Efficient Sparse Representation Response Surface Modelling and Simulation Method for Complex Products

Pu Li[1], Changzheng Huang[1], Songxi Hu[1], Senquan Yang[1], Haoxiang Wen[1] and Lin Zhou[2, *]

[1]School of Physics and Mechanical & Electrical Engineering, Shaoguan University, Shaoguan, China
[2]School of Education, Shaoguan University, Shaoguan, China

*Corresponding author e-mail: zling0126@126.com

Abstract. With the composition and control of modern electromechanical products more and more complex, the time of solving simulation model is longer and longer. The response surface simulation optimization method is an effective method to reduce the simulation time, however, the number of design sampling points of response surface approximation is still large. In this paper, we propose a sparse representation response surface model to accurately reconstruct the source model with small amount sampling points. By means of the sparse representation of the source model on a specific basis, most of the coefficients are zero which can be solved by the equations constructed from small number of sampling. Sparse representation response surface models include sparse response surface and quasi-sparse response surface which are respectively applied to the case where the number of sampling points is greater than and less than the degree of sparse. Sparse response surface runs quicker and quasi-sparse response surface has higher accuracy. Two test functions and one engineering practice problem are employed to compare the performance between sparse representation response surface model and other common response surface models. The results show that the sparse representation response surface model has better performance in approximate accuracy and simulation efficiency.

1. Introduction

With the complex function of modern electromechanical products (such as vehicles, engines, CNC machines and so on), the degree of intelligence is improving, the system composition and control are becoming more and more complicated. Therefore, modelling and simulation technology has been widely used in the design process to improve the overall performance of the product [1].

In the process of products modelling and simulation, simulation models often reflect the multi-disciplinary, non-line and other significant features, which cause a long time to solve the simulation, maybe take as many as 30 to 160 hours [2]. As the product simulation optimization needs to call the simulation process several times, which takes longer time, it is difficult to achieve the objective of product optimal design. Therefore, the efficient simulation of complex products has become an urgent problem in the process of engineering product design.

Because the complex product model has a long computation time and the simulation model may not be a complete explicit mathematical equation, the existing method usually regards the simulation model as a black box function, and uses the optimization algorithm to search globally in the design

space to expect to reduce the call times of black box function [3]. Among those methods, the response surfaces simulation optimized method based on computer experiment design is an effective method [4]. For the source model with unknown shape in advance, the response surface of the approximation target is constructed by random sampling and target estimation in the design space, and the accuracy of the response surface is improved by increasing adaptive sampling points. The following response surface models are mainly used at present[2]: (1) Polynomial Response Surface (PRS), (2)Multivariate Adaptive Regression Splines (MARS), (3) Kriging (Ordinary Kriging (OK) and Blind Kriging (BK)), (4) Radial Basis Functions (RBF) and Extended Radial Basis Functions (ERBF), (5)Support Vector Regression (SVR). By means of iterative adaptive sampling and response surface approximation in the design space, these response surface simulation optimization methods establish the intrinsic function relation between design variable and the objective function value to search the optimal design point quickly and improve the simulation optimization efficiency. However, when the response surface shape is more complex, the number of sampling points used to construct the response surface and the number of black box function's calling time will be correspondingly more.

Compressed sampling theory is a new theoretical framework of information acquisition and processing proposed in the field of signal processing [5; 6]. Based on the sparse representation, the signal (or image) can be acquired and reconstructed with high fidelity through a very small amount of non-adaptive sampling. In the simulation optimization process, the target (or constraint) estimate in the design space is similar with the image and signal, and the compression sampling can be used as a new response surface constructing method which we call sparse representation response surface (SRRS). We first construct a set of function bases, if the source model can be sparsely represented on the set of function basis (coefficients are mostly zeros), then the non-zero coefficients can be solved only by a small number of sampling points, so as to build sparse or quasi-sparse response surface which can reduce the calculation cost and improve the simulation optimization efficiency.

The rest of the paper is organized as follow: in section 2, compressed sampling method is introduced; we introduce the method of SRRS include sparse response surface and quasi-sparse response surface corresponding applied to the number of sampling points is greater than and less than the degree of sparse degree in section 3 and section 4; the effect of SRRS is shown and verified in section 5; at last, the conclusion is given.

