

Research Article

Kernel Parameter Optimization for Kriging Based on Structural Risk Minimization Principle

Hua Su, Chunlin Gong, and Liangxian Gu

Shaanxi Aerospace Flight Vehicle Design Key Laboratory, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China

Correspondence should be addressed to Hua Su; su@nwpu.edu.cn

Received 26 May 2017; Accepted 10 August 2017; Published 11 September 2017

Academic Editor: Sebastian Heidenreich

Copyright © 2017 Hua Su et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An improved kernel parameter optimization method based on Structural Risk Minimization (SRM) principle is proposed to enhance the generalization ability of traditional Kriging surrogate model. This article first analyses the importance of the generalization ability as an assessment criteria of surrogate model from the perspective of statistics and proves the applicability to Kriging. Kernel parameter optimization method is used to improve the fitting precision of Kriging model. With the smoothness measure of the generalization ability and the anisotropy kernel function, the modified Kriging surrogate model and its analysis process are established. Several benchmarks are tested to verify the effectiveness of the modified method under two different sampling states: uniform distribution and nonuniform distribution. The results show that the proposed Kriging has better generalization ability and adaptability, especially for nonuniform distribution sampling.

1. Introduction

Computer-intensive optimization problem becomes more and more as the increasing requirement for high-fidelity model in industry area, especially for aerospace engineering. It is very urgent to improve computation efficiency [1, 2]. Surrogate model is the comprehensive application of the Design of Experiment (DOE), mathematical statistics, and optimization techniques. It approximates the complicated and time-consuming physical model by building some analytical mathematical model, to reduce the analysis process and smooth the design space. It becomes one of the most effective methods for computer-intensive problem and has been widely applied to high-fidelity design and optimization [3].

The fitting precision of surrogate model for physical model is one of the most important indicators. For computational resource limitation, it is not rational to assess the entire design space of the physical model directly; using the error between surrogate model and physical model at the sample points to estimate the fitting precision is the most effective way. Fitting precision is a function of the sample capacity. It converges to the true precision only when the

sample capacity satisfies the large number theorem. Actually sample capacity and fitting precision are conflicting. We hope less sample points with higher fitting precision. Vapnik proposed the statistical learning theory (SLT) in the late 1970s [4, 5] and became a machine learning method for limiting sample points in the 1990s. The core idea is to control the generalization ability of the learning machine by control the capacity of the machine. That is to improve the fitting precision of the surrogate model by applying SRM principle.

There are many surrogate models, such as Response Surface Model (RSM), Kriging, Radial Basis Function (RBF), and Support Vector Regression (SVR). Kriging is an interpolation method based on statistical theory; the main idea is to evaluate the approximate function of the object based on the dynamic construction of design space to predict the information of unknown points [6, 7]. For Kriging, linear weighted combination of the information of nearby points is also needed for predicting unknown points, which can be determined by minimizing the variance of the estimated value error, which means Kriging is also a best linear unbiased estimator problem [8]. Kriging surrogate model is simple and stable compared to other methods and has been widely used in many fields [9–12]. Some improved Kriging models have

also been studied, such as Gradient-Enhanced Kriging [13], CoKriging [14], and Hierarchical Kriging [15].

Kriging is sensitive to sample points. It has worse predicting result when with less sample points. The quantity of information of the correlation matrix and the parameters of the basis function have obvious influence to fitting precision. It is very necessary to optimize the parameters of the basis function to improve fitting capacity. SRM is the most important theory to evaluate the generalization ability of the surrogate model, which has been widely used in kernel parameter optimization of SVM [16], fuzzy model [17], and chaotic system [18]. But there are a few studies for other surrogate models. Zhu [19] proposed an automatic method to optimize basis function for RBF using the SRM principle. Chen [20] applied to SRM principle to RBF network study to improve the generalization ability. These previous researches indicate that using SRM principle to optimize the parameters of the surrogate model really can improve the fitting precision of surrogate model.

This paper proposes an anisotropic basis function parameter optimization method for Kriging based on SRM principle. The parameters of the correlation functions of Kriging are optimized using the smoothness measure as objective function. Some benchmarks with different scale are used to verify the proposed methods in this paper. The influence of the distribution pattern of sample points with uniform distribution and nonuniform distribution is also studied.

2. Assessment Criteria of Surrogate Model

The purpose of machine learning is to find the internal dependencies by learning given data to predict unknown data or estimate model characteristic. From the perspective of statistics, machine learning can be viewed as to predict the unknown output relation by some given training samples as accurate as possible. To minimize the risk functional, we should find the optimal function $f(x, \alpha^*)$ in function set $\{f(x, \alpha)\}$:

$$R(\alpha) = \int (y - f(x, \alpha))^2 dF(x, y). \quad (1)$$

$\{f(x, \alpha)\}$ is the prediction function set; α is the parameter of the prediction function. So the machine learning is designed to minimize the risk functional of the squared error loss function.

