\alpha_i\\ (1-x_i)/(\alpha_i-1)+0.8 & 0.2+0.8\alpha_i<x_i\le 1\end{cases}$	$0\le x_i\le 1$	4
Rastrigin	$f(\mathbf{x})=10k+\sum_{i=1}^{k}\left(x_i^2-10\cos(2\pi x_i)\right)$	$-1\le x_i\le 1$	5
Schwefel	$f(\mathbf{x})=\sum_{i=1}^{k}\left(-x_i\sin\left(\sqrt{\lvert x_i\rvert}\right)\right)$	$-100\le x_i\le 100$	6

The maximum allowable sampling size to construct the HOPSM is set as 200, 400, 600, 800, and 1000, respectively. The R2 and NMAE shown in Section 2.3 are used to validate the approximation accuracy of HOPSM. The well-known RBF and Kriging model will be employed to compare with the proposed HOPSM. For the sampling methods, the proposed Chebyshev incremental sampling strategy will be compared with the Halton sequential sampling in [45]. To keep the uniformity of test points, we use the Hammersely [49] sequence to produce 10000 test points. The plot of the R-square and NAME are shown in Figs. 6 to 10, respectively.

The legend is expressed as 'surrogate model - sampling method', where 'C' denotes the Chebyshev sampling and 'H' denotes the Halton sequential sampling, e.g. the 'HOPSM-C' means the HOPSM based on the Chebyshev sampling.

For the Michalewicz function in Fig. 6, the Kriging-Halton gives the best R-square, followed by the HOPSM-Chebyshev. The RBF-Halton and RBF-Chebyshev have close R-square, while the Kriging-Chebyshev is not good when the sampling size is large. For the NMAE, the HOPSM-Chebyshev has the smallest error, followed by the Kriging-Halton. RBF-Chebyshv and RBF-Halton still give close NMAE, but the Halton sampling provides better result when the sampling size is relatively small. HOPSM-Halton gives the worst result, so it is not shown in the figure to make other curves can be distinguished easily.

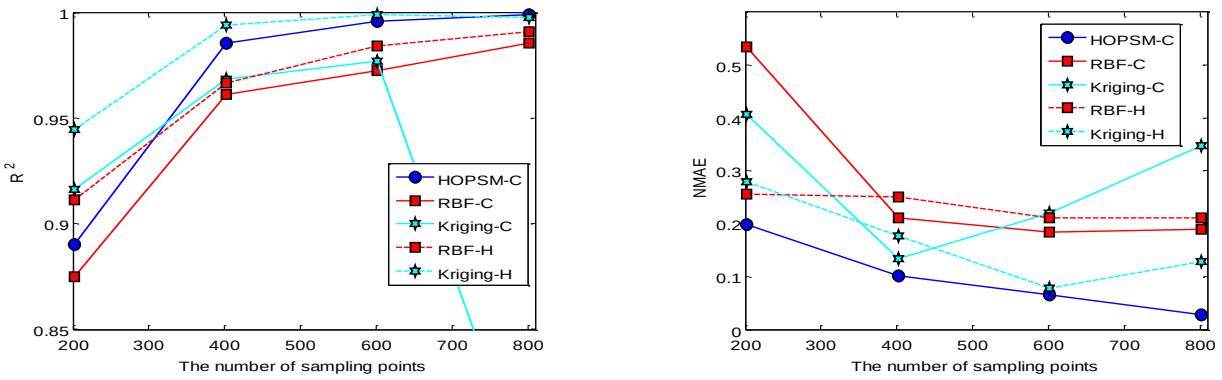


Figure.6 R^2 and NMAE for the Michalewicz function

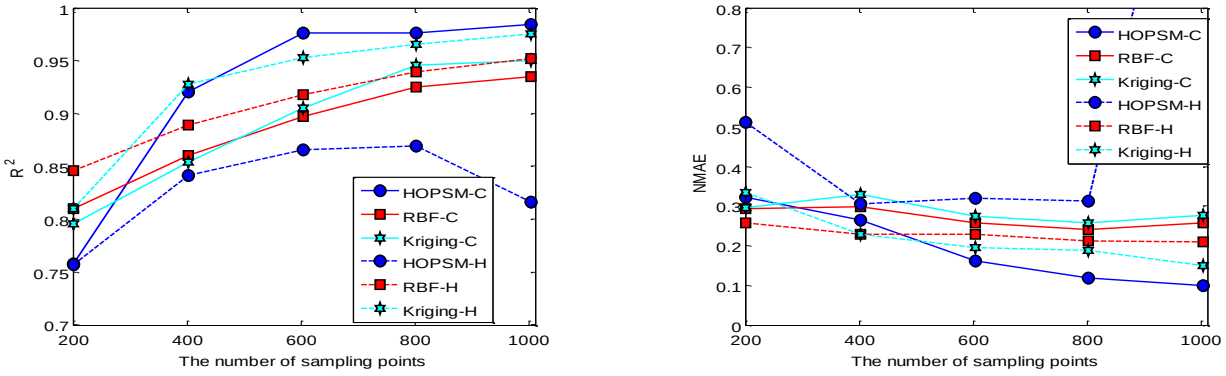


Figure.7 The R^2 and NMAE for the Ackley function

For the Ackley function, both the R-square and NMAE show that the HOPSM-Chebyshev has the best performance, and then the Kriging-Halton has good accuracy, following with the RBF-Halton, RBF-Chebyshev and Kriging-Chebyshev. The HOPSM-Halton provides the worst accuracy.