2. Compressed sampling method

Consider a *n*-dimension signal $\boldsymbol{x} \in \boldsymbol{R}^{n\times 1}$ expended in a set of orthogonal basis $\{\varphi_i\}_{i=1}^{n}$, which is:

$$\boldsymbol{x} = \sum_{i=1}^{n} \theta_i \varphi_i. \tag{1}$$

Written in matrix form:

$$\boldsymbol{x} = \boldsymbol{\Phi\Theta}, \tag{2}$$

where $\boldsymbol{\Phi} = [\varphi_1, \varphi_2, \cdots, \varphi_n] \in \boldsymbol{R}^{n\times n}$ is the sparse matrix (also known as dictionary), φ_i is the atom, $\boldsymbol{\Theta} = [\theta_1, \theta_2, \cdots \theta_n]^T \in \boldsymbol{R}^{n\times 1}$ is expansion coefficients vector. Assuming that the majority of the elements in the coefficient vector $\boldsymbol{\Theta}$ are zero, or close to zero, that is, the number of nonzero elements $k \ll n$, then it can be said that the signal $\boldsymbol{x}$ can be sparse on the basis $\boldsymbol{\Phi}$. k is called sparse degree which means the number of nonzero elements. Compressed observing the signal with an observed matrix $\boldsymbol{\Psi} \in \boldsymbol{R}^{m\times n}$ which is not related to the dictionary $\boldsymbol{\Phi}$:

$$\boldsymbol{y} = \boldsymbol{\Psi x}. \tag{3}$$

We can get m linear observation responses $\boldsymbol{y} \in \boldsymbol{R}^{m}$. Recovering the signal $\boldsymbol{x}$ from compressed observation response $\boldsymbol{y}$ is a problem of solving linear equations. However, since $m < n$, Eq.(3) is invalid linear equations which has infinite solutions. By combining Eq.(2) and Eq.(3), we obtain:

$$\boldsymbol{y} = \boldsymbol{\Psi x} = \boldsymbol{\Psi\Phi\Theta} = \boldsymbol{A\Theta}. \tag{4}$$

Although Eq.(4) is still invalid linear equations, since coefficient vector $\boldsymbol{\Theta}$ is k sparse, we can recover the coefficients $\boldsymbol{\Theta}$ from compressed observation response $\boldsymbol{y}$ by a mathematical method, then obtain signal $\boldsymbol{x}$ from $\boldsymbol{x} = \boldsymbol{\Phi\Theta}$. Literature [5] [6] pointed out that the correlation between observation matrix and sparse basis should be low, that is, to meet the condition of Restricted Isometry Property

(RIP). So that we can solve the sparse representation coefficients and reconstruct the source model if the number of sampling points met a certain condition.

Solving the sparse representation coefficients based on the observational system equation model Eq.(4) is essentially an optimization problem that search the easiest solution with the constraint of the equation model, that is to solve Eq.(5):

$$\min\|\boldsymbol{\Theta}\|_0 \quad s.t.\, \boldsymbol{y} = \boldsymbol{\Psi x} = \boldsymbol{A\Theta}, \tag{5}$$

where $\|\boldsymbol{\Theta}\|_0$ is the number of nonzero elements in $\boldsymbol{\Theta}$. It can be found that solving Eq.(5) is a combinatorial optimization problem of NP hard.

At present, the method of greedy and relaxation can be used to solve this problem. The greedy algorithm is represented by matching tracing [7] (MP), orthogonal matching tracing [8] (OMP), generalized orthogonal matching tracing [9] (GOMP), etc. The greedy algorithm is easy to use, and its algorithm complexity is only O(*kmn*), however, it is less adaptive to noise. The relaxation algorithm, such as FOCUSS (Focal Underdetermined System Solver) [10] and BP (Basis Pursuit) algorithm, is computationally larger than the greedy algorithm, but with higher accuracy and more adaptive to noise. The iterative threshold algorithm [11] greatly improves the accuracy by iteration, but the number of iterations is higher and the convergence rate is late.

From the compressed sampling method, it can be found that the higher the representation sparsity of response surface (the smaller the sparse degree *k*), the less the nonzero coefficient of response surface on the representation basis, the smaller the number of sampling points needed for the response surface approximation. The number of calling to the black box function will be less, and the optimization efficiency of the product simulation will be higher. Therefore, we propose a constructing method of sparse representation response surface model in this paper, which includes sparse response surface (SRS) model constructed based on sparse representation theory and quasi-sparse response surface (QSRS) model to apply in the case of source model is not strictly sparse on the basis.