2.1. Empirical Risk Minimization Principle. Usually we cannot predict the specific form of probability distribution and only know there is a proper probability distribution which can fit the training samples. The expected risk depends on probability distribution $F(x, y)$; it is hard to minimize the risk functional directly by constructing predictive function of minimized expected risk with several samples. Therefore, Empirical Risk Minimization (ERM) principle is proposed to minimize risk functional in the traditional machine learning method,

$$R_{\text{emp}}(\alpha) = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i, \alpha))^2. \quad (2)$$

ERM has been widely used in the least square method for regression problem, maximum likelihood method for probability density estimation, and neural network learning [21]. But ERM is based on the large number theorem, completely dependent on the sum of squared error of the samples. With less samples, small empirical risk cannot guarantee minimizing expected risk. There may happen oversmoothing for the RBF and neural network, which is based on ERM principle [22]: little training error for sample points but large testing error for other points.

2.2. Structural Risk Minimization Principle. Generalization ability means the capacity to predict or estimate unknown phenomenon by machine learning. ERM principle cannot guarantee minimizing expected risk. Therefore, SRM principle is proposed based on the statistical learning theory. It provides the relationship between empirical risk and real risk, which is known as the bound of the generalization ability. SRM principle divides the real risk of machine learning into empirical risk and confidence interval [4]:

$$R(\alpha) \leq R_{\text{emp}}(\alpha) + \Phi\left(\frac{h}{n}\right). \quad (3)$$

h is the Vapnik-Chervonenkis (VC) dimension, which is the most important theoretical basis of the statistical learning theory. It defines the capacity of the function set and reflects the generalization ability of the machine learning. It is the best descriptive indicators for the capacity of function set learning until now.

From (3), we know ERM principle is unconscionable with limited samples. When h/n becomes larger, empirical risk decreases, but real risk does not always become lower. When the dimension of samples is fixed, reduced h (VC) may decrease confidence interval, so the empirical risk comes close to real risk. In this case, less empirical risk stands for less expected risk. Increasing the capacity of samples n also decreases confidence interval to reduce real risk, but it is difficult to get plenty of samples because of the limitation of computational cost.

Therefore, decreasing VC dimension is the most suitable way to reduce real risk. For a given observation set, SRM principle chooses the proper function $(y - f(x, \alpha_n^k))^2$ from the subset S_k with minimized risk to minimize the empirical risk. So SRM principle is a compromise method to balance the fitting precision and the complexity of fitting function.

2.3. SRM Index Based on Smoothness Measure. Smoothness measure is used to evaluate the generalization ability. In the projection space formed by kernel function, smoothness measure has its more direct and natural definition. It can be described as the norm of Hilbert space of the kernel function, which is defined as follows.

Every function $f(x)$ in Hilbert space of kernel function K can be expressed as

$$f(x) = \sum_{i=1}^n \alpha_i k(x, x_i). \quad (4)$$

Then the norm of $f(x)$ in Hilbert space is

$$\|f\|_R^2 = \sum_{i=1}^n \frac{\alpha_i^2}{\lambda_i}. \quad (5)$$

λ_i is the eigenvalue of kernel K . The generalization ability can be measured by evaluating $\|W\|^2$. The optimal regression problem becomes

$$\begin{aligned} \min \quad & \|W\|^2 \\ \text{s.t.} \quad & W\phi + b = Y, \end{aligned} \quad (6)$$

where $\|W\|^2$ is the smoothness measure of sample points.

2.4. Mercer Theorem. The criteria to determine whether a function is kernel function is as follows.

I is the domain of the independent variables u and u' . There is a nonnegative real function $K(u, u')$; it is continuous in $I \times I$. The sufficient and necessary condition for $K(u, u')$ to be expressed as a series

$$K(u, u') = \sum_i \lambda_i \psi_i(u) \psi_i(u') \quad (7)$$

is that all functions satisfy $\int g^2(u) du < \infty$, $g \neq 0$ and also satisfy the following condition:

$$\iint k(u, u') g(u) g(u') du du' > 0, \quad (8)$$

where the series is convergence in $I \times I$. λ_i is the eigenvalue ($\lambda_i > 0$) of $K(u, u')$. $\psi_i(u)$ is the eigenvector of $K(u, u')$. A function satisfying the above definition is known as Mercer kernel and can be expressed as the inner product in Hilbert space,

$$K(u, u') = \langle k(\cdot, u), k(\cdot, u') \rangle = \langle \phi(u), \phi(u') \rangle, \quad (9)$$

where $\langle \cdot, \cdot \rangle$ is the operator of inner product. $\phi : u \rightarrow k(\cdot, u)$ is a nonlinear transformation of kernel function.

3. Kriging Based on SRM Principle

3.1. Kriging Model. Kriging model includes two parts:

$$y(x) = f^T(x) \beta + z(x), \quad (10)$$

where $f^T(x)\beta$ is a polynomial regression model and $z(x)$ is a deviation model. The covariance matrix of $z(x)$ is

$$\text{Cov}[Z(x_i), Z(x_j)] = \sigma^2 K(\theta, x_i, x_j), \quad (11)$$

where K is the correlative coefficient matrix. $K(\theta, x_i, x_j)$ is the space correlation function of any two sample points x_i and x_j in sample points set. It plays a decisive role to fitting precision. θ is the parameters of K . $d_k = x_i^{(k)} - x_j^{(k)}$ and the common correlation functions include the following [23].