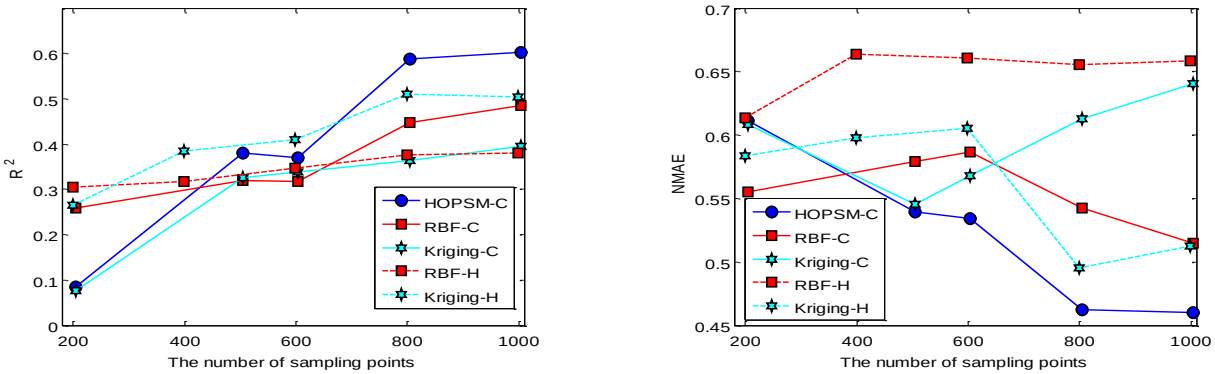


Figure.8 The R^2 and NMAE for the Deceptive function

The performance of the Deceptive function can be ranked from high to low as follows: HOPSM-Chebyshev, Kriging-Halton, RBF-Chebyshev, Kriging-Chebyshev, RBF-Halton, HOPSM-Halton (which is not shown in the figure).

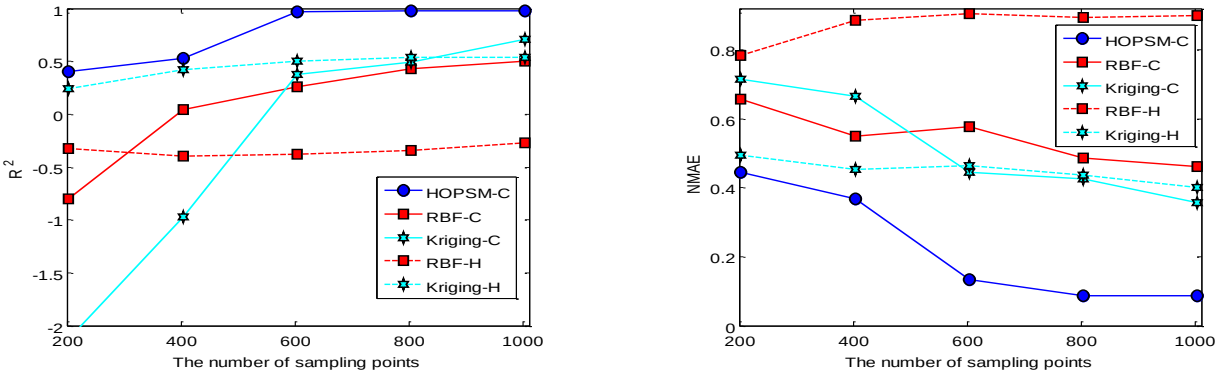


Figure.9 The R^2 and NMAE for the Rastrigin function

For the Rastrigin function, the convergence ratio can be ranked from high to low as follows: HOPSM-Chebyshev, Kriging-Halton, Kriging-Chebyshev, RBF-Chebyshev, RBF-Halton, HOPSM-Halton (which is not shown in the figure).

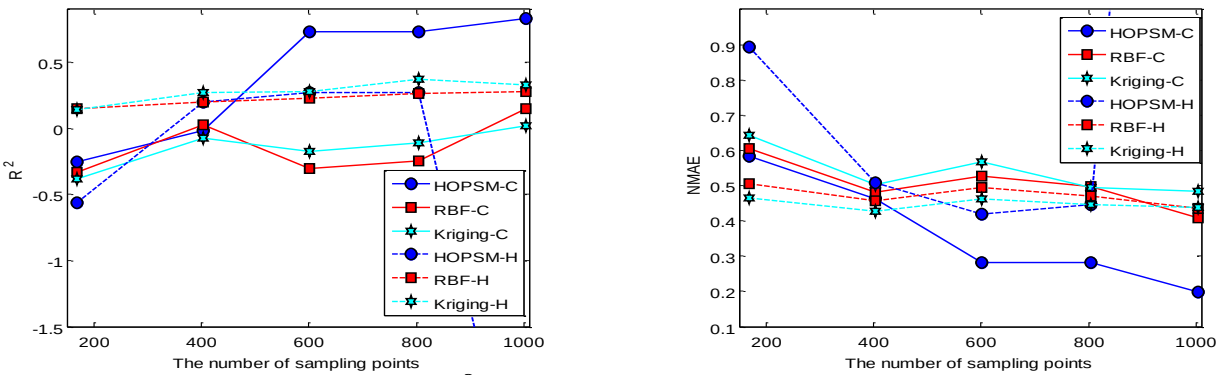


Figure.10 The R^2 and NMAE for the Schwefel function

The HOPSM-Chebyshev has the highest convergence ratio for the Schewefel function. The Kriging and RBF model using the Halton sampling have slightly better accuracy than using the Chebyshev sampling.

To compare these models comprehensively, we plot the mean value and standard deviation of R-square and NMAE based on these test functions, shown in Fig. 11. Both the mean and standard deviation of the two indexes show that the HOPSM-Chebyshev has the best performance, when the sampling size is relatively large (e.g. > 500), while the Kriging-Halton performs the best accuracy and robustness when the sampling size is relatively small (e.g. < 400), but the HOPSM-Chebyshev also has better performance than other methods (RBF and Kriging-Chebyshv) with the sampling size. The NMAE of both the Kriging-Halton and Kriging-Chebyshev become worse when the sampling size close to 1000, so the Kriging model may be instable for problems with large sampling sizes. For the mean value, the RBF-Halton is better than the RBF-Chebyshev for a small number of samples, but the result is contrary when the sampling size becomes large. The standard deviation of RBF-Chebyshev is smaller than that of RBF-Halton, so the former should be more robust.

