Applied Soft Computing 46 (2016) 26–36

Contents lists available at ScienceDirect

Applied Soft Computing

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

Two-level Multi-surrogate Assisted Optimization method for high dimensional nonlinear problems

Enying Li [a], Hu Wang [b,c,*], Fan Ye [b,c]

[a] School of Logistics, Central South University of Forestry and Teleology, Changsha 41004, PR China
[b] State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University, Changsha 410082, PR China
[c] Joint Center for Intelligent New Energy Vehicle, Shanghai 201804, PR China

ARTICLE INFO

Article history:
Received 6 November 2015
Received in revised form 22 March 2016
Accepted 26 April 2016
Available online 4 May 2016

Keywords:
SAO
HDMR
Separable
GMDH
Multi-surrogate

ABSTRACT

Curse of dimensionality is a key issue in engineering optimization. When the dimension increases, distribution of samples becomes sparse due to expanded design space. To obtain accurate and reliable results, the amount of samples often grows exponentially with the dimensions. To improve the efficiency of the surrogate with limited samples, a Two-level Multi-surrogate Assisted Optimization (TMAO) is suggested. The framework of the TMAO is to decompose a complicated problem into separable and non-separable components. In the first-level, High Dimensional Model Representation (HDMR) is utilized to determine the correlations among input variables. Then, a high dimensional problem can be decomposed into separable and non-separable components. Thus, the dimension of the original problem might be reduced significantly. Moreover, considering noises and outliers, Support Vector Regression (SVR)-HDMR is utilized to obtain more reliable surrogate. Expected Improvement (EI) criterion is suggested to generate new samples to save computational cost. In the second-level, to handle the non-separable component, a multi-surrogate assisted sampling strategy is suggested. Compared with other methods, the distinctive characteristic of the suggested sampling strategy is to use different surrogates to search potential uncertain regions. Considering the diversity of surrogates, more feature samples might be generated close to the local optimum. Even though it is still difficult to find a global solution, it could help us to find a feasible solution in practice. To verify the performance of the suggested method, several high dimensional mathematical functions are tested by the suggested method. The results demonstrate that all test functions can be successfully solved.

1. Introduction

In the past 30 years, Surrogate Assisted Optimization (SAO) methods have been extensively used in multidiscipline. The SAO is invoked as a substitution for physical models or simulation-based evaluations, improving the efficiency of optimization procedure significantly. However, with the development of complexity of practical engineering design, it is difficult for the SAO to handle such complicated problems, especially for high dimensional problems. Generally, the essential of SAOs is approximation. To improve the performance of surrogate in terms of accuracy and reliability, various surrogate modeling methods have been developed and applied in various disciplines in the past 30 years, such as Polynomial Regression (PR) [1], Radial Basis Function (RBF) [2–4], Moving Least Square (MLS) [5,6], Kriging [7,8] and SVR [9,10]. The most influential SAO might be Efficient Global Optimization (EGO) developed by Jones et al. [11], which can be utilized to find global optimization assisted by Kriging. Chen et al. used heuristics method

Abbreviations: AI, Artificial Immune; EICAM, EI criterion assisted modeling; EGO, Efficient Global Optimization; EI, Expected Improvement; GA, Genetic Algorithm; GMDH, Group Method of Data Handling; GSS, Golden Section Sampling; HDMR, High Dimensional Model Representation; LOOCV, Leave-One-Out Cross Validation; MAS, Multi-surrogate Assisted Sampling algorithm; EImax, maximum EI; MLS, Moving Least Square; NTS, Number of Test Samples; PR, Polynomial Regression; PSO, Particle Swarm Optimization; RAAE, Relative Average Absolute Error; RBF, Radial Basis Function; RMAE, Relative Maximum Absolute Error; SA, Simulated Annealing; STD, Standard Deviation; SAEO, Surrogate Assisted Evolutionary Optimization; SAO, Surrogate Assisted Optimization; SVR, Support Vector Regression; TMAO, Two-level Multi-surrogate Assisted Optimization; TLBO, Teaching Learning-Based Optimization.

* Corresponding author at: State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University, Changsha 410082, PR China. Fax: +86 731 88821807.
E-mail address: wanghu@hnu.edu.cn (H. Wang).

http://dx.doi.org/10.1016/j.asoc.2016.04.035
1568-4946/

to lead the surface refinement to a smaller design space [12]. Wujek and Renaud compared a number of move-limit strategies which focused on controlling function approximation in a more "meaningful" design space [13,14]. Alexandrov et al. advocated the use of a sequential modeling approach using Trust Region Method (TRM) [15]. Rodríguez employed trust region augment Lagrangian method for Sequential Response Surface Method (SRSM) [16]. Wang et al. developed an Adaptive RSM (ARSM), which systematically reduced the size of design space [17]. Wang et al. proposed a Boundary and Best Neighbor Sampling (BBNS)-based ARSM, which is integrated with the MLS approximation for nonlinear problems [18]. Another sub-branch of SAOs is Surrogate Assisted Evolutionary Optimization (SAEO). Studies on SAEOs began over two decades ago and have received considerably increasing interest in recent years. For the SAEOs, surrogates are commonly integrated with Evolutionary Algorithms (EAs), such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Teaching Learning-Based Optimization (TLBO) [19–21]. Theoretically, the surrogate can be applied to almost all operations of the EA, such as population initialization, cross-over, mutation and local search and fitness evaluations [22]. The SAEO commonly uses surrogates in the local search for both single and multi-objective optimization methods [23,24]. In such cases, sophisticated model management methods developed in traditional design optimization, such as TRM can be employed directly [25]. The surrogate has also been used in stochastic search methods other than EAs, such as surrogate assisted Simulated Annealing (SA) [26] or Artificial Immune (AI) systems [27]. Although SAOs have achieved good results for various disciplines, most of high dimensional nonlinear problems still cannot be handled, especially for costly simulation evaluations. According to Wang and Shan [28] and Chen et al.s' [29] suggestions, 10 or more dimensional problems can be considered high if the corresponding evaluation is time consuming, and such problems widely exist in various disciplines.

For a practical industrial design application, if all input variables are independent, each parameter can be designed individually. Ideally, under such circumstance, a complicated case can be decomposed easily. Practically, for most of the problems, design parameters are commonly correlated. However, according to application experiences, some of them are weak correlated. If the correlations among input variables can be identified, a high dimensional problem might be decomposed into a combination of some low or medium problems and can be solved easier than the original one. The HDMR is a particular family of representations where each term reflects the independent and cooperative contributions of inputs upon the output. The HDMR was elaborated by Rabitz et al. [30]. Similar to Taylor expansion, a HDMR expansion expresses a high dimensional function as a finite hierarchical correlated function expansion in terms of inputs, which can efficiently reduce sampling effort for learning the behavior of a high dimensional system and automatically identify the correlated relationship among input variables. Recently, Shorter et al. utilized the HDMR to build an efficient chemical kinetics solver [31]. Li et al. proposed a HDMR-based random sampling method and developed an approach to approximate its different component functions [32]. Shan and Wang integrated RBF with cut-HDMR for high dimensional expensive black-box problems [33]. Wang et al. utilized the MLS as a basis function for cut-HDMR and applied to high dimensional problems [34]. However, the HDMR is still rather a modeling technique than an optimizer. If a practical case is optimized by HDMRs, the computational cost is still expensive.

To improve the performance of HDMRs and make them feasible for optimizer, a Two-level Multi-surrogate Assisted Optimization (TMAO) method is suggested. Compared with other HDMRs, the HDMR is only a decomposer for the original problem in the TMAO. Even for a non-separable problem, it still can be solved by other strategies in such framework. Generally, in the first-level, the HDMR is utilized to decompose the original input variables by considering noises and outliers. In this level, low correlated term (two-order) is considered. In the second-level, an ensemble sampling strategy is suggested for uncoupled variables based sub-problems. Therefore, the sub-solution can be obtained based on each decomposed sub-problem. To verify the feasibility and performance of the TMAO, several test functions are employed.

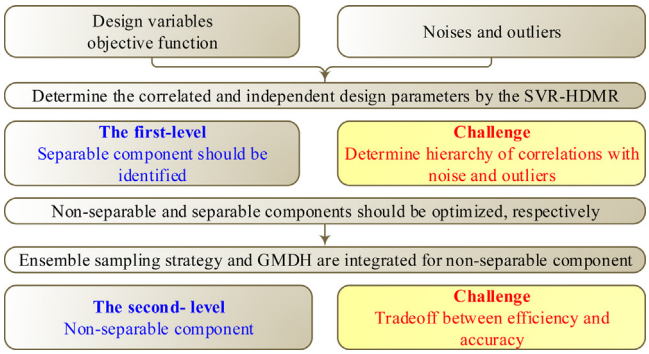


Fig. 1. The framework of TMAO and corresponding challenges.