Exponential function is

$$K(\theta_k, d_k) = \exp(-\theta_k |d_k|). \quad (12)$$

Gaussian function is

$$K(\theta_k, d_k) = \exp(-\theta_k |d_k|^2). \quad (13)$$

Linear function is

$$K(\theta_k, d_k) = \max\{0, 1 - \theta_k |d_k|\}. \quad (14)$$

Spherical function is

$$\begin{aligned} K(\theta_k, d_k) &= 1 - 1.5\xi_k + 0.5\xi_k^3, \\ \xi_k &= \min\{1, \theta_k |d_k|\}. \end{aligned} \quad (15)$$

Cube function is

$$K(\theta_k, d_k) = 1 - 3\xi_k^2 + 2\xi_k^3, \quad \xi_k = \min\{1, \theta_k |d_k|\}. \quad (16)$$

Applying the linear combination of the response \hat{y} of known samples to predict the response $y(x)$ of any given samples,

$$\hat{y} = C^T Y. \quad (17)$$

The interpolated coefficients of Kriging were obtained by the condition of unbiasedness and the principle of minimum variation. The mean value of error must be zero with minimized variance σ^2 to guarantee the unbiasedness of the fitting process. The following is the corresponding Lagrangian minimization problem:

$$L(C, \lambda) = \sigma^2 (1 + C^T K C - 2C^T r) - \lambda^T (F^T C - f). \quad (18)$$

The parameter of correlative coefficient matrix K can be obtained by minimizing the Maximum Likelihood Estimation (MLE) [23].

3.2. Kriging Process. The output of Kriging model can be expressed as

$$\hat{y} = f^T \beta^* + r^T K^{-1} (Y - F \beta^*), \quad (19)$$

$$r^T = [K(x, x_1), K(x, x_2), \dots, K(x, x_n)]. \quad (20)$$

From Section 2.4, if $K(x, x_i)$ satisfies Mercer theorem, it can be expressed as an inner product in Hilbert space

$$K(x, x_i) = \langle \phi(x), \phi(x_i) \rangle. \quad (21)$$

Equation (19) is replaced with

$$\hat{y} = f^T \beta^* + \phi(x) \cdot \left(\sum_{i=1}^n \phi(x_i) \right) \cdot K^{-1} (Y - F \beta^*). \quad (22)$$

Equation (22) is simplified as

$$\hat{y} = W \phi(x) + b, \quad (23)$$

where

$$W = \left(\sum_{i=1}^n \phi(x_i) \right) \cdot K^{-1} (Y - F \beta^*), \quad (24)$$

$$b = f^T \beta^*.$$

The transformed equation (23) has the same form as the kernel function of SVR. It constructs a constant nonlinear mapping $\phi : R^n \rightarrow Z$ from input space R^n to feature space Z . Input variable x_i has been mapped into feature space Z , and a linear model is generated for linear learning in feature space. The correlative coefficient matrix $K(\theta, x_i, x_j)$ of Kriging has the same mean to the kernel matrix of SVR; it has also been called kernel matrix in this paper. Then $\|W\|^2$ can be used to evaluate the smoothness of the fitting function; it has been called smoothness measure. From the above derivation we knew that SRM principle also can be used to minimize the VC dimension of Kriging model to obtain minimized expected risk [4].

Kernel matrix K must be positive definite symmetric matrix or conditioned positive definite symmetric matrix. The kernel matrix formed by Gauss basis function satisfies Mercer theorem. There are also many other functions which can be found in [24].

3.3. Anisotropic Kernel Function. The basis function of Kriging can be expressed as

$$\varphi(c, r) = f(c_j, r_{i,j}), \quad i, j = 1, 2, \dots, n. \quad (25)$$

The parameter c of basis function means the contribution rate of every sample point to Kriging model. For a problem with n samples, the size of parameter c also is n . As the size of samples increases, the scale of the kernel parameter optimization problem also becomes larger. These may remarkably increase Kriging modeling time. Usually the parameter c_j is considered as the same value for all samples; the basis function reduces to an isotropic model. This simplification reduces the design space of Kriging model and weakens fitting effect. The following anisotropic kernel function with different contribution by each dimension is used to balance the fitting effect and computational cost:

$$\begin{aligned} \varphi(\theta, r) &= f(r(\theta_t)_{i,j}), \\ & \quad t = 1, 2, \dots, m, \quad i, j = 1, 2, \dots, n, \\ r(\theta_t)_{i,j} &= \sqrt{\sum_{t=1}^m \theta_t (x_i^{(t)} - x_j^{(t)})^2}, \end{aligned} \quad (26)$$

$$\frac{\partial \|W\|_2^2(\theta_t)}{\partial \theta_t} = \begin{cases} -(\alpha)^T \cdot \left\{ \frac{dk(x_i^{(t)}, x_j^{(t)})}{dr} \cdot \frac{(x_i^{(t)} - x_j^{(t)})^2}{2r} \right\}_{\substack{i=1, \dots, n \\ j=1, \dots, n}} \cdot (\alpha), & j \neq i, \\ 0, & j = i, \end{cases} \quad (32)$$

where $\{(i, j)\}$ is the matrix composed by every derivative value at (i, j) . dk/dr is the derivative of basis function for the distance of samples. Based on this derivative information, the

where m is the dimension of sample x . θ_t is the contribution of the t th dimension. This improved kernel function has the advantage of constant scale of kernel parameter optimization problem for a given problem with any size of samples.