3. Sparse response surface

3.1. The mathematical model

SRS is based on the polynomial response surface model, which is represented as linear model:

$$\hat{\boldsymbol{y}}(\boldsymbol{x}) = \sum_{i=1}^{p} \theta_i \varphi_i(\boldsymbol{x}). \tag{6}$$

It also can be written as matrix form

$$\hat{\boldsymbol{y}}(\boldsymbol{x}) = \boldsymbol{\Phi\Theta}, \tag{7}$$

where $\boldsymbol{x} = [x_1 \cdots x_m]$ is the design sampling point, and m is the number of variables. The atoms of SRS are constructed with $\{\varphi_i(\boldsymbol{x})\}_{i=1,2,\cdots p}$,and p is the number of atoms. $\boldsymbol{\Phi}$ is called dictionary, which is a set of basic functions. $\{\theta_i\}_{i=1,2,\cdots p}$ are the atoms coefficients and $\boldsymbol{\Theta}$ is the coefficient vector. We select polynomials functions as the basic atoms of SRS, and atoms is tensor product of the whole univariate polynomials. The univariate polynomials are decided by the definition of polynomials and the index vector. Thus , $\varphi_i(\boldsymbol{x})$ may be defined as

$$\varphi_i(\boldsymbol{x}) = L\left(\boldsymbol{x}, \boldsymbol{\eta}^{(i)}\right) = \prod_{j=1}^{m} l_j\left(x_j, \eta_j^{(i)}\right), i = 1, \cdots, p. \tag{8}$$

where $\boldsymbol{\eta}^{(i)} = \left[\eta_1^{(i)}, \cdots, \eta_m^{(i)}\right]^T$ is the exponent vector of $\varphi_i(\boldsymbol{x})$, presented the order of polynomial. $L\left(\boldsymbol{x}, \boldsymbol{\eta}^{(i)}\right)$ is the polynomial of basis, which is calculated through taking variable $\boldsymbol{x}$ into the polynomial definition and the order number. $l_j\left(x_j, \eta_j^{(i)}\right)$ is the $\eta_j^{(i)}$-order variable polynomial with respect to x_j.

Given a set of sampling points $\boldsymbol{x} = [x^{(1)}, \cdots, x^{(n)}]^T$, $x^{(k)} \in R^m, k = 1,2,\cdots, n$, and the corresponding actual response $\boldsymbol{y} = [y^{(1)}, \cdots, y^{(n)}]^T$, then the dictionary can be defined as

$$\boldsymbol{\Phi} = \begin{bmatrix} \varphi_1\left(x^{(1)}\right) & \cdots & \varphi_p\left(x^{(1)}\right) \\ \vdots & \ddots & \vdots \\ \varphi_1\left(x^{(n)}\right) & \cdots & \varphi_p\left(x^{(n)}\right) \end{bmatrix}. \tag{9}$$

Thus, the model of sparse response surface may be noted as

$$\widehat{\boldsymbol{y}}(\boldsymbol{x}) = \boldsymbol{\Phi\Theta}, \quad s.t.\ \|\boldsymbol{\Theta}\|_0 \leq s. \tag{10}$$

3.2. Dictionary construction

Dictionary determined the expression ability of response surface in SRS. In the theory of compression sensing, dictionary may be constituted of a set of orthogonal sparse representation basis like Fourier, discrete cosine, wavelets etc. A set of Legendre polynomial function is employed as basis in the SRS. And the Legendre polynomial is obtained by the circular definition. Assuming that $L_0(x) = 1, L_1(x) = x,$, define the following function as

$$(n+1)L_{n+1}(x) = (2n+1)L_n(x) - nL_{n-1}(x),\ n = 1,2,\cdots. \tag{11}$$

where the subscript n represents the order of Legendre polynomials. The reason why selected Legendre polynomials as basis function is that it can be orthogonal with respect to ℓ2 norm meaning in the interval [-1, 1], and this excellent feature may be extended from univariate to multivariate. In addition, Legendre polynomial is convenient to be constructed and calculated.

In addition, the choice of dictionary scale, which determines the numbers of atoms, need to be trade off between the dictionary expression ability and the solution stability. If dictionary scale is too small, it will be not enough to embody the model characteristics in sampling points and the function relationship. Whereas the dictionary is too big, calculating coefficient will be a serious underdetermined problem, and solution stability will become poorer. According to the literature [12], the number of sampling points *n* should be at least 15% of the number of atoms *p*. In this paper, we set $p = 6n$. In the meanwhile, to avoid overfitting and reduce computation, it prefers to select low-order function when choosing the Legendre polynomial function to construct atoms To guarantee the total order is as small as possible, the atomic structure of multivariate model also obey this criterion.