2. Brief view of framework of TMAO and challenges

Mathematically, a high dimensional problem can be decomposed into two components, separable and non-separable components. For a separable component f_{sp}, it can be expressed as

$$\left\{ arg\left(\min_{x_1} f(x_1, \ldots)\right), \ldots, arg\left(\min_{x_n} f(\ldots, x_n)\right)\right\} = arg\left(\min_{x_1,\ldots,x_n} f(x_1 \ldots, x_n)\right). \quad (1)$$

It means that f_{sp} can be decomposed into n functions independently, and each function can be modeled or optimized individually. In other words, input variables of f_{sp} function are independent and can be expressed as

$$f_{sp}(x_1, \ldots, x_n) = \sum_{i=1}^{n} f_i(x_i). \quad (2)$$

If Eq. (2) is satisfied, it suggests that all input variables are independent and can be designed independently. Generally, the HDMR can help the designer to decompose a high dimensional function into several combinations of low or medium dimensional functions. Thus, the physical essential of an underlying system can be disclosed partly. Such a decomposed problem is easy to be solved compared with the original one. Therefore, the purpose of the first-level is reduction of dimensionality as shown in Fig. 1.

As shown in Fig. 1, the cut-HDMR is utilized to determine the correlations of input variables $x_i x_j$. Theoretically, all hierarchy of correlations can be identified by the cut-HDMR, such as $x_i x_j x_k$, $x_i x_j x_k \ldots x_z$. Obviously, the number of samples should be dramatically increased correspondingly. Fortunately, most of the practical problems are commonly low-order or weak high-order correlated, only $x_i x_j$ correlated terms are considered in most of HDMRs. The assumption is also used in the first-level of the TMAO. Compared with other HDMR-based optimization method, the errors due to loss of high-order correlated term can be compensated in the second level and should be discussed later.

For any function which is a combination of two input variables $f(x_i, x_j)$, if it satisfies following equation

$$f(x_i + \Delta x_i, x_j + \Delta x_j) = f(x_i + \Delta x_i, x_j) + f(x_i, x_j + \Delta x_j) - f(x_i, x_j). \quad (3)$$

If input variables x_i, x_j are independent of each other, it means the correlative term of x_i, x_j must be zero. Sequentially, when $f(x_i, x_j)$ is filtered, $f(x_i, x_j, x_k)$ should be filtered correspondingly. However, it should be noted that it is difficult to strictly satisfy the criterion as Eq. (3). Practically, if an input variable is weak correlated with others, it should be assumed to be independent and can be decoupled. Moreover, for most of HDMRs, noises and outliers are commonly not considered while modeling. Under such circumstance, some characteristics of a physical model cannot be disclosed, separable and non-separable components might not be recognized. Therefore, in TMAO, noises and outliers should be considered in the first-level while modeling and should be discussed in Section 3.

The second-level is utilized to handle the non-separable part of the original problem. The major concern is the way to construct reliable and feasible sampling strategy for the global solution. In this level, Group Method of Data Handling (GMDH) [35] method is utilized due to its property of self-organized. Considering robustness of modeling, an ensemble sampling strategy is suggested. Unlike ensemble surrogate method suggested by Goel et al. [36], each modeling technique builds the surrogate independently. Therefore, multiple optima and error distributions should be obtained. It suggests that more characteristics of a physical essential might be captured and the corresponding surrogate might be more reliable. The details of the ensemble sampling strategy are presented in Section 4.

3. Robust high-dimensional model representation

According to hierarchy framework of the TMAO as shown in Fig. 1, the key issue is to decompose separable and non-separable components. Therefore, the HDMR is a crucial issue. If the HDMR determines the correlations among input variables accurately, the difficulty of the original problem should be reduced significantly.

3.1. Basic theories of HDMR

The HDMR is a general set of quantitative model assessment and analysis tools for recognizing the high dimensional relationships between input variables and corresponding outputs. Let the n-dimensional vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T \in R^n$ with n which denotes the number of input variables of an underlying problem, and output function $f(\mathbf{x})$ can be expressed as a finite expansion in terms of input variables as

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1<i<j\le n} f_{ij}(x_i, x_j) + \sum_{1<i<j<k\le n} f_{ijk}(x_i, x_j, x_k) + \ldots + \sum_{1<i_1<\ldots<i_l\le n} f_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) + \ldots + f_{12\ldots n}(x_1, x_2, \ldots, x_n) \quad (4)$$

where f_0 denotes a constant term representing the zero-th order effect to $f(\mathbf{x})$. The function $f_i(x_i)$ is the first-order term expressing the effect of variable x_i acting alone. The function $f_{ij}(x_i, x_j)$ is the second-order term that describes the cooperative effect of variables x_i and x_j upon $f(\mathbf{x})$. The high-order terms give the cooperative effects of increasing numbers of input variables acting together to $f(\mathbf{x})$. The last term $f_{12\ldots n}(x_1, x_2, \ldots, x_n)$ contains any residual dependence of all input variables correlated together in a cooperative way to $f(\mathbf{x})$. As mentioned in Section 1, high-order terms in Eq. (4) should be negligible such that the HDMR with only low order correlations to the second-order among input variables as

$$f(\mathbf{x}) \approx f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1<i<j\le n} f_{ij}(x_i, x_j). \quad (5)$$

Furthermore, according to Shan and Wang's [33] suggestions, it often provides a satisfactory description of $f(\mathbf{x})$ for many high dimensional cases. Practically, most of the HDMRs often only consider low order correlative terms. For the HDMR expression Eqs. (4) and (5), each component function has distinctive mathematical meaning. At each new level of a HDMR model, high-order correlated effects of input variables should be introduced. While there is no correlated term between input variables, only the constant component f_0 and function terms $f_i(x_i)$ will exist in the HDMR expansion. Moreover, it can be proved that $f_0 = f(\mathbf{x}_0)$ is a constant term of Taylor series; the first-order function $f_i(x_i)$ is the sum of all Taylor series terms with input variable x_i, while the second-order function $f_{ij}(x_i, x_j)$ is the sum of all Taylor series terms variables x_i and x_j, and so on. There are two particular HDMR expansions: ANOVA-HDMR (ANalysis Of VAriance-HDMR) and cut-HDMR. The cut-HDMR expansion is an exact representation of $f(\mathbf{x})$ in the hyperplane passing through a cut point in design space. In this study, the cut-HDMR procedure is used to disclose the second-order correlations.

Using the cut-HDMR, a cut sample $\mathbf{x}^c$ is defined and the cut-HDMR is invariant to the choice of cut point $\mathbf{x}^c$. In practice, $\mathbf{x}^c$ can be the center point of design space as

$$\mathbf{x}^c = \left\{ \frac{\mathbf{x}^{\text{Upper}} - \mathbf{x}^{\text{Lower}}}{2} \right\} \quad (6)$$

where $\mathbf{x}^{\text{Upper}}$ and $\mathbf{x}^{\text{Lower}}$ denote upper and lower bounds of each input variable, respectively. The output function is determined by evaluating input variables/output function of the system relative to the defined cut point $\mathbf{x}^c$ along associated lines, planes and hypercubes, etc. This procedure reduces to the optimal function of the cut-HDMR in Eq. (4) as following form.

$$f_0 = f(\mathbf{x}^c), \quad (7)$$

$$f_i(x_i) = f(x_i, \mathbf{x}_i^c) - f_0, \quad (8)$$

$$f_{ij}(x_i, x_j) = f(x_i, x_j, \mathbf{x}_{ij}^c) - f_i(x_i) - f_j(x_j) - f_0, \quad (9)$$

where $\mathbf{x}_i^c$, $\mathbf{x}_{ij}^c$, are respectively $\mathbf{x}$ without terms x_i and $\left(x_i, x_j\right)$. The first-order term $f_i(x_i)$ is evaluated along its variable axis through the cut point. The second-order term $f_{ij}(x_i, x_j)$ is evaluated in a plane defined by the binary set of input variables (x_i, x_j) through the cut sample. The procedure of subtracting low-order expansion functions removes their dependence to assure a unique contribution from the new function.