3.4. SRM-Kriging. The kernel parameter optimization problem of Kriging model with smoothness measure is

$$\min R(\theta) = \|W\|^2, \quad (27)$$

where θ is the kernel parameter. $\|W\|^2$ is the smoothness measure as objective function:

$$\begin{aligned} \|W\|^2 &= W^T W \\ &= (K^{-1}(Y - F\beta^*))^T \left(\sum_{i=1}^n \phi(x_i) \right)^T \left(\sum_{i=1}^n \phi(x_i) \right) \\ & \quad \cdot K^{-1}(Y - F\beta^*) \\ &= (Y - F\beta^*)^T \cdot K^{-1} \cdot (Y - F\beta^*), \end{aligned} \quad (28)$$

where β^* is the generalized least squares solution of the polynomial problem $F\beta = Y$. The derivative of $\|W\|^2$ for θ is

$$\begin{aligned} \frac{\partial \|W\|_2^2(\theta)}{\partial \theta} &= -(Y - F\beta^*)^T \cdot K^{-1} \frac{\partial K}{\partial \theta} K^{-1} \\ & \quad \cdot (Y - F\beta^*). \end{aligned} \quad (29)$$

Let $\alpha = K^{-1}(Y - F\beta^*)$; the derivative can be

$$\frac{\partial \|W\|_2^2(\theta)}{\partial \theta} = -\alpha^T \frac{\partial K}{\partial \theta} \alpha. \quad (30)$$

The derivative for each θ_t is

$$\begin{aligned} \frac{\partial K(i, j)}{\partial \theta_t} &= \frac{dK(i, j)}{dr} \cdot \frac{\partial r}{\partial \theta_t} \\ &= \begin{cases} \frac{dk(x_i^{(t)}, x_j^{(t)})}{dr} \cdot \frac{(x_i^{(t)} - x_j^{(t)})^2}{2r}, & j \neq i, \\ 0, & j = i. \end{cases} \end{aligned} \quad (31)$$

Then we get

efficient Sequential Quadratic Programming (SQP) method is used to optimize the kernel parameter optimization problem of Kriging model.

4. Test Cases

Nine standard test functions with different dimension are used to validate the effectiveness of the proposed method. The improved Kriging model also has been evaluated by sample points with uniform distribution and nonuniform distribution.

4.1. *Standard Test Functions.* There are nine standard test functions with different dimension.

Fun1 is

$$f(x) = \frac{1}{1 + 3x^2}. \quad (33)$$

Fun2 is

$$f(x) = \frac{\sin(x)}{x}. \quad (34)$$

$$f(x) = \sum_{i=1}^{10} x_i \left(c_i + \ln \frac{x_i}{x_1 + \dots + x_{10}} \right), \quad (38)$$

$$x \in [1, 5] \quad c = [-6.089, -17.164, -34.054, -5.914, -24.721, -14.986, -24.100, -10.708, -26.662, -22.179].$$

Fun7 is

$$f(x) = \sum_{i=1}^{16} \sum_{j=1}^{16} a_{i,j} (x_i^2 + x_i + 1) (x_j^2 + x_j + 1). \quad (39)$$

Fun8 is

$$f(x) = \sum_{i=1}^{30} x_i^2. \quad (40)$$

Fun9 is

$$f(x) = \frac{1}{100} \sum_{i=1}^{30} 100 (x_i^2 - x_{i+1})^2 + (x_i - 1)^2. \quad (41)$$

These functions are from one dimension to thirty dimensions with minor different nonlinearity, which can be used to evaluate the performance of the surrogate model under different scales of variables.

These test functions are divided into three scales according to their dimension: small scale, middle scale, and large scale. The sampling information of the above functions is in Table 1.

Kriging with SRM (SRM-Kriging), Kriging with MLE (MLE-Kriging), and Kriging with constant kernel parameter (CON-Kriging) are three methods tested using the above test cases. Gauss function is chosen as the kernel function for all the three methods. Multiple correlation coefficient R^2 and root-mean-square error (RMSE) are used to evaluate the fitting precision.

Fun3 is

$$f(x_1, x_2) = (30 + x \sin(x)) (4 + e^{-y^2}). \quad (35)$$

Fun4 is

$$f(x_1, x_2) = 2x_1^2 - 1.05x_1^4 + \frac{x_1^6}{6} - x_1x_2 - x_2^2. \quad (36)$$

Fun5 is

$$\begin{aligned} f(x) = & x_1^2 + x_2^2 + x_1x_2 - 14x_1 - 16x_2 + (x_3 - 10)^2 \\ & + 4(x_4 - 5)^2 + (x_5 - 3)^2 + 2(x_6 - 1)^2 + 5x_7^2 \\ & + 7(x_8 - 11)^2 + 2(x_9 - 10)^2 + (x_{10} - 7)^2 \\ & + 45. \end{aligned} \quad (37)$$

Fun6 is

4.2. *Samples with Uniform Distribution.* The optimal Latin hypercube sampling method is used to generate uniform distributed samples. Figure 1 and Table 2 are the comparison of the generalization ability between different test functions with uniform distribution sampling. It is observed that, as the dimension of variables increases, the fitting precision of CON-Kriging decreases rapidly, especially for the large-scale cases (Fun7~Fun9); MLE-Kriging has better fitting precision than CON-Kriging via using optimized kernel parameter, but still with poor fitting precision for large-scale cases. SRM-Kriging has the best fitting precision than CON-Kriging and MLE-Kriging; even for the large-scale cases, it also has the smallest RMSE.