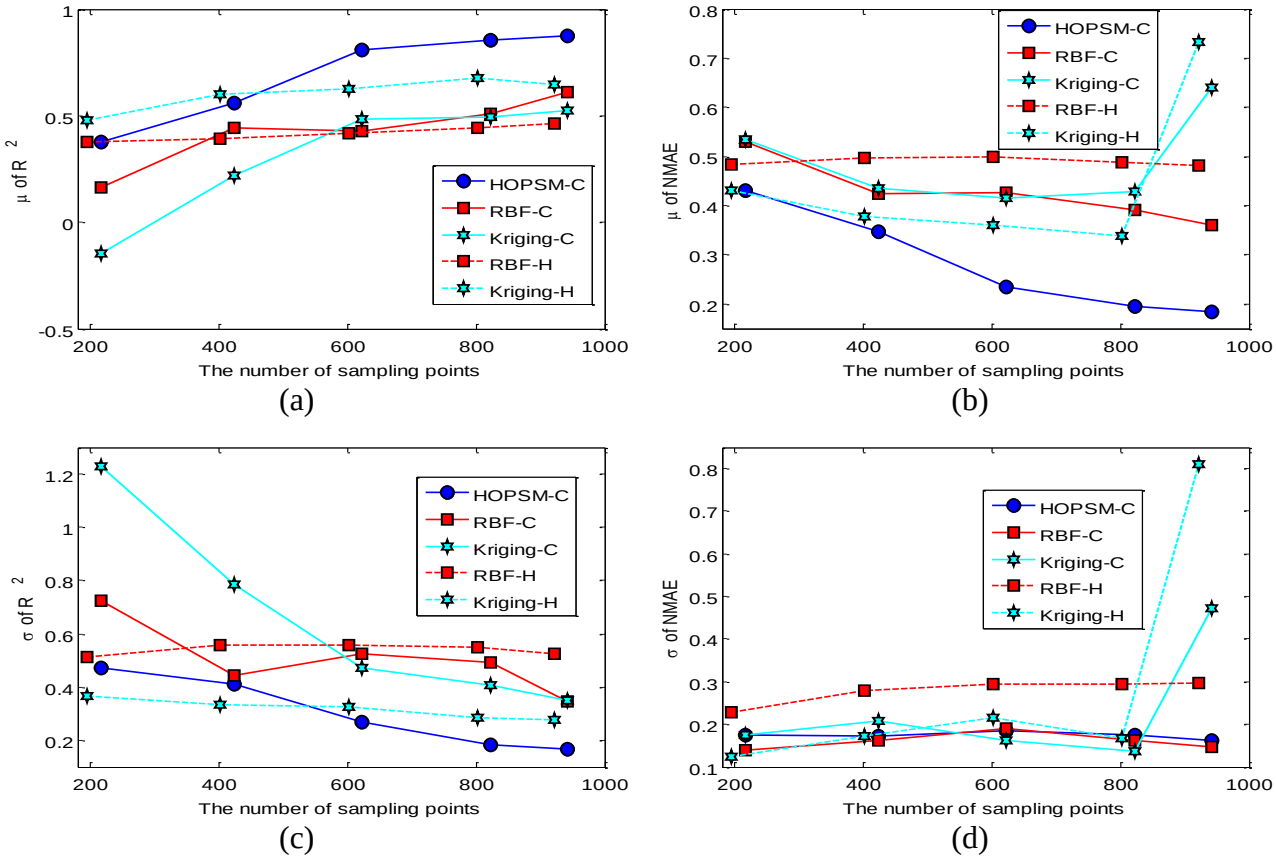


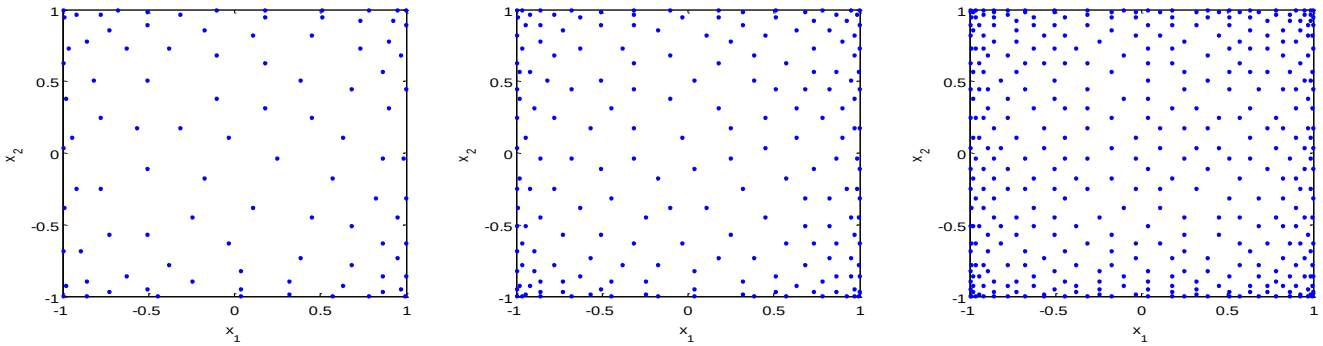
Figure.11 The robustness of the test functions

The HOPSM-Halton (not shown in the figure) has the worst performance with respect to both R-square and NMAE. The worst performance of HOPSM-Halton may be explained by the instability of high-order polynomials, e.g. the Runge phenomenon, but it can be overcome by using the zeros of Chebyshev