If each input variable takes s values, the required model runs to construct $f_i(x_i), f_{ij}(x_i, x_j), \ldots$, which can be obtained by

$$1 + n(s-1) + \frac{n(n-1)(s-1)^2}{2} + \ldots, \quad (10)$$

which grows only polynomially with n and s. As only low-order component functions of cut-HDMR are used, the sample savings for large n are significant compared to traditional s^n exponential sampling. Thus, the cut-HDMR renders the original exponential difficulty to a problem of polynomial complexity.

3.2. EI criterion assisted modeling strategy

For most of the cut-HDMRs, the most popular way is to generate samples randomly. To reduce the number samples, an adaptive strategy [37] was proposed. However, noises and outliers are not

considered in their work. It might lead to misleading estimates of the underlying problem. To enhance the reliability of HDMR model, an EI Criterion Assisted Modeling (EICAM) strategy for cut-HDMR is suggested especially for the cut-HDMR. Compared with other HDMRs, the exploration and exploitation are both considered during sampling, the efficiency of modeling cut-HDMR can be improved. Moreover, the SVR considering empirical and structural risks is employed to be the basis for the cut-HDMR. Thus, the reliability of cut-HDMR can be enhanced. Details of the EICAM are described as follows.

1 Select center point (cut sample) $\mathbf{x}^c$ as the initial sample;
2 Generate two quarter sample set $\mathbf{x}^q$ on the each variable axial and evaluated corresponding response values, $\mathbf{x}^q$ is defined as

$$\mathbf{x}^q = \left\{ \frac{\mathbf{x}^{\text{Upper}} - \mathbf{x}^c}{2}, \frac{\mathbf{x}^c - \mathbf{x}^{\text{Lower}}}{2} \right\} \tag{11}$$

3 Construct the SVR surrogate $\bar{f}(x_i)$ with a single input variable, x_i denotes the ith design variable. It suggests that n surrogates should be built. In this procedure, the SVR is employed because both empirical and structural risks are considered as mentioned before. Therefore, the effects from noises and outliers can be reduced.
4 Locate maximum EI (EImax) samples based on the present n constructed surrogates. It means that n new samples should be generated in this step.

For popular SAOs, such as EGO [11], the probability of improvement indicates how the magnitude of improvement may be. It suggests that some potential improvement region may be located and EI can be expressed as

$$E(I(\mathbf{x})) = \begin{cases} (f_{\min} - \bar{y}(\mathbf{x}))\Phi\left(\frac{f_{\min} - \bar{y}(\mathbf{x})}{\bar{\sigma}(\mathbf{x})}\right) + \bar{\sigma}\phi\left(\frac{f_{\min} - \bar{y}(\mathbf{x})}{\bar{\sigma}(\mathbf{x})}\right) & \text{if} \bar{\sigma} > 0 \\ if & \text{if} \bar{\sigma} = 0 \end{cases}, \tag{12}$$

where I is a measure of improvement, $\bar{\sigma}(\mathbf{x})$ is the predicted variance in the prediction of Gaussian process based model, $\Phi(\bullet)$ and $\phi(\bullet)$ are the Cumulative Density Function (CDF) and the Probability Density Function (PDF) of a normal distribution, respectively. Unfortunately, with the multiple updated new samples, it is difficult to predict the EI because it is actually a sequential process. Theoretically, for k updates the samples $X^{(m+1)}, \ldots, X^{(m+k)}$ can be obtained by maximizing

$$E_{Y_1}[I_1(\mathbf{x})], \ldots, E_{Y_k}[I_k(\mathbf{x})]. \tag{13}$$

When $k > 2$, it is mathematically unrealistic without using statistical estimate techniques such as Monte Carlo Simulation (MCS) commented by Ginsbourger et al. [38]. Therefore, sequential design suggested by Schonlau [39] is used in this study given as

$$E^{n+i}(I(\mathbf{x})) = \begin{cases} \bar{\sigma}^{n+i}(\mathbf{x}) \left[\frac{(f^m_{\min} - \bar{y}^n(\mathbf{x}))}{\bar{\sigma}^n(\mathbf{x})} \Phi\left(\frac{f^m_{\min} - \bar{y}^n(\mathbf{x})}{\bar{\sigma}^n(\mathbf{x})}\right) + \bar{\sigma}\phi\left(\frac{f^m_{\min} - \bar{y}^n(\mathbf{x})}{\bar{\sigma}^n(\mathbf{x})}\right) \right] & \text{if } \bar{\sigma}^{(m+i)} > 0 \\ 0 & \text{if} \bar{\sigma}^{(m+i)} = 0 \end{cases} \tag{14}$$

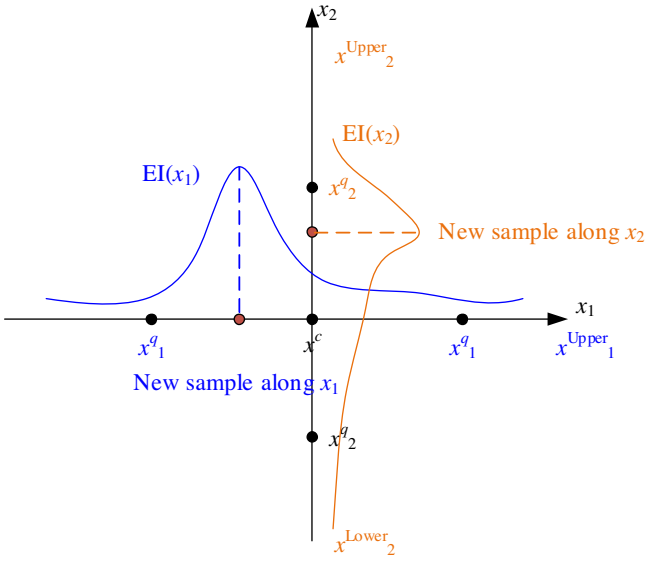


Fig. 2. Pattern of sampling strategy for the 2D problem.

where $\mathbf{x}^{m+i}$ in $\bar{\sigma}^{m+i}(\mathbf{x})$ is the estimated error of prediction at $\mathbf{x}^{m+i}$. The estimate depends on $\mathbf{x}^{m+i}$ and updates correlation parameters, not the unknown output y^{m+i}. It permits multiple regions for updated samples to be identified before any expensive evaluations are carried out. Fortunately, the SVR model is only one-dimension and can be well approximated by using EI criterion even if highly nonlinear problems. Therefore, the Number of Function Evaluations (NFEs) can be well controlled. The pattern of sampling strategy for the 2D problem is demonstrated in Fig. 2.

5 Check the accuracy of surrogate by using R^2. If $R^2 > 0.9$, the procedure goes to Step 6, else the new samples should be generated by the updated surrogate and the procedure goes back to Step 4. It should be noted that the new sample should be evaluated by the real function in this step. In order to predict the accuracy without adding new samples, Leave-One-Out Cross Validation (LOOCV) is used to obtain R^2.
6 Generate test samples. Because the second-order correlative term is considered, when a test sample $\mathbf{x}^q_i$ is generated, the difference between $\mathbf{x}^q$ and $\mathbf{x}^q_i$ on the variable axial can be obtained. The correlation criterion can be rewritten as

$$\begin{cases} R_{\text{avg}} = \dfrac{\sum_{k=1}^{M} R^k_{ij}}{M}, R_{\text{dev}} = \dfrac{\sum_{k=1}^{M} \left| \dfrac{R^k_{ij} - R_{\text{avg}}}{R_{\text{avg}}} \right|}{M} \\ \text{where } R^k_{ij} = \left| \dfrac{f(x_i + \Delta x^k_i, x_j + \Delta x^k_j) - \left(\bar{f}(x_i + \Delta x^k_i, x_j) + \bar{f}(x_i, x_j + \Delta x^k_j) - f(x_i, x_j) \right)}{f(x_i, x_j)} \right| \end{cases}, \tag{15}$$

where R_{avg} and R_{dev} denote mean and Standard Deviation (STD) of residual values, respectively, R^k_{ij} denotes residual value with ith and jth design parameters and M is the Number of Test Samples (NTS). If $R_{\text{avg}} < 0.05$ and $R_{\text{dev}} < 0.05$, x_i and x_j can be regarded as a weak correlation pair. In Eq. (15), $f(x_i + \Delta x^k_i, x_j + \Delta x^k_j)$ and $f(x_i, x_j)$ should be evaluated by real function evaluations. Moreover, these