Figure 2 shows the fitting results of the Fun1 and Fun2 under uniform distribution sampling. From the fitting result of Fun1 we knew that CON-Kriging and MLE-Kriging have poor fitting precision, especially for CON-Kriging; it had oscillated badly. This illustrated kernel function optimization is very important for improving fitting precision of surrogate model. Compared with CON-Kriging and MLE-Kriging, SRM-Kriging is almost closed to original function. It has better generalization ability than others.

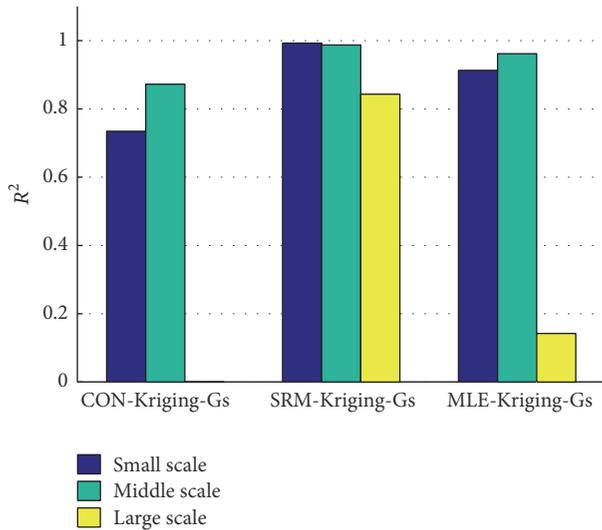
4.3. *Samples with Nonuniform Distribution.* In a practical application, the samples are always not satisfied uniform distribution standard. This will do some impact to surrogate model building. A typical example is the sequential surrogate model [25]; as adding points goes on, the distribution of samples changes continuously; an aggregation distribution of samples has appeared. The performance of surrogate model with nonuniform distribution samples is also important. At

TABLE 1: Sampling information of the test functions.

Test function	Dimension	Bound	Samples size	Test sample size	Scale
Fun1	1	[-6, 6]	25	600	Small
Fun2	1	[-12, 12]	25	600	Small
Fun3	2	[-3, 3]	50	1 000	Small
Fun4	2	[-5, 5]	50	1 000	Small
Fun5	10	[-10, 11]	300	2 000	Middle
Fun6	10	[1, 5]	300	2 000	Middle
Fun7	16	[-1, 1]	400	3 000	Middle
Fun8	30	[0, 10]	600	3 000	Large
Fun9	30	[-5, 5]	600	3 000	Large

TABLE 2: Comparison of the generalization ability between different test functions with uniform distribution sampling.

Test functions	R^2			RMSE		
	CON-Kriging	SRM-Kriging	MLE-Kriging	CON-Kriging	SRM-Kriging	MLE-Kriging
Fun1	Failure	0.999 39	0.672 59	0.372 85	0.005 82	0.134 61
Fun2	1.000 00	1.000 00	1.000 00	0.000 00	0.000 00	0.000 00
Fun3	0.988 30	0.991 93	0.999 20	1.197 99	0.994 75	0.313 37
Fun4	0.950 40	0.980 54	0.979 91	103.012 78	64.513 05	65.463 16
Fun5	0.894 21	0.993 22	0.986 45	326.976 85	82.749 71	117.016 48
Fun6	0.999 93	0.999 99	0.999 98	0.694 53	0.263 94	0.341 66
Fun7	0.723 31	0.969 00	0.899 10	11.309 11	3.783 29	6.824 92
Fun8	Failure	0.931 03	0.105 14	168.556 75	42.903 68	154.449 27
Fun9	Failure	0.755 16	0.179 30	1 010.540 9	499.434 93	913.770 52

FIGURE 1: Comparison of R^2 between different scale of test functions with uniform distribution sampling.

present there are a few researches for surrogate model with nonuniform distribution samples [26]. In this section, normal distribution is used to imitate and generate nonuniform distribution samples. The above three surrogate models are

tested using these nonuniform distribution points. The mean and standard deviation of the normal distribution of test functions are listed in Table 3.

30 times of random normal distribution sampling are generated to test these three methods and the average value of R^2 and RMSE is used to make a comparison. Figure 3 and Table 4 are the comparison of the generalization ability between different test functions with nonuniform distribution sampling.

It is observed that, for the test cases with nonuniform distribution sampling, the fitting precision of all methods decreased remarkably, especially for large-scale problem. CON-Kriging has very poor performance, even for small-scale problems. MLE-Kriging is better than CON-Kriging. SRM-Kriging still has the best fitting precision than the others. Although all three methods are failed for large-scale problems, but SRM-Kriging still has the smallest RMSE.