The choice of sampling position commands that the sampling may express characterize of source model to the greatest extent. In general, sampling points should be uniform distribution on the entire design feasible region. Uniform design (UD) and Latin hypercube design (LHD) are meet the demand, since both of them can divide design space into *n* sampling interval uniformly, and choose one sampling points in each interval. The difference is that sampling position of UD is fixed in each interval, while sampling position of LHD is random in the corresponding interval. In the SRS, LHD is employed for sampling, because the randomness in LHD can shrink off the correlation of atoms in the dictionary and facilitate the resolution of atoms expressing different characteristics.

Due to the definition intervals of Legendre polynomials function is not the same as the design space of design model, it is necessary to map sampling points of design space to the definition intervals of Legendre polynomials function $[-1,1]^m$. Then, calculate value of atoms by taking sampling points $\boldsymbol{x}$ into Legendre polynomials function, and dictionary $\boldsymbol{\Phi}$ can be constructed.

3.3. Construction of sparse response surface

The sparseness of source model in polynomial basis function is unknown. To build the model of SRS, the coefficients of the basis function can be calculated through the following function.

$$\boldsymbol{\Theta} = min\|\boldsymbol{\Theta}\|_0.\ s.t.\|\boldsymbol{y} - \boldsymbol{\Phi\Theta}\|_2^2 \leq \epsilon. \tag{12}$$

Due to the object ℓ0 norm is a nonconvex problem, we can relax the sparse object condition to ℓ1 norm. Eq.(12) is translated to :

$$\boldsymbol{\Theta} = min\|\boldsymbol{\Theta}\|_1.\ s.t.\|\boldsymbol{y} - \boldsymbol{\Phi\Theta}\|_2^2 \leq \epsilon. \tag{13}$$

ℓ1 norm is a convex problem, which is the closest to ℓ0 norm, and it's Lagrange multiplier can be

$$\widehat{\boldsymbol{\Theta}}(\lambda) = \arg\min_{\boldsymbol{\Theta}}\ \|\boldsymbol{y} - \boldsymbol{\Phi\Theta}\|_2^2 + \lambda\|\boldsymbol{\Theta}\|_1. \tag{14}$$

Eq.(14) is a famous regression model names least absolute shrinkage and selection operator (LASSO), which may provide both variable selection and coefficient shrinkage functions, is commonly applied in statistics. Variable selection can be used to generate the model of sparse response surface, and coefficient shrinkage can achieve better prediction performance through compromising the model training accuracy. The LASSO model can be solved by least angle regression (LAR). And author of literature [13] provides a MATLAB toolbox for using to calculate.

The process of constructing SRS is shown as following Table 1.

Table 1. The process of construction sparse response surface.

Step1:	**Generate sampling points.** LHD method is applied to sampling with the lhsdesign of Latin hypercube function in MATLAB. Since the sampling interval of lhsdesign function is [0,1]m, it need to be mapped the sampling points towards design space of source model.
Step2:	**Estimate objective value.** Obtain the objective value $\boldsymbol{y}$ of corresponding sampling points by simulating the black box function of source model.
Step3:	**Construct dictionary.** The sampling points sequence is mapped to Legendre polynomial definition interval[-1,1]m. Then obtain index vector $\{\boldsymbol{\eta}^{(i)}\}_{i=1,2,\cdots p}$, and calculate the value of dictionary.
Step4:	**Calculate coefficients.** Solve Eq. (13) through invoking the LASSO solver solution.
Step5:	**The end of construction.**

4. Quasi-sparsity response surface

The ability of LASSO to provide sparse models has been demonstrated in the field of statistics, but the number of atoms chosen by LASSO cannot exceed the number of rows or columns of the dictionary. For the sparse representation of the response surface, if the number of sampling points n is greater than the sparse degree k, then LASSO will select k atoms to construct a sparse response surface, which can be seen as the precise surrogate model of the source model. However, the source model is complicated in real application, its sparse degree based on the selected basis functions may be not small enough, and the increase of the sampling points will consume the expensive calculation cost, then the number of sampling points is less than sparse degree. Since the model performance is limited in the number of atoms selected by LASSO, we consider the Elastic net regression to construct a quasi-sparsity response surface model which will be more stable by selecting more atoms.