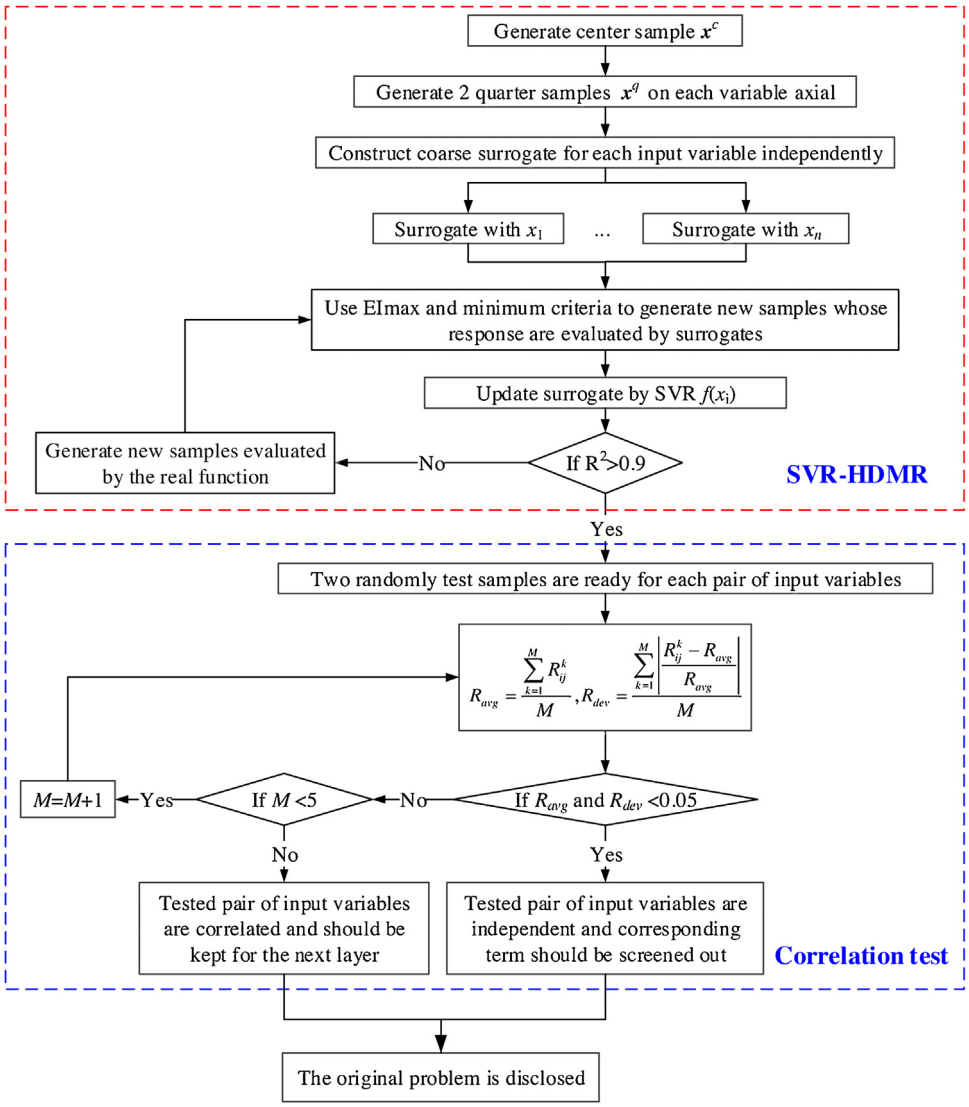


Fig. 3. Flowchart of the EICAM strategy.

test samples can be reused for constructing the initial surrogate in the second-level. Considering computational cost, the NTS needs to be well controlled. The EICAM commonly uses 2 samples to test correlations among input variables. If correlation criteria are not satisfied, one more sample should be generated. However, if more samples are used, the efficiency of the EICAM should be decreased dramatically. Although the threshold of NTS is difficult to be determined, according to tests for less than 50 dimensional problems, the NTS of correlations for each pair of input variables is suggested to be less than 5. To clearly describe the suggested algorithm, the flowchart of EICAM is demonstrated in Fig. 3.

3.3. Separable function test

To verify the performance of the proposed strategies, three different dimensional ($10d, 20d, 30d$) nonlinear functions (Eq. (16)) are tested. To test the reliability of the suggested method, noise level is set to 5% of each response. For each case, the maximum NFEs is set to $(3(d+1)(2d+1))/2$. To obtain reprehensible results, the NTS is set to $100d$ and test samples are generated randomly. In this test, SVR-HDMR, MLS-HDMR, Kriging-HDMR and RBF-HDMR modeling methods are tested. In addition, two sampling methods, Golden Section Sampling (GSS) and EICAM strategies are employed. For each function, to obtain representative test results, 50 replications are performed and the initial cut sample is generated randomly. Compared with the real function evaluation, the computational cost of a surrogate assisted evaluation is far less. Therefore, the number of surrogate evaluations is not presented. Means and Standard Deviations (STDs) values of R^2, Relative Average Absolute Error (RAAE) and Relative Maximum Absolute Error (RMAE) are used to evaluate the accuracy of the surrogate with noises. R^2 and RAAE indicate the overall accuracy of a surrogate over the entire design space. A high RMAE value indicates large error in a region of the design space.

$$f(x) = \left[\sum_{i=1}^{d} i^3 (x_i - 2)^2 \right]^3 ; x_i \in [-3, 3], d = 10, 20, 30 \quad (16)$$

According to results listed in Table 1, approximation methods, MLS and SVR obviously outperform interpolation methods, Kriging and RBF-based HDMR by considering noises. The SVR-HDMR with EICAM outperforms most of the others in terms of mean and STD of all criteria. With the increase of dimension, the accuracy of each surrogate obviously decreases. However, for $30d$ function, mean of R^2 of the surrogate constructed by the suggested strategy is still more than 0.8; RAAE and RMAE are less than 7.0. Therefore, the SVR-

Table 1
Comparisons among all benchmark algorithms for Eq. (16).

(1) R^2 Modeling method	Sampling strategy	Mean of R^2	STD of R^2	Mean of R^2	STD of R^2	Mean of R^2	STD of R^2
Dimension		10*d*		20*d*		30*d*	
SVR-HDMR	GSS	0.856	0.108	0.754	0.120	0.665	0.122
SVR-HDMR	EICAM	0.934	0.065	0.897	0.107	0.820	0.076
MLS-HDMR	GSS	0.756	0.124	0.681	0.210	0.610	0.136
MLS-HDMR	EICAM	0.864	0.087	0.815	0.098	0.789	0.101
Kriging-HDMR	GSS	0.602	0.112	0.561	0.291	0.302	0.223
Kriging-HDMR	EICAM	0.771	0.087	0.623	0.156	0.412	0.178
RBF-HDMR	GSS	0.587	0.221	0.501	0.264	0.342	0.188
RBF-HDMR	EICAM	0.754	0.108	0.612	0.201	0.436	0.156

(2) RAAE Modeling method	Sampling strategy	Mean of RAAE	STD of RAAE	Mean of RAAE	STD of RAAE	Mean of RAAE	STD of RAAE
Dimension		10*d*		20*d*		30*d*	
SVR-HDMR	GSS	0.812	0.143	1.076	0.167	1.543	0.300
SVR-HDMR	EICAM	0.312	0.034	0.342	0.078	0.487	0.089
MLS-HDMR	GSS	0.892	0.234	1.132	0.211	1.891	0.321
MLS-HDMR	EICAM	0.382	0.091	0.396	0.087	0.543	0.101
Kriging-HDMR	GSS	1.091	0.187	1.291	0.239	3.123	0.872
Kriging-HDMR	EICAM	0.391	0.087	0.501	0.110	0.619	0.131
RBF-HDMR	GSS	1.483	0.493	2.103	0.781	3.897	1.071
RBF-HDMR	EICAM	0.401	0.111	0.551	0.138	0.701	0.138

(3) RMAE Modeling method	Sampling strategy	Mean of RMAE	STD of RMAE	Mean of RMAE	STD of RMAE	Mean of RMAE	STD of RMAE
Dimension		10*d*		20*d*		30*d*	
SVR-HDMR	GSS	4.234	0.893	5.512	1.451	5.973	1.092
SVR-HDMR	EICAM	2.732	0.463	2.892	0.662	2.487	0.551
MLS-HDMR	GSS	4.877	1.292	6.239	1.881	6.221	1.223
MLS-HDMR	EICAM	3.182	0.532	3.102	1.012	2.812	0.723
Kriging-HDMR	GSS	6.399	2.231	8.291	2.121	8.123	1.456
Kriging-HDMR	EICAM	4.293	1.232	5.123	0.872	3.612	1.154
RBF-HDMR	GSS	20.483	3.457	21.133	3.495	23.121	2.987
RBF-HDMR	EICAM	6.341	1.983	5.251	1.056	4.722	1.138

HDMR integrated with EICAM is utilized to the suggested TMAO method.