Figure 4 shows the fitting results of the Fun1 and Fun2 with nonuniform distribution sampling. From the fitting result we knew that all fitting curves for Fun1 and Fun2 are distorted. For Fun1, the fitting curves generated by CON-Kriging and MLE-Kriging even cannot fit the basic trends; SRM-Kriging obviously is better than the others and is in line with the basic trends except the left region with little samples. It has been proved that generalization ability is more suitable for evaluating comprehensive performance of surrogate model. For Fun2, three methods are all distorted at the right

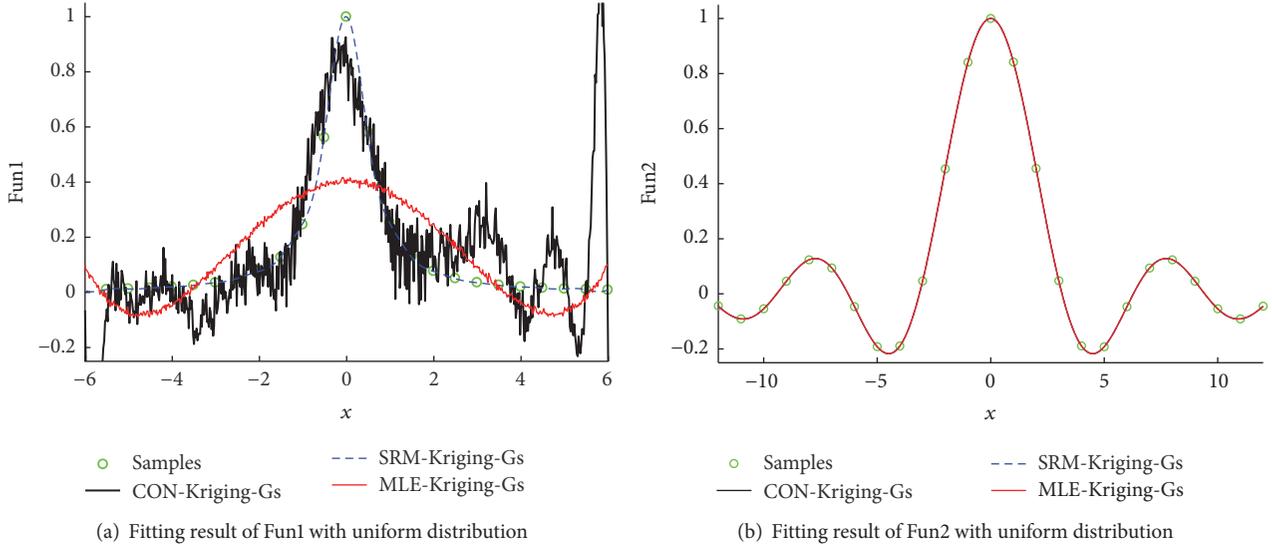


FIGURE 2: Fitting results of the Fun1 and Fun2 with uniform distribution sampling.

TABLE 3: Mean and standard deviation of the normal distribution of test functions.

Test functions	Dimension	Bound	Mean	Standard deviation
Fun1	1	[-6, 6]	0.0	3.0
Fun2	1	[-12, 12]	0.0	6.0
Fun3	2	[-3, 3]	0.0	1.5
Fun4	2	[-5, 5]	0.0	2.5
Fun5	10	[-10, 11]	0.0	5.0
Fun6	10	[1, 5]	3.0	1.0
Fun7	16	[-1, 1]	0.0	0.25
Fun8	30	[-10, 10]	0.0	5.0
Fun9	30	[-5, 5]	0.0	2.5

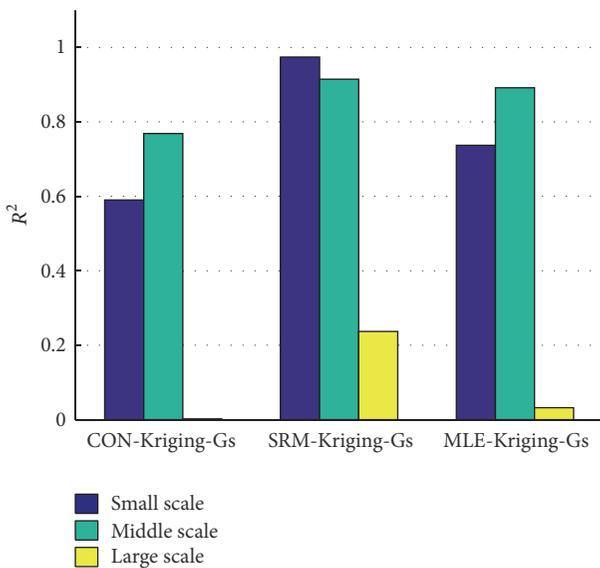


FIGURE 3: Comparison of R^2 between different scale of test functions with nonuniform distribution sampling.

region with less samples, but the fitting performance of SRM-Kriging and MLE-Kriging is much better than CON-Kriging, which proves the importance of kernel parameter optimization.