4.1. Definition of QSRS model

Consider the cost function

$$\mathcal{L}(\lambda_1, \lambda_2, \theta) = \|\boldsymbol{y} - \boldsymbol{\Phi\Theta}\|^2 + \lambda_1\|\boldsymbol{\Theta}\|_1 + \lambda_2\|\boldsymbol{\Theta}\|_2^2, \tag{15}$$

where

$$\|\boldsymbol{\Theta}\|_1 = \sum_{j=1}^{p}|\theta_j|,\ \|\boldsymbol{\Theta}\|_2^2 = \sum_{j=1}^{p}\theta_j^2.$$

The Elastic net estimator $\widehat{\Theta}$ is the minimize of Eq.(14):

$$\widehat{\boldsymbol{\Theta}} = argmin_{\boldsymbol{\Theta}}\{\mathcal{L}(\lambda_1, \lambda_2, \theta)\}. \tag{16}$$

From the Eq.(15), we can see that the elastic net is a regularized regression method that linearly combines the ℓ1 and ℓ2 penalties of LASSO regression and ridge regression.

4.2. Solution of Elastic net regression

Theorem. Given data set (**y**,**X**) and (λ_1, λ_2), the elastic net estimation is

$$\boldsymbol{\Theta}(elastic\ net) = argmin_{\boldsymbol{\Theta}}\,\boldsymbol{\Theta}^T\left(\frac{\boldsymbol{X}^T\boldsymbol{X}+\lambda_2\boldsymbol{I}}{1+\lambda_2}\right)\boldsymbol{\Theta} - 2\boldsymbol{y}^T\boldsymbol{X\Theta} + \lambda_1|\boldsymbol{\Theta}|_1. \tag{17}$$

It is easy to see that

$$\boldsymbol{\Theta}(lasso) = argmin_{\boldsymbol{\Theta}}\,\boldsymbol{\Theta}^T(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{\Theta} - 2\boldsymbol{y}^T\boldsymbol{X\Theta} + \lambda_1|\boldsymbol{\Theta}|_1. \tag{18}$$

The proof of theorem is simple, and we omit it here. We can find there is a convenience for elastic net that the estimation process can be turned to the LASSO solution from above subsection. We introduce a lemma as below.

Lemma. Given data set (**y**, **X**) and (λ_1, λ_2), define an artificial data set (**y***, **X***) by

$$\boldsymbol{X}^*_{(n+p)\times p} = (1+\lambda_2)^{-\frac{1}{2}}\left(\frac{\boldsymbol{X}}{\sqrt{\lambda_2}\boldsymbol{I}}\right),\ \boldsymbol{y}^*_{(n+p)} = \begin{pmatrix}\boldsymbol{y}\\ \boldsymbol{0}\end{pmatrix}. \tag{19}$$

Let $\gamma = \lambda_1/\sqrt{(1+\lambda_2)}$, $\Theta^* = \sqrt{(1+\lambda_2)}\Theta$. Then LASSO question is equaled with

$$\mathcal{L}(\gamma, \boldsymbol{\theta}) = \mathcal{L}(\gamma, \boldsymbol{\theta}^*) = |\boldsymbol{y}^* - \boldsymbol{X}^*\boldsymbol{\theta}^*|^2 + \gamma|\boldsymbol{\theta}^*|_1. \quad (20)$$

Let $\Theta^* = \text{argmin}_{\Theta^*}\{\mathcal{L}(\gamma, \Theta)\}$, then

$$\boldsymbol{\theta}(elasticnet) = \sqrt{1+\lambda_2}\boldsymbol{\theta}^*. \quad (21)$$

We omit the proof for it is just simple algebra. For the fixed λ_2, the elastic net problem can be changed to a simple LASSO problem which can be solved by the "LAR-EN" solver. In Ref [14], the author gives a "LAR-EN" solver that can directly find Elastic net solutions for fixed λ_2.

4.3. Construction of QSRS

There are two main differences between the construction of QSRS and SRS construction: the choice of parameter λ_2; the sampling mode is changed from LHD sampling to UD sampling.

In the construction of the QSRS, the parameter λ_2 is determined by the method of cross-validation. The specific flow is shown in Table 2.

Elastic net will select a group of similar atoms, regardless of the correlations between atomics, and the sampling points of UD is fixed, then, the response surface is fixed, the reconstruction effect is stably. Thus, UD sampling is employed in the constructing of quasi-sparsity response surface.

Table 2. The process of sparse response surface construction.