4. A multi-surrogate assisted sampling GMDH strategy

4.1. Basic theories of GMDH

In the second-level, any high performance modeling techniques can be employed for the rest of non-separable terms theoretically. In this work, a self-organized model method, GMDH is utilized. The GMDH is based on sorting-out of gradually complicated models and the evaluation of them by external criterion on a separate part of data sample [40,41]. Compared with other methods, the distinctive feature of GMDH is to measure of selected variables significance by itself. The GMDH increases the number of partial model components and finds a model structure with optimal complexity indicated by the minimum value of an external criterion. This process is called self-organization of the model. This inductive strategy is different from commonly used deductive technique. Most of GMDHs use the polynomial reference functions which can be expressed by Volterra functional series [42] as:

$$y = a_0 + \sum_{i=1}^{N} a_i x_i + \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij} x_i x_j + \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} a_{ijk} x_i x_j x_k, \quad (17)$$

where $\mathbf{x}(x_1, x_2, \ldots, x_N)$ is a input variable vector; $\mathbf{a}(a_1, a_2, \ldots, a_N)$ is a vector of coefficients or weights. Components of an input variable vector x can be independent variables, functional forms or finite difference terms. It allows finding simultaneously the structure of model and the dependence of modeled system output on the values of the most significant inputs of the system.

Let the original data set consists of a column of the observed values of output function **y** and N columns of the values of independent input variables. A primeval equation forms an expression which comes in the form of a quadratic regression polynomial as

$$z = A + Bu + Cv + Duv + Eu^2 + Fv^2, \quad (18)$$

where A, B, C, D, E, and F are parameters of a model, u; v are pairs of variables standing in **x** whereas z is the best fit of the dependent variable **y**.

Details of the GMDH modeling procedure can be briefly described as

1 In the first step, an estimate of **y** is determined by primeval equations. Here, u and v are taken out of all independent input variables $\mathbf{x}(x_1, x_2, \ldots, x_m)$. In this way, the total number of polynomials constructed by Eq. (17) is equal to $m(m-1)/2$. The resulting columns zn of values $s = 1, 2, 3, \ldots, m(m-1)/2$, contain estimates of **y** resulting from each polynomial interpreted as new variables. It may exhibit a higher predictive power than the original variables being just the input variables of a system, $\mathbf{x}(x_1, x_2, \ldots, x_m)$.
2 The second step is to determine the best of new variables and eliminate those which are the worst ones. Several specific selection criteria can be used for this selection. All of them are based on some performance indices (such as R^2, RAAE, RMAE) which express how the values zn follow the output **y**. This step returns a list of design variables. Columns of $\mathbf{x}(x_1, x_2, \ldots, x_m)$ are replaced by the retained columns of $\mathbf{z}(z_1, z_2, \ldots, z_k)$, where k denotes the

total number of the retained columns. If Step 2 is completed within the generation of the present layer, the iteration of next iteration begins immediately by repeating Step 1, otherwise procedure goes to Step 3.

3 This step consists of testing whether the set of equations of the model can be further improved. The lowest value of the selection criterion obtained in this iteration is compared with the smallest value obtained at the previous one. If an improvement is achieved, the procedure goes back to Step.1, otherwise the procedure terminates and a surrogate has been completed.

The GMDH is close in structure to self-organizing algorithm of multilayer modeling system. The structure of the GMDH model becomes closer to the real physical model as samples grow. The main drawbacks of the GMDH can be described as

1 Firstly, the GMDH tends to generate quite complex polynomial for relatively simple systems;
2 Secondly, owing to its limited generic structure (quadratic two-variable polynomial), the GMDH tends to produce an overly complex model when it comes to a highly nonlinear system;
3 Thirdly, it is difficult to select combinations in each cycle. Accuracy for each combination actually might be improved further with more samples. For the GMDH, the training and testing samples are required before construction.

According to these drawbacks, following strategies are employed for overcoming.

Compared with regular GMDH methods, two important issues should be considered. As mentioned in previous section, the cut-HDMR is utilized to determine the separable and non-separable components in the first-level of the framework. Some separable design variables have been screened out. Therefore, the final expression of the GMDH should be simplified significantly. Secondly, since it is difficult for a surrogate to build a highly nonlinear model, the surrogate model should only be the guide for achieving a global optimum efficiently. Reliability of surrogate is more important than other criteria. Therefore, multi-surrogate sampling strategy is suggested to construct more reliable model.

4.2. Multi-surrogate Assisted Sampling algorithm (MAS)

It is well known that different surrogates have been shown to perform well in different criteria. Although various surrogates have been tested by several scholars, it is difficult to claim which one has better performance. Briefly, it is a “case by case” problem. Kammer and Alvin presented a multi-surrogate method for combining a set of alternative surrogate models into a single master response surface called as Super-Metamodel that is valid over all of design space [43]. Zerpa et al. presented one application of using an ensemble of surrogates to construct a weighted average surrogate model for the optimization of alkali–surfactant–polymer flooding process [44]. They indicated that the weighted average surrogate model has better modeling capability compared to individual surrogates. Stone presented a multi-surrogates assisted memetic algorithm for solving optimization problems with computationally expensive fitness functions [45]. Goel et al. explored the possibility of using the best surrogate or a weighted average surrogate model instead of individual surrogate models [36]. They found the use of an ensemble of surrogates may prove a robust approximation method.

For most of the ensemble surrogate modeling methods, a weighted average surrogate model is commonly expressed as follows

$$\bar{y}_{\text{wt.avg.}}(\mathbf{x}) = \sum_{i}^{N_{\text{SM}}} \omega_i(\mathbf{x}) \bar{y}_i(\mathbf{x}) \tag{19}$$

where $\bar{y}_{\text{wt.avg.}}(\mathbf{x})$ is the predicated objective function by the weighted average of surrogate models, $\bar{y}_i(\mathbf{x})$ is the predicted response by the ith surrogate model, $\omega_i(\mathbf{x})$ is the weight associated with the ith surrogate model with design point x, and N_{SM} denotes the number of surrogate models. The sum of weights must be one $\sum_{i}^{N_{\text{SM}}} \omega_i(\mathbf{x}) = 1$. For each ensemble surrogate, weight functions can be determined by different ways, such as data-based error measure, best PRESS (Prediction Residual Error Sum of Squares) and others.

According to theories of ensemble surrogate, the most critical issue is to determine weight coefficients. Generally, whatever criterion it uses, it is commonly determined by the distribution of samples. For different sampling strategies, weight coefficients and corresponding performances are diverse. Moreover, for some multimodal cases, such as Schwefel function, it is difficult to build an accurate surrogate. According to the test results obtained by Li et al. [32], the performance of ensemble surrogate actually did not get a significantly improvement. Since it is difficult to build a highly accurate global surrogate and the performance is determined by the sampling strategy, a multi-surrogate assisted sampling strategy might be suitable than a global ensemble surrogate in term of performance.

Generally, the surrogate is constructed by the GMDH structure and the samples are generated by the MAS. To guarantee quality of a surrogate, the surrogate should be tested in each iteration. Moreover, considering the efficiency of optimization, the NTS should be controlled. Therefore, the LOOCV is suggested to be the accuracy criterion without adding new test samples.

Details of the MAS method are described as follows.

Stage 1. The SVR is chosen as basis function for the GMDH.

Commonly, the PR is used for the GMDH. Although the PR can filter noises and outliers, the PR is difficult to handle nonlinear functions. Similar to the basis of HDMR, the SVR is also integrated with the GMDH. Briefly, empirical and structural risks should considered in the SVR-GMDH.

Stage 2. If

$$\left| \frac{\mathbf{x}^i - \mathbf{x}^{i-1}}{\mathbf{x}^{i-1}} \right| \le \gamma \tag{20}$$

is satisfied in 5 successive iterations, the procedure terminates, else it goes to Stage 3. In Eq. (20) $\mathbf{x}^i$ and $\mathbf{x}^{i-1}$ denote the solutions in the present and the last iterations, respectively.

Stage 3. PR, Kriging, RBF, MLS and SVR are employed to construct surrogates based on the present sample set. Theoretically, any surrogates can be employed. However, the pre-test demonstrates that the quality of the surrogate would not be improved significantly even if more surrogate modeling methods are employed. In this stage, each method constructs a surrogate individually. The EImax sample of each surrogate can be obtained by importing the Kriging. Sequentially, absolute errors of each sample based on different surrogates should be obtained by the LOOCV. The sum of error estimation E_j for each surrogate can be defined as

$$E_j = \frac{\sum_{i=1}^{n} E(x_j^i)}{n} \tag{21}$$

Table 2
Test functions used in experimental studies.