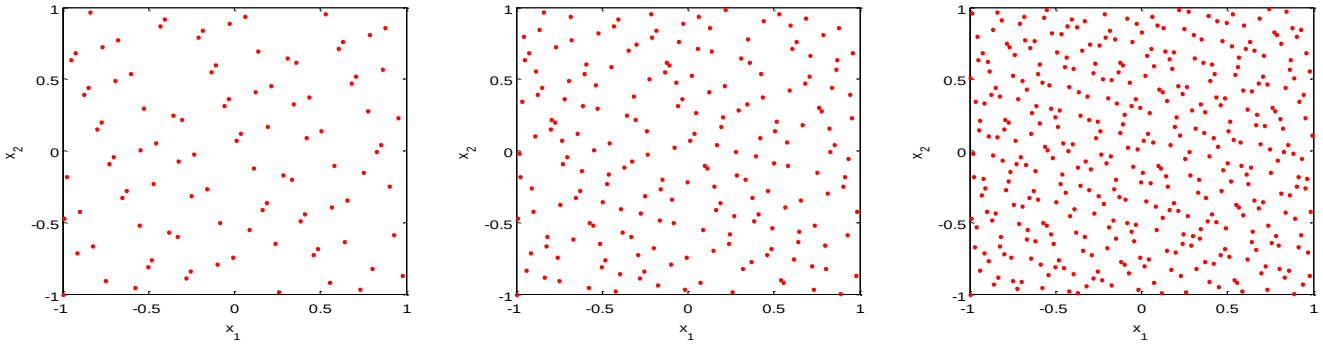
polynomials as the sampling points, shown as the best performance of HOPSM-Chebyshev. The Kriging model gives an opposite conclusion that the Halton sampling is much better than the Chebyshev sampling. The RBF model is insensitive to the sampling method compared with HOPSM and Kriging model.

For the computational cost, the five test functions takes 155s, 165s, 129s, 120s, and 145s on the computer equipped with a 2.60 GHz i5-2540M CPU, respectively. In surrogate modelling, this computational cost is quite low compared with the practical engineering problems, since each running of the engineering model may take more than several minutes or even several hours. Therefore, the computational cost for surrogate modelling is very cheap, while the most expensive part is the sampling for original complicated simulation model.

To show the two sampling methods more directly, the locations of the 2-dimensional sampling points are shown in Fig. 12, considering a sampling size 100, 200, and 400, respectively. It can be found that the samples of the Chebyshev sequential sampling method distribute more uniformly than that of the Halton sequential sampling method when the sampling size is small. There are more samples located in the regions closing to the boundary of the design space for the Chebyshev sequential sampling scheme with the increase of the sampling size.



(a) 100 samples - Chebyshev (b) 200 samples - Chebyshev (c) 400 samples - Chebyshev



(d) 100 samples - Halton (e) 200 samples - Halton (f) 400 samples - Halton

Figure.12 The location of samples for Chebyshev sampling and Halton sampling

4.2 Engineering Applications

Application 1: Vehicle Handling

The vehicle handling stability affects the safety of vehicles, especially when the vehicle speed is high. Figure 13 shows a state of the vehicle turning [50]. The vehicle handling characteristic can be studied by simplifying the vehicle model to a 2-DOF bicycle model without considering the influence of steering system and suspension system, and assuming that the longitudinal velocity of vehicle keeps constant, as shown in Fig. 14. O is the center of gravity of the vehicle, and O' is the turning center of vehicle. F_{Y1} and F_{Y2} are the lateral force of front and rear wheels, respectively. a and b are the distance from the front and rear axles to the center of gravity, respectively. u denotes the longitudinal velocity of the vehicle which is set as 110 km/h, v denotes the lateral velocity for the center of gravity, respectively. u_1 and u_2 are the velocity for front and rear wheels, respectively. α_1, α_2 and β denote the slip angle for the front wheel, rear wheel and center of gravity. δ is the steering angle for the front wheel and ξ is the angle between the direction of velocity for front wheel and x axis. ω is the yaw velocity of the vehicle.