Step1:	Generate sampling points by Uniform sampling.
Step 2:	Get the responding value ***y*** of the sampling points.
Step 3:	Divide the sampling points into K sets. (K is priority to use 10, or other values that can divide the number of sampling points)
Step 4:	Respectively, fix$\lambda_2$to0, 0.0001,0.001,0.01,0.1,1and 10, run steps 5 to 7.
Step 5:	K-1 copies of the sampling points and responding values are taken as the training set, and the remaining one is the prediction set. The Quasi-sparse response surface model is constructed by using the LAR-EN solver with the training set, the prediction accuracy of the model is evaluated on the prediction set.
Step 6:	Repeat step 5 to ensure that each set is used for predicting.
Step 7:	Calculate the mean of the predictor error as the prediction ability of every λ_2.
Step 8:	Take $\lambda_2$with thebest prediction ability.
Step 9:	Construct the QSRS using the LAR-EN solver with the training set and the determined λ_2 .

5. Numerical experiments

In order to verify the appreciate effect of SRS and QSRS, 2 simulation test functions and one practical engineering design problem are chosen in this section as the object to be reconstructed, and some commonly used response surface models are chose as a comparison.

5.1. Experiments problems

1) Test function 1 (10 variables)

$$f(x) = \sum_{i=1}^{m}[0.6 + sin\left(\frac{16*x_i}{15} - 1\right) + sin\left(\frac{16*x_i}{15} - 1\right)^2 + sin(4 * \left(\frac{16*x_i}{15} - 1\right))], \quad (22)$$

where $x_i \in [-1,1]$, i=1,2,3,…,m, m=10.

2) Test function 2 (20 variables)

$$f(x) = 1 + \sum_{i=1}^{m}\{8\sin^2[7(x_i - 0.9)]^2 + 6\sin^2[14(x_i - 0.9)]^2 + (x_i - 0.9)^2\}, \quad (23)$$

where $x_i \in [-500,500]$, i=1,2,3,…,m, m=20.

3) Engineering design problem

Reference [15] provides a few design studies based on finite element analysis and natural experiment design of composite sheet structures. The analysis process is complicated in the optimization process. The response surface method can provide an approximate model for this kind of problem, save the computation time and improve the simulation optimization efficiency. We choose a global deflection structure optimization example of a square composite sheet as the validated case of SRRS model. The specific description and data of the case can be downloaded from

http://www.cs.rtu.lv/jekabsons/datasets.html. The optimization target of this case is decided by six variables, and the variable description and range of values are shown in Table 3.

Table 3. Variables description of engineering design problem.

Variable	Description	Minimum	Maximum	Units
L	Panel length	3	7	m
h	Panel height	4	16	mm
t1	Top and bottom plate thickness	2	4	mm
t1	Core stiffener thickness	1.5	4	mm
kh	Core stiffener spacing factor	1.5	4	
n	Symmetrical number of core stiffeners	2	6	

5.2. Accuracy assessment criteria

The following metrics are used to assess the generalization capabilities of the response surface models compared in this work:

1) Root mean squared error(RMSE)

The RMSE is given by

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y^{(k)} - \hat{y}^{(k)})^2}. \tag{24}$$

where $y^{(k)}$ is the actual response and $\hat{y}^{(k)}$ is the predicted response at the k-th test point, n is the number of test points.

2) Maximum absolute error(MAE)

The maximum absolute error is given by

$$MAE = max\left|y^{(k)} - \hat{y}^{(k)}\right|, k = 1,2,\cdots,n. \tag{25}$$

3) Correlation coefficient (R)

The correlation coefficient between actual response and predicted response is expressed as

$$R = \frac{(\sum_{k=1}^{n}(y^{(k)}-\bar{y})(\hat{y}^{(k)}-\bar{\hat{y}}))}{\left(\sqrt{\sum_{k=1}^{n}(y^{(k)}-\bar{y})^2}\sqrt{\sum_{k=1}^{n}(\hat{y}^{(k)}-\bar{\hat{y}})^2}\right)}. \tag{26}$$

where $\bar{y}$ is the mean of actual response and $\bar{\hat{y}}$ is the mean of prediction responses.

5.3. Models settings

SRS and QSRS are compared with the models listed in introduction section including PRS, OK, BK, RBF, ERBF, MARS, and SVR. The parameters of these models are listed in Table 4.

Table 4. The detail settings of different surrogate models.