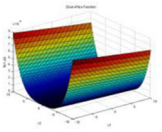
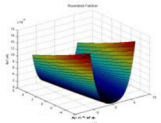
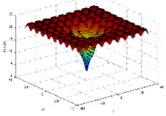
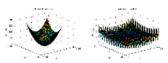
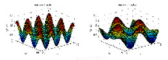
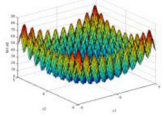
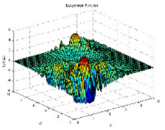
Functions	Objective function	Number of variables	Analytical optimum	Property	2*d* illustration
F1	Dixon-price	20	0	Multimodal with narrow valley	
F2	Dixon-price	30	0	Multimodal with narrow valley	
F3	Dixon-price	50	0	Multimodal with narrow valley	
F4	Rosenbrock	20	0	Multimodal with narrow valley	
F5	Rosenbrock	30	0	Multimodal with narrow valley	
F6	Rosenbrock	50	0	Multimodal with narrow valley	
F7	Ackley	20	0	Multimodal	
F8	Ackley	30	0	Multimodal	
F9	Ackley	50	0	Multimodal	
F10	Griewank	20	0	Multimodal	
F11	Griewank	30	0	Multimodal	
F12	Griewank	50	0	Multimodal	
F13	Rastrigin	20	0	Multimodal	
F14	Rastrigin	30	0	Multimodal	
F15	Rastrigin	50	0	Multimodal	
F16	Langermann	10	−1.4	Complicated multimodal	

Table 3
Statistics of the best solutions obtained by the TMAO limited in 1000 exact function evaluations.

Function	$Best_{1000}$	$Worst_{1000}$	$Mean_{1000}$	STD_{1000}	Number of real or near global optimum
F1	13.45	56.32	18.83	17.28	0
F2	18.54	86.92	34.23	29.56	0
F3	131.21	383.12	282.17	87.15	0
F4	16.14	76.81	24.13	19.32	35
F5	25.71	98.82	51.08	29.15	38
F6	177.73	409.48	232.30	90.07	0
F7	0.0025	1.07	0.15	0.37	50
F8	1.73	3.74	2.89	0.81	46
F9	7.32	12.81	11.04	1.25	0
F10	0.00021	0.20	0.036	0.056	50
F11	0.36	1.39	0.91	0.20	43
F12	23.9	47.98	37.81	12.92	0
F13	0.00031	0.21	0.046	0.067	50
F14	0.23	1.42	0.79	0.65	46
F15	32.65	67.49	37.90	12.41	0
F16	9.55	6.96	8.39	6.92	0

where n is the number of sample points, $E(x_j^i)$ denotes the error of jth sample point based on the ith surrogate. Then, the minimum of E_j is the most accurate surrogate, the EImax and Present Best Sample (PBS) should be selected for updating the SVR-GMDH.

It also should be noted that if a new sample is close to another, this sample should be ignored and the second better sample should be the substitution. Therefore, a distance criterion is defined as

$$\eta = \sqrt{\sum_{j=1}^{m} \left(\frac{x_j^{\text{Best}} - x_j^{\text{Closest}}}{x_j^U - x_j^L} \right)^2} \tag{22}$$

where m denotes the number of design variables, x^{Best} denotes the value of EImax or PBS sample points, x^{Closest} denotes the closet one from EImax or PBS samples. In this study, if $\eta \leqq 0.05$, this sample should be filtered.

5. Numerical examples

5.1. Test functions

For high dimensional problems, Rugh [42] used the GP model with the probability of improvement prescreening as a global surrogate and Lamarckian evolution as local surrogate model to improve the efficiency of evaluation optimization. Emmerich et al. [47] investigated GP-based local surrogate models and different prescreening methods to improve the performance of evaluation optimization. The test functions used in above mentioned two literatures are 20 variables. Emmerich used Ackley function; Zhou

Table 4
Comparisons of TMAO with SAGA-GLS [46] and MAES [47].

Problem ID	$TMAO_{1000}$	$SAGA\text{-}GLS_{1000}$	$MAES_{1000}$
F4	16.14	64.00	NA
F7	0.0025	5.50	3.00
F10	0.00021	2.12	NA

et al. [46] used Ackley, Rosenbrock, Griewank and Rastringin functions. To compare the performance of the TMAO, 16 nonlinearly mathematical functions are tested as shown in Table 2.

5.2. Performance and comparisons with other high dimensional optimizers

For most of the test problems, the NFEs are commonly set to 1000*N*. For practical problems, it is impossible to use such a lot of evaluations. Therefore, in this study, the NFE is set to 1000. Moreover, to obtain the representive results, the statistics of the best function values obtained by the TMAO with 1000 function evaluations on 50 independent runs for F1-F16 are reported in Table 3.

For the cut-HDMR in the first-level, the initial sample should be randomly generated and all sample points should be used for the initial sample points for the GMDH in the second-level.

According to Table 3, for most of the multimodal functions with 20 and 30 variables, the suggested method can find good solutions within 1000 exact function evaluations, such as Ackley, Griewank and Rastrigin functions. For multimodal with narrow valley and complicated multimodal functions, such as Dixon-price, Rosenbrock and Langermann functions, it is difficult to achieve the real global optimums. Even so, the EGO with 10,000 exact evaluations cannot find good solutions for F1, F2, F4, F5 and F16. For 50-variable functions, although the final solutions listed in Table 3 are not very close to the real global optima, their qualities are reasonably better than others. To check whether the TMAO really find the global optimization, the best solutions of them are also investigated. According to Table 3, it can be found that for F4, F5, F7, F8, F10, F11, F13, F14, the solutions obtained by the TAMO are real global optima or are very close to the global optima. For other test functions, the solutions fall into the local optima. For 50 dimensional test functions, the TMAO still cannot find the real global optimum in the limited NFEs.

Moreover, to compare the suggested TMAO in this study, SAGA-GLS (Surrogate-Assisted Evolutionary Algorithm with Global-Search) [46] and MAES (Metamodel Assisted Evolutionary Strategy) for the same functions are listed. SAGA-GLS on F4 (Rosenbrock), F7 (Ackley) and F10 (Griewank), and MAES on F7 are compared in Table 4. The best solutions by TMAO, SAGA-GLS and MAES in 1000 iterations for F4, F7 and F10 are presented. Obviously, the results for the selected test functions are better than the SAGA-GLS and the MAES.

5.3. Performance and comparisons with EGO

To investigate the benefits of the key techniques in the suggested method, the EGO is also compared with the TMAO. For the EGO, considering the dimensionalities of test functions, the number of the initial samples for each function is 10*n* and generated by Latin hyper design. For 20, 30, 50 input variables functions, the number of initial samples are 200, 300 and 500, respectively. Similar to Table 3, the best, the worst and mean values of all optima and STD are also presented as shown in Table 5.

To clarify the comparison, the histograms of mean and STD values of the TMAO and the EGO are also presented in Figs. 4 and 5. It is easy to observe that the convergence rate of the TMAO is far better than the EGO. However, for some cases, such as F1 and F2,

Table 5
Test results by the EGO (initial samples are not involved).

Function	$Best_{1000}$	$Worst_{1000}$	$Mean_{1000}$	STD_{1000}
F1	38.92	56.33	67.67	15.54
F2	95.98	123.92	108.87	23.43
F3	274.56	367.53	381.51	65.45
F4	54.32	89.45	67.87	25.43
F5	183.12	234.81	192.23	37.79
F6	509.15	563.48	522.53	67.23
F7	18.23	26.032	18.15	15.32
F8	20.12	54.74	40.12	23.21
F9	43.12	90.87	78.03	34.67
F10	84.23	128.20	98.08	45.23
F11	252.89	284.33	269.23	18.78
F12	349.43	393.12	357.78	32.16
F13	49.83	74.45	55.94	15.23
F14	89.23	103.23	92.45	22.56
F15	124.45	178.47	149.67	32.23
F16	73.59	98.67	89.15	23.56

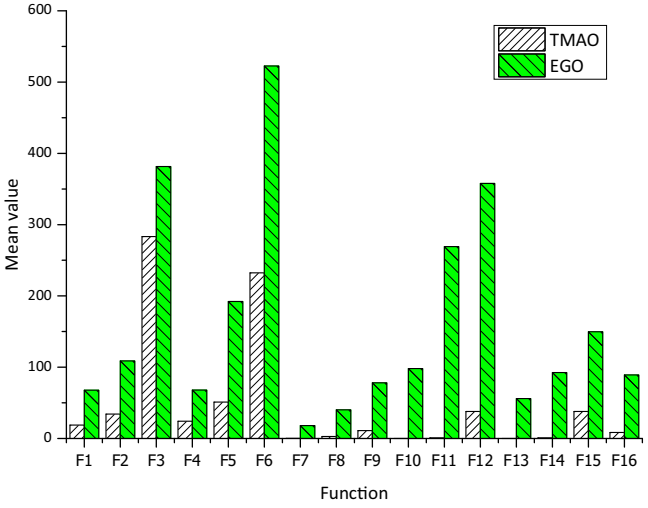


Fig. 4. Comparsions of means of test functions achieved by TMAO and EGO.