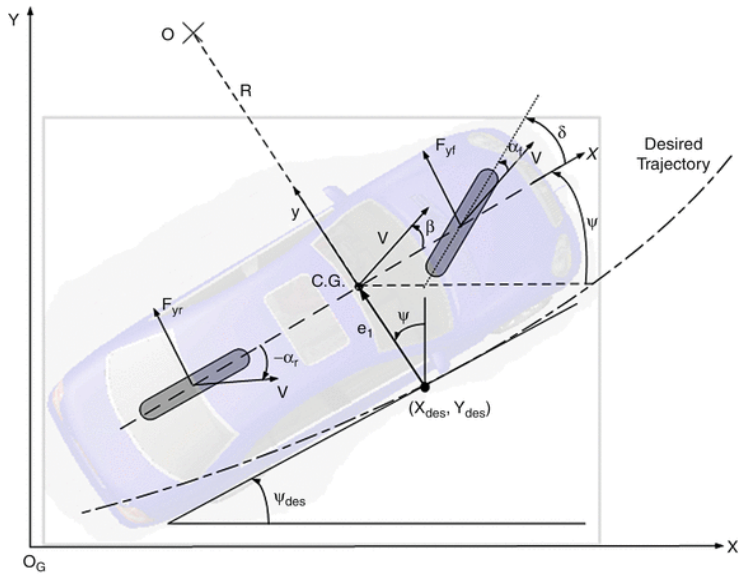


Figure. 13 Vehicle turning

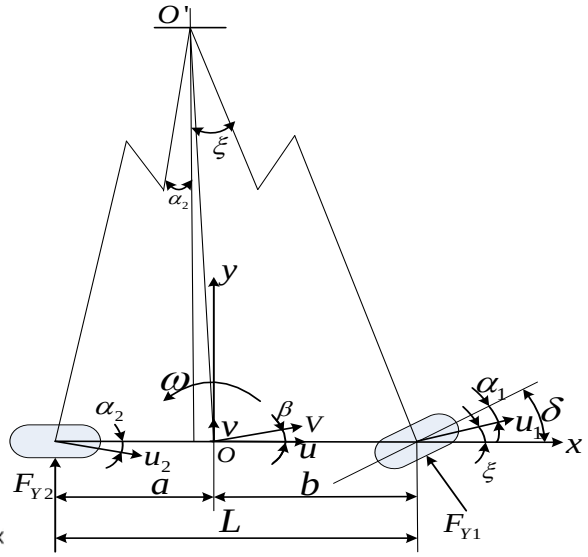


Figure. 14 The 2DOF Bicycle model for vehicle

We consider the dynamic response of the steer angle step input, which increases the steering angle and then keeps the steering angle unchanged. The step input of the front tire turning angle is shown in Fig. 15. More detailed expressions can be found in reference [50].

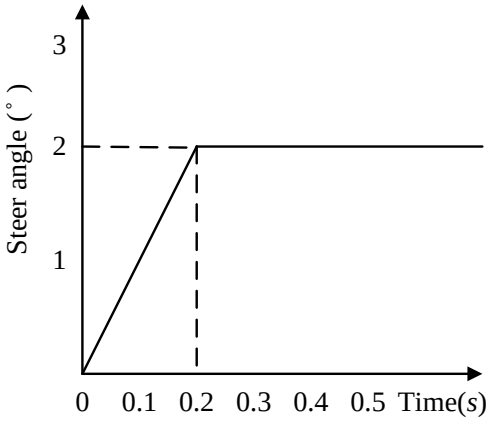


Figure. 15 Steering angle step input

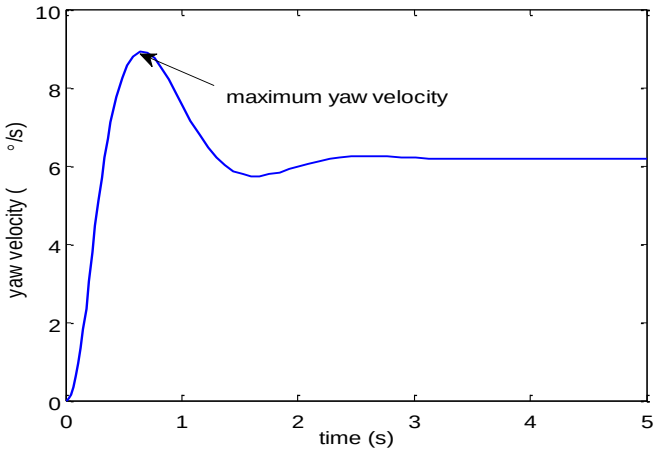


Figure. 16 The yaw velocity

As shown in Table 6, we identify 6 design variables and will build a surrogate model of the maximum yaw velocity, which has large influence on the vehicle handling characteristic. Solving the 2-DOF model, the yaw velocity when the design variables are equal to the mean value is shown as Fig. 16. The HOPSM, RBF, and Kriging model are used as the surrogate model, respectively, while both the Chebyshev and Halton sampling schemes are applied to produce 50, 100, 150, 200, and 250 sampling points, respectively. 10000 Hammersely points are used as the test points.

Table 6 Design variables of a car

Parameters	Mass (kg)	Mass moment of inertia $(kg.m^2)$	a (m)	b (m)	Cornering Stiffness of front tire (N/rad)	Cornering Stiffness of rear tire (N/rad)
Range	[1100, 1300]	[1900, 2100]	[1.02, 1.22]	[1.12, 1.32]	[-3.5,-2.5]×10^4	[-5,-4]×10^4

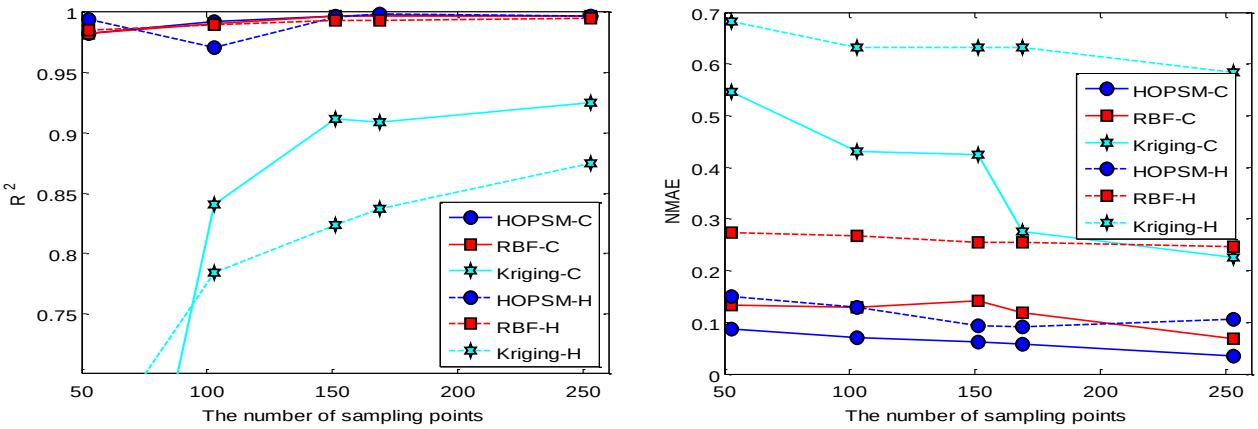


Figure. 17 R^2 and NMAE for the maximum yaw velocity

The R-square and NMAE of the maximum yaw velocity are given in Fig. 17. For the R-square, the HOPSM and RBF have very close performance and are much better than the Kriging model. For the

NMAE, the HOPSM-Chebyshev has the highest accuracy, and then the RBF-Chebyshev and HOPSM-Halton. The Kriging model still gives the worst result.