Surrogate model	Details
PRS	p=2
OK	A constant regression function and a Gaussian correlation model are employed in the mode. In all cases, $\theta_0=1_{m*1}$, and $0.1\leq \theta_i \leq 20$, for i = 1, … ,m, where *m* and $1_{m*1}$are the number of variables and the vector whose entries are all equal to 1, respectively
BK	The same as OK.
RBF	The form of basis functions is the multi quadric function and we set c = 0.9.
ERBF	The parameter n is set equal to 2 and the parameter c is set to be approximately 1/3 of the average domain size. The remaining options are the same as RBF.
MARS	The piecewise-cubic regression functions are adopted in this approach and the stopping criterion for the forward phase is set to 1e-4. Additionally, the maximal number of basis functions in the forward model building phase is 81 and the Generalized Cross-Validation penalty per knot is 3.
SVR	We choose the RBF kernel function which is set with σ = 4 and the e-insensitive loss function. The parameters C, e are selected as C =100max($|\bar{y} + 3\sigma\ y|$,$||\bar{y} - 3\sigma\ y|$) and $\varepsilon = \delta_y/\sqrt{ns}$, respectively. Here, $\bar{y}$ and σ_y the mean value and the standard deviation of the function values at the training points

5.4. Experiment settings

For all validated objects, the sampling points and test points are generated by LHD in the design domain, and the sampling points are used to construct the response surfaces. The test points are used to test the model approximation accuracy. To reduce the effect of random sampling on the test results, the test is to be repeated 300 or 100 times. The average of the accuracy criteria is used to measure the approximation of the response surface model. Detailed can be seen in Table 5.

Table 5. Numerical setup for test problems.

Test problem	NO. of variables	No. of training points	No. of test points	No. of test times
Test function 1	10	100	4500	100
Test function 2	20	120	5000	100
Engineering problem	6	50	450	300

5.5. Experiments analysis and discussion

5.5.1. RMSE. Table 6 shows the comparison of the average RMSE results of different response surface models. It can be seen that the RMSEs of both SRS and QSRS models are significantly better than other common response surface models in this test. From the test results of test function 1, all tested response surface models' RMSEs are low. The performance of QSRS is the best, which is 11.6% higher than that of SRS. In addition to SRS and QSRS, ERBF and MARS also have a good approximation effect. For test function 2, because the number of variables reaches 20 whereas the number of sampling points is only 120, so most of the response surface models can not be a good object to the reconstruction of the approximation, however, both SRS and QSRS are doing well. For the engineering design problem, due to the small number of variables, all the models are doing well, SRS and QSRS have the advantage relative to other models. In general, SRRS has a significant advantage in approximating RMSE, and this advantage is more pronounced in the case of multi-variable and less sampling points. In response to SRS and QSRS, the performance of QSRS is better than SRS. Because when the parameter $\lambda_2 = 0$, the QSRS degrades to the SRS, so the QSRS has performance that is not weaker than the SRS. However, it must be noted that the construction of the QSRS is more expensive than the SRS due to the cross validation process

Table 6. Mean of RMSEs for different surrogate models, the optimal value in each column is shown in bold for ease of comparison.

Surrogate model	Test function 1	Test function 2	Engineering design problem
PRS	3.4667	9.658e5	0.0065
OK	2.3096	3.332e5	0.0122
BK	4.0665	1.786e5	0.0061
RBF	2.2124	1.711e5	0.0143
ERBF	1.6026	5.013e3	0.0203
MARS	1.6193	8.776e4	0.0081
SVR	2.8475	4.176e5	0.0060
SRS	1.4967	34.821	0.0041
QSRS	**1.3415**	**33.826**	**0.0037**

5.5.2. MAE. Table7 compared the mean MAE results of all models. It can be seen that the trend of MAE is consistent with the trend of RMSE from the results. SRS and QSRS are also better than the other response surface models, and the advantage is more obvious in the case of multi-variable and less sampling points such as test function 2 which have 20 variables.

Table 7. Mean of MAEs for different surrogate models, the optimal value in each column is shown in bold for ease of comparison.

Surrogate model	Test function 1	Test function 2	Engineering design problem
PRS	15.9729	4.231e6	0.0401
OK	8.3281	1.303e6	0.0834
BK	18.3734	8.611e5	0.0378
RBF	8.4164	7.504e5	0.0834
ERBF	6.9643	1.990e4	0.0963
MARS	7.4095	3.345e5	0.0537
SVR	11.7081	1.827e6	0.0478
SRS	6.2611	146.447	0.0330
QSRS	**6.0611**	**140.756**	**0.0313**

5.5.3. Correlation coefficient. Table 8 shows the comparison of the mean of correlation coefficient results for different surrogate models. From the results of test function 1, the performance of the QSRS is still the best, the performance of ERBF and MARS are also good. It is worth noting that the performance of SRS is not as good as ERBF and MARS, the reason should be the sparse degree of test function 1 is greater than the number of sampling points, then the number of selected atoms is too small and the response surface is not enough stable. For the test function 2, SRS, QSRS, and MARS all behave well, and the mean R of MARS is even higher than the SRS. For the engineering design problem, the performance of QSRS and SRS response are in the top two. It can be seen that although there are other response surface models may be better than SRRS in the correlation coefficient, but on the whole, both SRS and QSRS play a stable and excellent performance.