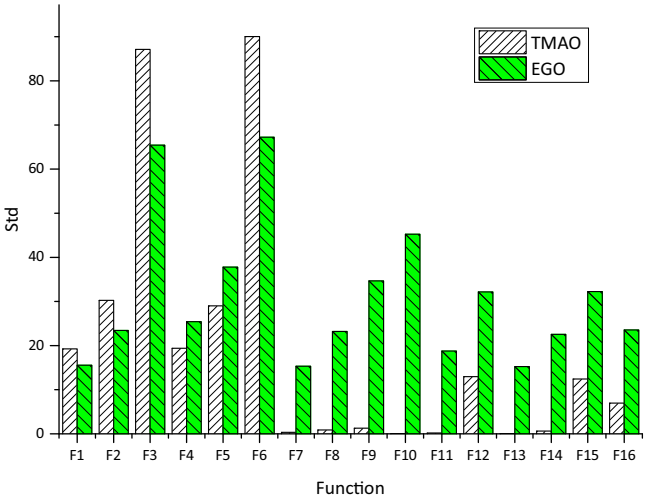


Fig. 5. Comparsions of STDs of test functions achieved by TMAO and EGO.

the EGO is more stable than the TMAO. This is because the EGO is easy to converge to the local optimum, especially for complicated functions. Moreover, to observe convergence trend of the objective functions, the best four convergence curves of F7, F10, F13 functions by the TMAO and EGO are presented in Fig. 6. For all test functions, the EGO cannot find the real global optima. Actually, all solutions

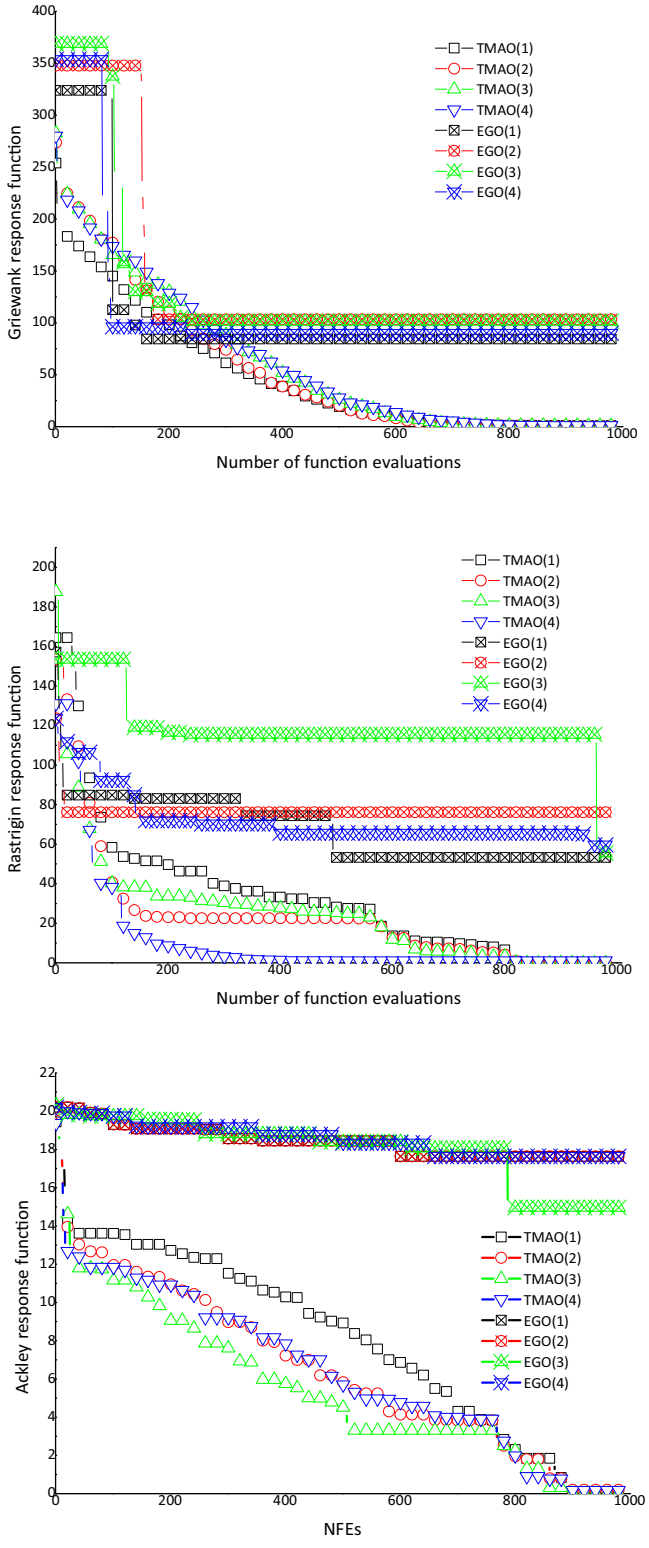


Fig. 6. Convergence curves of objective functions by TMAO and EGO.

are far from the real ones. Moreover, it also can be found that the convergence ratio of the EGO is very slow. Although the EGO can obtain the real global optimum theoretically, the computational cost is very expensive.

6. Conclusions

In this study, an alternative SAO method called TMAO is suggested. The major characteristics of the TMAO are summarized as follows.

TMAO is composed of two layers. In the first-level, the cut-HDMR modeling framework is used to decompose a high dimensional problem into combinations of low or medium dimensional ones. Compared with popular cut-HDMR, the SVR-HDMR is suggested to obtain more reliable surrogate. The test results also show that the SVR-HDMR achieves the best accurate model with noises and outliers compared with other cut-HDMRs. Moreover, to improve the efficiency of modeling, the EI criterion is also integrated. In the second-level, the GMDH framework is used to optimize the correlative input variables. Because some independent or weak correlative input variables might be filtered in the first-level, the GMDH is suitable. Furthermore, to enhance the robustness of modeling without adding more samples, the MAS is suggested. Compared with the MSEGO, the diversity of the sampling distribution can be promised by the LOOCV.

In order to verify the performance of the TMAO, three tests are performed. The first one is to compare the performance of modeling approaches and sampling strategies. Based on this test, we select SVR-HDMR and EI sampling strategy as the best combination in the cut-HDMR framework. The second test is used to compare the performance of the TMAO and other recently developed algorithms, such as SAGA-GLS and MAES. Because we hope use the TMAO to apply to the real-world problems, the NFEs are limited to 1000. According to the results, the advantage of the TMAO is obvious. The final one is the test for the EGO, a well-known SAO method. Compared with the EGO, the efficiency of the TMAO is significantly improved.

However, the TMAO has been investigated by several kinds of nonlinear functions. For some highly nonlinear problems (50d), the real global optimum is still difficult to be obtained in the limited NFEs. Moreover, the suggested TMAO is based on the assumption that the problem can be decomposed. Otherwise, only the second-level is available. Since it is difficult to determine the characteristic of the given problem before starting the TMAO, some samples might be wasted. We also hope to find a more efficient way to disclose the essential of the problem. Unfortunately, we still cannot apply the TMAO to a 50d problem yet because some geometrical parameters cannot be parametrized in practice and CAD and FE models cannot be integrated yet. In our opinion, this is another critical difficulty to prevent the application of the optimization algorithm.

Acknowledgment

This work has been supported by Project of the Program of National Natural Science Foundation of China under the Grant Numbers 11172097, 11302266 and 61232014.