Application 2: Engine Mount

In this section we use the HOPSM to study the ride comfort of automotive. A typical work condition for ride comfort is to drive the car across an obstacle in a constant speed, and then measure the acceleration of floor or the driver's seat. The memory shake of the driver's seat is a dimensionless number, which is used as the evaluation index to characterise the performance of ride comfort. We use the AES Bump as the obstacle, the section plot is given in Fig. 18, and the vehicle velocity is 20 mph.

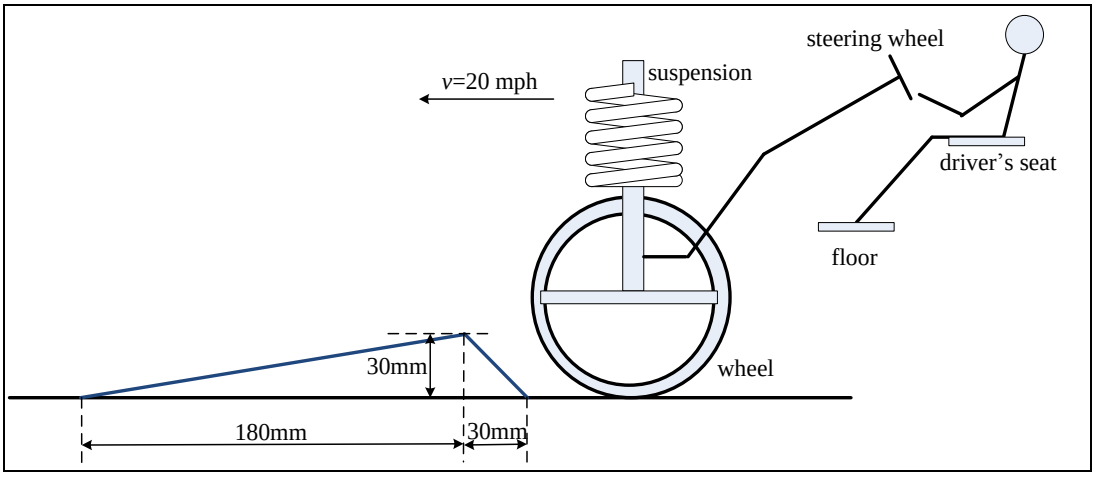


Figure.18 The schematic of ride comfort experiment

The simulation model is built by using Hyperworks MotionView, and the time for each simulation is about 450 seconds, which would be too long to analysis or optimization of this system directly. Therefore, we will use the proposed method to build a surrogate model of the ride comfort as the simulation model. Here five parameters of the engine mount are considered as the design variables, the description and range are shown in Table 7. There are three engine mounts, and the locations are shown in Fig. 19.

Table 7. The description of design variables

Parameters	Description	Lower bound	Upper bound
Keng	Engine mount vertical stiffness (*N/mm*)	200	340
Ktrans	Transmission mount vertical stiffness (*N/mm*)	200	410
Ceng	Engine mount vertical damping (*N s/mm*)	0.2	4.5
Ctrans	Transmission mount vertical damping (*N s/mm*)	0.2	4.5
Kts	Torque strut lateral stiffness (*N/mm*)	150	300

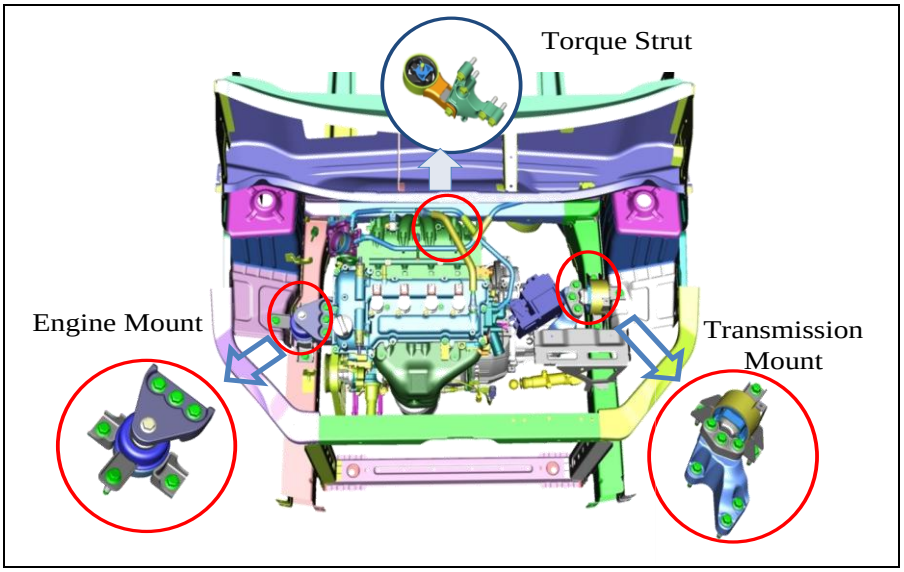


Figure.19 The collocation of mounts

Since the computational cost of the simulation model is expensive, 128 samples will be collected by using the proposed Chebyshev sampling method and the Latin Hypercube sampling (LHS) method, respectively. Use the two groups of 128 samples to construct the HOPSM of the simulation outputs (memory shake) with order 2 and 3, respectively. To save the cost of test points, 500 Hammersely points and additional 21 data points produced by the orthogonal table will be used as the test points. The results are shown in Table 8. It can be found that the Chebyshev sampling provides better result than the Latin Hypercube sampling for all the three surrogate models, especially for NMAE. The NMAE of HOPSM based on Chebyshev sampling has the best accuracy, while the Kriging model and RBF model are equivalent. For the R-square, the RBF and HOPSM are very close and have better performance than the Kriging model.