Table 8. Mean of Rs for different surrogate models, the optimal value in each column is shown in bold for ease of comparison.

Surrogate model	Test function 1	Test function 2	Engineering design problem
PRS	0.2480	0.3309	0.8838
OK	0.0179	1.81e-8	0.4748
BK	0.1864	0.8543	0.8832
RBF	0.4125	0.9365	0.4667
ERBF	0.8262	0.9399	0.4260
MARS	0.8125	0.9633	0.7986
SVR	0.2454	0.0778	0.8982
SRS	0.7758	0.9477	0.9491
QSRS	**0.8369**	**0.9499**	**0.9621**

6. Conclusions

In this paper we proposed a sparse representation response surface include SRS and QSRS based on compressed sampling method. The Legendre polynomial function is employed as the sparse basis, the SRS and QSRS models are constructed by LASSO or Elastic net regression method and LHD or UD sampling. The SRS and QSRS are compared with other commonly used response surface model in the numerical experiments of two standard test functions and one practical engineering problem. After the accuracy criteria evaluation of RMSE, MAE, and R, the results show that both SRS and QSRS have better approximation ability than other commonly used response surface models.

The conclusions can be given as below: 1)The sparse representation response surface can be used to approximate the simulation object, and only a small amount of sampling can obtain effective simulation and optimization efficiency, 2)Sparse representation response surface has a better approximation effect compared to the other commonly used response surfaces tested in this paper, and this advantage is more obviously in the case of multi-variable and less sampling.

Acknowledgments

This work was financially supported by Natural Science Foundation of Guangdong Province of China under Project 2015A030310510 and 2016KQNCX156, and the Science and Technology Program of Shaoguan City of China (441-99000318 and 441-99000319).

References

[1] Torry-Smith J M, Qamar A, Achiche S, et al. Challenges in Designing Mechatronic Systems[J]. Journal of Mechanical Design, 2013, 135(1): 184-194.
[2] Wang G G, Shan S. Review of Metamodeling Techniques in Support of Engineering Design Optimization[C]. ASME 2006 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 2006: 415-426.
[3] Hong L J, Nelson B L. A brief introduction to optimization via simulation[C]. Winter Simulation Conference, 2009: 75-85.
[4] Romero D A, Amon C H, Finger S. Multiresponse Metamodeling in Simulation-Based Design Applications[J]. Journal of Mechanical Design, 2012, 134(134): 091001.
[5] Candes E J, Wakin M B. An Introduction To Compressive Sampling[J]. IEEE Signal Processing Magazine, 2008, 25(2): 21-30.
[6] Donoho D L. Compressed sensing[J]. IEEE Transactions on Information Theory, 2006, 52(4): 1289-1306.
[7] Mallat S G, Zhang Z. Matching pursuits with time-frequency dictionaries[J]. IEEE Transactions on Signal Processing, 1993, 41(12): 3397 - 3415.
[8] Tropp J A, Gilbert A C. Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit[J]. IEEE Transactions on Information Theory, 2007, 53(12): 4655-4666.
[9] Jian W, Kwon S, Shim B. Generalized Orthogonal Matching Pursuit[J]. IEEE Transactions on Signal Processing, 2012, 60(12): 6202-6216.
[10] Gorodnitsky I F, Rao B D. Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm[J]. IEEE Transactions on Signal Processing, 2002, 45(3): 600-616.
[11] Blumensath T, Davies M E. Iterative Thresholding for Sparse Approximations[J]. Journal of Fourier Analysis and Applications, 2008, 14(5): 629-654.
[12] Gribonval R, Cevher V, Davies M E. Compressible Distributions for High-Dimensional Statistics[J]. IEEE Transactions on Information Theory, 2012, 58(8): 5016-5034.
[13] Efron B, Hastie T, Johnstone I, et al. Least Angle Regression[J]. Annals of Statistics, 2004, 32(2): 407-451.
[14] Zou H, Hastie T. Regularization and Variable Selection via the Elastic Net[J]. Journal of the Royal Statistical Society, 2005, 67(2): 301-320.
[15] Kalnins K, Eglitis E, Jekabsons G, et al. 3.2 – Metamodels for Optimum Design of Laser Welded Sandwich Structures[J]. Design Fabrication & Economy of Welded Structures, 2008: 119-126.