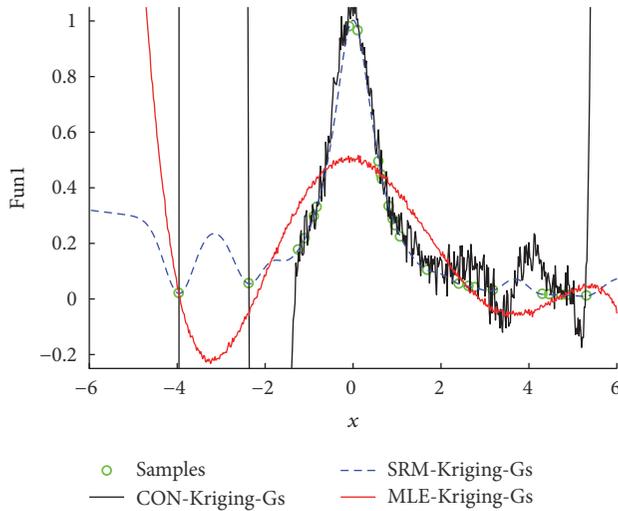
5. Conclusion

More and more attention is paid to the improvement of fitting precision as the widely spread and applied surrogate model. This paper studies the assessment criteria of surrogate model and discusses the importance of the generalization ability. Based on these, the kernel function optimization method is proposed to improve the fitting precision. Some standard benchmarks are tested to verify the effectiveness of the improved method based on sample points with uniform distribution and nonuniform distribution. In conclusion, we have the following.

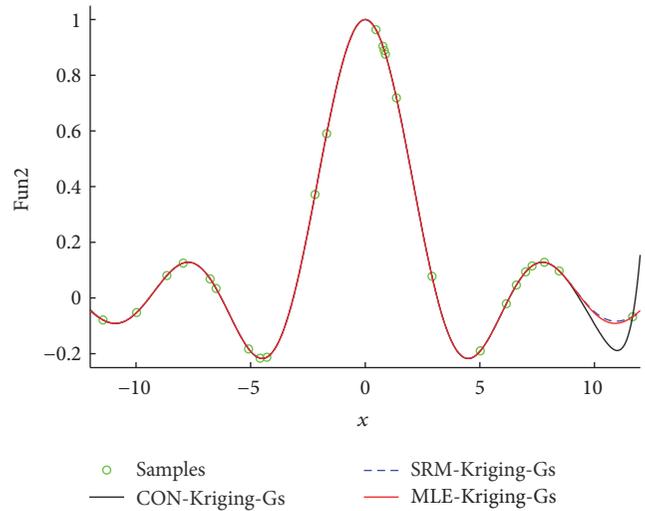
- (1) SRM-Kriging method based on anisotropic kernel function optimization is proposed to replace the coefficient of each samples with the coefficient of the component of samples. The computational cost is reduced.

TABLE 4: Comparison of the generalization ability between different test functions with nonuniform distribution sampling.

Test functions	R^2			RMSE		
	CON-Kriging	SRM-Kriging	MLE-Kriging	CON-Kriging	SRM-Kriging	MLE-Kriging
Fun1	Failure	0.955 64	failure	985.97 356	0.049 79	0.564 43
Fun2	1.000 00	0.999 99	0.999 99	0.000 15	0.000 83	0.000 79
Fun3	0.886 67	0.957 91	0.948 37	3.753 33	2.293 57	2.347 32
Fun4	0.474 05	0.983 03	0.999 97	336.163 41	60.180 79	2.715 26
Fun5	0.882 14	0.996 40	0.997 44	354.549 56	61.200 31	61.109 76
Fun6	0.999 90	0.999 98	0.999 99	0.840 81	0.378 94	0.318 47
Fun7	0.424 08	0.747 45	0.677 50	21.787 80	11.903 82	13.896 88
Fun8	Failure	0.338 04	0.009 51	457.409 39	190.136 04	394.725 25
Fun9	0.004 02	0.135 85	0.054 74	2 385.597 16	1 506.804 29	1 889.119 78



(a) The fitting result of Fun1 with nonuniform distribution



(b) The fitting result of Fun2 with nonuniform distribution

FIGURE 4: The fitting results of the Fun1 and Fun2 with nonuniform distribution sampling.

- (2) A comparison is carried out among CON-Kriging, SRM-Kriging, and MLE-Kriging. Results show that kernel function optimization is very important to improve the fitting precision of surrogate model. SRM principle provides a more effective evaluation for the generalization ability. SRM-Kriging has better performance than the others, especially for problems with nonuniform distribution sampling.
- (3) From the Kriging process we know that Kriging can be regarded as the special case of the SVR with zero empirical risk. By setting the insensitive loss function of SVR to zero, the SVR optimization problem with regularization parameters and kernel parameters is transformed to a simple optimization problem with independent kernel parameters and a certain problem to evaluate coefficients. It reduces the design dimension and simplifies the implement process.

The distribution of samples may become more and more complicated. This paper also studies the fitting precision of Kriging model with uniform sampling and nonuniform

sampling. The results show that Kriging with SRM principle has better fitting performance, but still with low precision for high dimension problems. Further research may perform full analysis of surrogate model with different nonuniform distribution sampling to find more suitable selection mechanism of the kernel function and efficient kernel parameter optimization method.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by funding from the National Natural Science Foundation of China (no. 51505385), the Shanghai Aerospace Science Technology Innovation Foundation (no. SAST2015010), and the Defense Basic Research Program (no. JCKY2016204B102 and no. JCKY2016208C001). The authors are also thankful to Shaanxi Aerospace Flight Vehicle Design Key Laboratory of NPU.