Table 8. Approximation accuracy of surrogate model for memory shake

	Chebyshev sampling		Latin Hypercube sampling	
	R^2	NMAE	R^2	NMAE
HOPSM	0.7385	0.3236	0.7193	0.5067
RBF	0.7435	0.3423	0.7352	0.5038
Kriging	0.7042	0.3423	0.6521	0.5038

5. Conclusions

This paper has proposed a new high-order polynomial surrogate model (HOPSM), based on a Chebyshev sequential sampling scheme and an incremental modelling process. There are two reasons accouting for the use of the Chebyshev polynomials: (1) the first of which is that we would employ the zeros of the

Chebyshev polynomials as the set of all sampling candidates. These samples can avoid the Runge phenomenon occurred in polynomials with high-order interpolation; (2) the second of which is that the truncated Chebyshev series is closer to the best uniform approximation, which may improve the approximation accuracy of the surrogate modelling. The sequential sampling and incremental modelling can flexibly control the sampling size and the approximation accuracy. To demonstrate the performance of the proposed HOPSM, the well-known Kriging model and RBF model are used as the references. Through the comparison by using several strong nonlinear mathematic testing examples and two engineering applications, the results show that the Chebyshev sequential sampling based HOPSM has good accuracy and robustness. However, it should be noted that the HOPSM is only suitable for the Chebyshev sampling method, because other sampling methods may make it unstable, while other surrogate models, e.g. RBF model, may be fit for various types of sampling methods.

ACKNOWLEDGEMENTS

This research is supported in part by Australian Research Council (Discovery Projects) (DP150102751; DP160102491), and the National Natural-Science-Foundation of China (51575204; 11502083), the China Postdoctoral Science Foundation (2014M562014), and the Science and Technology Support Program of Hubei Province of China (2015BHE026), as well as the China Scholarship Council (CSC).

Reference

[1] K. Crombecq, L. De Tommasi, D. Gorissen, T. Dhaene, A novel sequential design strategy for global surrogate modeling, in: Proceeding in 2009 Winter Simulation Conference, Austin, TX, USA, 2009, pp. 731-742.

[2] L.P. Swiler, R. Slekpoy, A.A. Giunta, Evaluation of sampling methods in constructing response surface approximations, in: the 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Newport, Rhode Island, 2006, pp. AIAA-2006-1827.

[3] B.M. Rutherford, L.P. Swiler, T.L. Paez, A. Urbina, Response Surface (Meta-model) Methods and Applications, in, SANDIA Nat. Lab., 2005.

[4] R.H. Myers, D. Montgomery, Response Surface Methodology: Process and Product Optimization Using Designed Experiments, John Wiley and Sons, Inc., Toronto, 1995.

[5] S. Shan, G.G. Wang, Metamodeling for high dimensional simulation-based design problems, Journal of Mechanical Design, 132 (2010) 051009:051001-051011.

[6] T.W. Simpson, T.M. Mauery, J.J. Korte, F. Mistree, Kriging models for global approximation in simulation-based multidisciplinary design optimization, AIAA Journal, 39 (2001) 2233-2241.

[7] L. Zhang, Z. Lu, P. Wang, Efficient structural reliability analysis method based on advanced Kriging model, Applied Mathematical Modelling, 39 (2015) 781-793.

[8] J.H. Friedman, Multivariate adaptive regression splines, Annals of Statistics 19 (1991) 1-141.

[9] E. Andrés, S. Salcedo-Sanz, F. Monge, A.M. Pérez-Bellido, Efficient aerodynamic design through evolutionary programming and support vector regression algorithms, Expert Systems with Applications, 39 (2012) 10700-10708.

[10] S.M. Clarke, J.H. Griebsch, T.W. Simpson, Analysis of Support Vector Regression for Approximation of Complex Engineering Analyses, Journal of Mechanical Design, 127 (2005) 1077.

[11] K.H. Hajikolaei, G.G. Wang, Adaptive orthonormal basis functions for high dimensional metamodeling with existing sample points, in: ASME 2012 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference Chicago, IL, USA, 2012, pp. 709-716.

[12] H. Wang, L. Tang, G.Y. Li, Adaptive MLS-HDMR metamodeling techniques for high dimensional problems, Expert Systems with Applications, (2011).

[13] E. Acar, Various approaches for constructing an ensemble of metamodels using local measures, Structural and Multidisciplinary Optimization, 42 (2010) 879-896.

[14] T. Goel, R.T. Haftka, W. Shyy, N.V. Queipo, Ensemble of surrogates, Structural and Multidisciplinary Optimization, 33 (2006) 199-216.

[15] E. Acar, Effect of error metrics on optimum weight factor selection for ensemble of metamodels, Expert Systems with Applications, 42 (2015) 2703-2709.

[16] R. Jin, W. Chen, T.W. Simpson, Comparative studies of metamodelling techniques under multiple modelling criteria, Structural and Multidisciplinary Optimization, 23 (2001) 1-13.

[17] T.W. Simpson, D.K.J. Lin, W. Chen, Sampling Strategies for Computer Experiments Design and Analysis, International Journal of Reliability and Applications, 2 (2001) 209-240.

[18] J.J.M. Rijpkema, L.F.P. Etman, A.J.G. Schoofs, Use of design sensitivity information in response surface and kriging metamodels, Optimization and Engineering, 2 (2001) 469-484.

[19] R.R. Barton, Metamodeling: a state of the art review, in: the 1994 Winter Simulation Conference, Lake Buena Vistak, Florida, 1994.

[20] R.R. Barton, Simulation metamodels, in: the 1998 Winter Simulation Conference, Washington D. C., 1998.

[21] A. Gallina, Response surface methodology as a tool for analysis of uncertainty in structural dynamics, in, AGH - University of Science and Technology, 2009.

[22] G.G. Wang, S. Shan, Review of metamodeling techniques in support of engineering design optimization, Journal of Mechanical Design, 129 (2007) 370-380.