References

[1] R. Myers, D. Montgomery, Response Surface Methodology: Process and Product Optimization using Designed Experiments, Wiley & Sons, NY, 1995.
[2] H.M. Gutmann, A radial basis function method for global optimization, J. Global Optim. 19 (3) (2001) 201–227.
[3] N. Dyn, D. Levin, S. Rippa, Numerical procedures for surface fitting of scattered data by radial basis functions, SIAM J. Sci. Stat. Comput. 7 (2) (1986) 639–659.
[4] N.V. Queipo, R.T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, P.K. Tucker, Surrogate-based analysis and optimization, Prog. Aerosp. Sci. 41 (1) (2005) 1–28.
[5] P. Lancaster, K. Salkauskas, Surfaces generated by moving least squares methods, J. Math. Comput. 37 (1981) 141–155.
[6] P. Breitkopf, H. Naceur, A. Rassineux, P. Villon, Moving least squares response surface approximation: formulation and metal forming applications, Comput. Struct. 83 (17–18) (2005) 1411–1428.

[7] D. Krige, A Statistical Approach to some Mine Evaluations and Allied Problems at the Witwatersrand, M.Sc. Thesis, University of Witwatersrand, Germany, 1951.
[8] Jack P.C. Kleijnen, Kriging metamodeling in simulation: a review, Eur. J. Oper. Res. 192 (3) (2009) 707–716.
[9] S.M. Clarke, J.H. Griebsch, T.W. Simpson, Analysis of support vector regression for approximation of complex engineering analyses, J. Mech. Des. 127 (6) (2005) 1077–1087.
[10] H. Wang, E. Li, G. Li, Probability-based least square support vector regression metamodeling technique for crashworthiness optimization problems, Comput. Mech. Arch. 47 (3) (2011) 251–263.
[11] D.R. Jones, M. Schonlau, W.J. Welch, Efficient global optimization of expensive black-box functions, J. Global Optim. 13 (1998) 455–492.
[12] W. Chen, J.K. Allen, D.P. Schrage, F. Mistree, Statistical experimentation methods for achieving affordable concurrent systems design, AIAA J. 35 (5) (1997) 893–900.
[13] B.A. Wujek, J.E. Renaud, New adaptive move-limit management strategy for approximate optimization, part 1, AIAA J. 36 (10) (1998) 1911–1921.
[14] B.A. Wujek, J.E. Renaud, New adaptive move-limit management strategy for approximate optimization, part 2, AIAA J. 36 (10) (1998) 1922–1934.
[15] N. Alexandrov, J.E.J. Dennis, R.M. Lewis, V. Torczon, A trust region framework for managing the use of approximation models in optimization, Struct. Optim. 15 (1) (1998) 16–23.
[16] J.F. Rodríguez, J.E. Renaud, L.T. Watson, Trust region augmented lagrangian methods for sequential response surface approximation and optimization, J. Mech. Des. 120 (1998) 58–66.
[17] G.G. Wang, Z. Dong, P. Aitchison, Adaptive response surface method—a global optimization scheme for computation-intensive design problems, Eng. Optim. 33 (6) (2001) 707–734.
[18] Hu Wang, Enying Li, G.Y. Li, Z.H. Zhong, Development of metamodeling based optimization system for high nonlinear engineering problems, Adv. Eng. Softw. 39 (8) (2008) 629–645.
[19] R.V. Rao, V.J. Savsani, D.P. Vakharia, Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems, Comput. Des. 43 (3) (2011) 303–315.
[20] R. Rao, V. Patel, An elitist teaching-learning-based optimization algorithm for solving complex constrained optimization problems, Int. J. Ind. Eng. Comput. 3 (4) (2012) 535–560.
[21] R.V. Rao, V.J. Savsani, D.P. Vakharia, Teaching–learning-based optimization: an optimization method for continuous non-linear large scale problem, Inf. Sci. 183 (1) (2012) 1–15.
[22] Jin Yaochu, Surrogate-assisted evolutionary computation: recent advances and future challenges, Swarm Evol. Comput. 1 (2011) 61–70.
[23] Y.S. Ong, P.B. Nair, A.J. Keane, Evolutionary optimization of computationally expensive problems via surrogate modeling, AIAA J. 41 (4) (2013) 687–696.
[24] Z. Zhou, Y.S. Ong, M.H. Lim, B.S. Lee, Memetic algorithm using multi-surrogates for computationally expensive optimization problems, Soft Comput. 11 (10) (2007) 957–971.
[25] J.-F.M. Barthelemy, Approximation concepts for optimum structural design—a review, Struct. Optim. 5 (1993) 129–144.
[26] H.K. Singh, T. Ray, W. Smith, Surrogate assisted simulated annealing (SASA) for constrained multi-objective optimization, IEEE Congr. Evol. Comput. (2010) 1–8.
[27] H.S. Bernardino, H.J.C. Barbosa, L.G. Fonseca, A faster clonal selection algorithm for expensive optimization problems, in: Artificial Immune Systems, Springer, 2010, pp. 130–143.
[28] G.G. Wang, S. Shan, Review of metamodeling techniques in support of engineering design optimization, ASME J. Mech. Des. 129 (2007) 370–389.
[29] V.C.P. Chen, K.-L. Tsui, R.R. Barton, M. Meckesheimer, A review on design, modeling and applications of computer experiments, IIE Trans. 38 (2006) 273–291.
[30] H. Rabitz, F. Ömer, J. Shorter, K. Shimd, Efficient input–output model representations, Comput. Phys. Commun. 117 (1–2) (1999) 11–20.
[31] J.A. Shorter, P.C. Ip, H. Rabitz, An efficient chemical kinetics solver using high dimensional model representation, J. Phys. Chem. A 103 (36) (1999) 7192–7198.
[32] G.Y. Li, C. Rosenthal, H. Rabitz, High dimensional model representations, J. Phys. Chem. A 105 (33) (2001) 7765–7777.
[33] S.Q. Shan, G.G. Wang, Metamodeling for high dimensional simulation-based design problems, J. Mech. Des. 132 (5) (2010) 051009.
[34] H. Wang, L. Tang, G.Y. Li, Adaptive MLS-HDMR metamodeling techniques for high dimensional problems, Expert Syst. Appl. 38 (11) (2011) 14117–14126.
[35] A.G. Ivakhnenko, G.A. Ivakhnenko, The review of problems solvable by algorithms of the group method of data handling (GMDH), Pattern Recognit. Image Anal. 5 (1995) 527–535 (C/C Of Raspoznavaniye Obrazov I Analiz Izobrazhenii).
[36] T. Goel, R.T. Haftka, W. Shyy, N.V. Queipo, Ensemble of surrogates, Struct. Multidiscip. Optim. 33 (2007) 199–216.
[37] E. Li, H. Wang, G. Li, High dimensional model representation (HDMR) coupled intelligent sampling strategy for nonlinear problems, Comput. Phys. Commun. 183 (9) (2012) 1947–1955.
[38] D. Ginsbourger, R.L. Riche, L. Carraro, Kriging is well-suited to parallelize optimization, in: Computational Intelligence in Expensive Optimization Problems, Springer, Berlin, 2010, pp. 131–162.
[39] M. Schonlau, Computer Experiments and Global Optimization, University of Waterloo, 1998.
[40] A.G. Ivakhnenko, G.A. Ivakhnenko, J.A. Muller, Self-organization of neural networks with active neurons, Pattern Recognit. Image Anal. 4 (2) (1994) 185–196.
[41] A.G. Ivakhnenko, G.A. Ivakhnenko, The review of problems solvable by algorithms of the group method of data handling (GMDH), Pattern Recognit. Image Anal. 5 (4) (1995) 527–535.
[42] W.J. Rugh, Nonlinear System Theory: The Volterra–Wiener Approach, Johns Hopkins Univ. Press, Baltimore, 1981.
[43] D.C. Kammer, K.F. Alvin, Component metamodel synthesis for the construction of master response surfaces, Comput. Methods Appl. Mech. Eng. 191 (2002) 4555–4568.
[44] L. Zerpa, N.V. Queipo, S. Pintos, J. Salager, An optimization methodology of alkaline–surfactant–polymer flooding processes using field scale numerical simulation and multiple surrogates, J. Pet. Sci. Eng. 47 (2005) 197–208 (also presented at SPE/DOE 14th Symposium on improved oil recovery, 2004, April 17–21, Tulsa).
[45] M. Stone, An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion, J. R. Stat. Soc. B 38 (1977) 44–47.
[46] Z. Zhou, Y. Ong, P. Nair, A. Keane, K. Lum, Combining global andlocal surrogate models to accelerate evolutionary optimization, IEEE Trans. Syst. Man Cybern. C: Appl. Rev. 37 (1) (2007) 66–76.
[47] M. Emmerich, K. Giannakoglou, B. Naujoks, Single and multiobjective evolutionary optimization assisted by Gaussian random field metamodels, IEEE Trans. Evol. Comput. 10 (4) (2006) 421–439.