References

- [1] M. Ma, C. Wang, J. ZHANG, and Z. Huang, "Multidisciplinary Design Optimization for Complex Product Review," *Chinese Journal of Mechanical Engineering*, vol. 44, no. 06, pp. 15–26, 2008.
- [2] O. de Weck, J. Agte, J. Sobieszcanski-Sobieski, P. Arendsen, A. Morris, and M. Spieck, "State-of-the-Art and Future Trends in Multidisciplinary Design Optimization," in *Proceedings of the 48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Honolulu, Hawaii, USA, 2007.
- [3] D. G. Ullman, "Toward the ideal mechanical engineering design support system," *Research in Engineering Design*, vol. 13, no. 2, pp. 55–64, 2002.
- [4] V. N. Vapnik, *Statistical Learning Theory*, Wiley, 1998.
- [5] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.
- [6] W. C. M. Van Beers, "Kriging metamodeling in discrete-event simulation: An overview," in *Proceedings of the 2005 Winter Simulation Conference*, pp. 202–208, Orlando, Fla, USA, December 2005.
- [7] T. W. Simpson, T. M. Mauery, J. J. Korte, and F. Mistree, "Kriging models for global approximation in simulation-based multidisciplinary design optimization," *AIAA Journal*, vol. 39, no. 12, pp. 2233–2241, 2001.
- [8] N. Cressie, "Spatial prediction and ordinary kriging," *Mathematical Geology*, vol. 21, no. 4, pp. 493–494, 1989.
- [9] J. P. Kleijnen, "Kriging metamodeling in simulation: a review," *European Journal of Operational Research*, vol. 192, no. 3, pp. 707–716, 2009.
- [10] S. Jeong, M. Murayama, and K. Yamamoto, "Efficient optimization design method using kriging model," *Journal of Aircraft*, vol. 42, no. 2, pp. 413–420, 2005.
- [11] J. P. Kleijnen, "Simulation-optimization via Kriging and bootstrapping: a survey," *Journal of Simulation*, vol. 8, no. 4, pp. 241–250, 2014.
- [12] Z. Chen, H. Qiu, L. Gao, X. Li, and P. Li, "A local adaptive sampling method for reliability-based design optimization using Kriging model," *Structural and Multidisciplinary Optimization*, vol. 49, no. 3, pp. 401–416, 2014.
- [13] S. Ulaganathan, I. Couckuyt, T. Dhaene, J. Degroote, and E. Laermans, "Performance study of gradient-enhanced Kriging," *Engineering with Computers*, vol. 32, no. 1, pp. 15–34, 2016.
- [14] P. Goovaerts, "Ordinary Cokriging Revisited," *Mathematical Geology*, vol. 30, no. 1, pp. 21–40, 1998.
- [15] Z.-H. Han and S. Görtz, "Hierarchical kriging model for variable-fidelity surrogate modeling," *AIAA Journal*, vol. 50, no. 9, pp. 1885–1896, 2012.
- [16] J. Xiao, L. Yu, and Y. F. Bai, "Survey of the Selection of Kernels and Hyper-parameters in Support Vector Regression," *Journal of Southwest Jiaotong University*, vol. 43, no. 3, pp. 297–303, 2008.
- [17] X. Liu, S. Zhou, Z. Xiong, L. Chen, and C. Yan, "Research on Identifying TS Fuzzy Model Based on Structural Risk Minimization," *Journal of Guizhou University*, 2016.
- [18] D. Berestin, M. Zimin, T. Gavrilenko, and N. Chernikov, "Determination of an object as chaotic system on the basis of structural risk minimization," *Complexity. Mind. Postnonclassic*, vol. 3, no. 4, pp. 73–87, 2014.
- [19] X. Zhu, W. Luo, Y. Wei, and X. Chen, "Structural Risk Minimization Based Radial Basis Function Surrogate Mode," *Journal of Projectiles, Rockets, Missile and Guidance*, vol. 31, no. 5, pp. 169–173, 2011.
- [20] C. Wei, L. I. Qian, and T. Wei, "Radial Basis Function Networks Based on Structural Risk Minimization Principle," *Control and Instruments in Chemical Industry*, vol. 36, no. 3, pp. 34–37, 2009.
- [21] I. J. F. Alexander, S. Andras, and J. K. Andy, *Engineering Design via Surrogate Modelling*, Wiley, 2008.
- [22] X. F. Zhu, *Surrogate Model Theory and Application in the Aircraft MDO*, National University of Defense Technology, 2010.
- [23] S. N. Lophaven, H. B. Nielsen, and J. Søndergaard, "DACE A MATLAB Kriging Toolbox," Tech. Rep., Technical University of Denmark, 2002.
- [24] Z. C. Xiong, *Some Problems in Approximation Theory with Radial Basis Function*, Fudan University, 2007.
- [25] B. Glaz, *Active/passive optimization of helicopter rotor blades for improved vibration, noise, and performance characteristics*, University of Michigan, 2008.
- [26] D. Zhang, Z. Gao, J. Li, and L. Huang, "Study of Metamodel Sampling Criterion," *Acta Aerodynamica Sinica*, vol. 29, no. 6, pp. 719–725, 2011.




Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