[23] J.G. Lin, Modeling Test Responses by Multivariable Polynomials of Higher Degrees, SIAM Journal on Scientific Computing, 28 (2006) 832-867.

[24] S.A. Yousefi, M. Behroozifar, M. Dehghan, Numerical solution of the nonlinear age-structured population models by using the operational matrices of Bernstein polynomials, Applied Mathematical Modelling, 36 (2012) 945-963.

[25] S.A. Yousefi, M. Behroozifar, M. Dehghan, The operational matrices of Bernstein polynomials for solving the parabolic equation subject to specification of the mass, Journal of Computational and Applied Mathematics, 235 (2011) 5272-5283.

[26] A. Saadatmandi, M. Dehghan, The numerical solution of problems in calculus of variation using Chebyshev finite difference method, Physics Letters A, 372 (2008) 4037-4040.

[27] M.R. Eslahchi, M. Dehghan, S. Amani, The third and fourth kinds of Chebyshev polynomials and best uniform approximation, Mathematical and Computer Modelling, 55 (2012) 1746-1762.

[28] K. Parand, M. Dehghan, F. Baharifard, Solving a laminar boundary layer equation with the rational Gegenbauer functions, Applied Mathematical Modelling, 37 (2013) 851-863.

[29] R.R. Barton, Metamodels for simulation input-output relations., in: the 24th conferencee on Winter simulation, New York, 1992, pp. 289-299.

[30] A. Gil, J. Segura, N.M. Temme, Numerical Methods for Special Functions, SIAM, 2007.

[31] L. Fox, I.B. Parker, Chebyshev Polynomials in Numerical Analysis, Oxford University Press, London, 1968.

[32] J. Wu, Y. Zhang, L. Chen, Z. Luo, A Chebyshev interval method for nonlinear dynamic systems under uncertainty, Applied Mathematical Modelling, 37 (2013) 4578-4591.

[33] J. Wu, Z. Luo, Y. Zhang, N. Zhang, L. Chen, Interval uncertain method for multibody mechanical systems using Chebyshev inclusion functions, International Journal for Numerical Methods in Engineering, 95 (2013) 608-630.

[34] J. Wu, Z. Luo, Y. Zhang, N. Zhang, An interval uncertain optimization method for vehicle suspensions using Chebyshev metamodels, Applied Mathematical Modelling, (2014).

[35] D.C. Montgomery, Design and Analysis of Experiments, 6th ed., Post & Telecom Press, Beijing, 2007.

[36] A.A. Giunta, J.S.F. Wojtkiewicz, M.S. Eldred, Overview of Modern Design of Experiments Methods for Computational Simulations, in: 41st AIAA Aerospace Sciences Meeting and Exhibit, Reno, NV, 2003, pp. AIAA Paper 2003-0649.

[37] M.D. Mckay, R.J. Beckkman, W.J. Conover, A comparison of three methods for selceting values of input variables in the analysis of output from a computer code, Technometrics, 21 (1979) 266-294.

[38] A.B. Owen, Orthogonal arrays for computer experiments, integration and visualiztion, Statistica Sinca, 2 (1992) 439-452.

[39] K.T. Fang, D.K.J. Lin, P. Winker, Y. Zhang, Uniform Design: Theory and Application, Technometrics, 39 (2000) 237-248.

[40] A. Ajdari, H. Mahlooji, An Adaptive Exploration-Exploitation Algorithm for Constructing Metamodels in Random Simulation Using a Novel Sequential Experimental Design, Communications in Statistics - Simulation and Computation, 43 (2014) 947-968.

[41] D. Gorissen, I. Couckuyt, P. Demeester, T. Dhaene, K. Crombecq, A Surrogate Modeling and Adaptive Sampling Toolbox for Computer Based Design, Journal of Machine Learning Research, 11 (2010) 2051-2055.

[42] V.J. Romero, L.P. Swiler, A.A. Giunta, Construction of response surfaces based on progressive-lattice-sampling experimental designs with application to uncertainty propagation, Structural Safety, 26 (2004) 201-219.

[43] K. Crombecq, D. Gorissen, D. Deschrijver, T. Dhaene, A Novel Hybrid Sequential Design Strategy for Global Surrogate Modeling of Computer Experiments, SIAM Journal on Scientific Computing, 33 (2011) 1948-1974.

[44] V.J. Romero, R. Slepoy, L.P. Swiler, A.A. Giunta, Error estimation approaches for progressive response surfaces - more results, in: 46th AIAA/ASME/ASCE/AHS/ASC Structural Dynamics and Materials Conference, Austin, TX, 2005, pp. 269-289.

[45] T.-T. Wong, W.-S. Luk, P.-A. Heng, Sampling with Hammersley and Halton Points, Journal of Graphics Tools, 2 (1997) 9-24.

[46] M.E. Jonson, L.M. Moore, D. Yalvisaker, Minimax and maximin distance designs, Journal of Statistical Planning and Inferences, 26 (1990) 131-148.

[47] X. Qu, G. Venter, R.T. Haftka, New formulation of a minimum-bias central composite experimental design and Gauss quadrature, Structural and Multidisciplinary Optimization, 28 (2004) 231-242.

[48] M.D. Morris, T.J. Mitchell, Exploratory design for computational experiments, Journal of Statistical Planning and Inferences, 43 (1995) 381-402.

[49] J.R. Kalagnanm, U.M. Diwekar, An Efficient Sampling Technique for off_line quality control, American Statistical Association and the American Society for Quality Control Technometrics, 39 (1997).

[50] D. Soudbakhsh, D.S. Azim Eskandarian, Vehicle lateral and steering control, in: D.S. Azim Eskandarian (Ed.) Handbook of intelligent vehicles, Springer-Verlag, London, 2012, pp. 209-232.