8th World Congress on Structural and Multidisciplinary Optimization
June 1 - 5, 2009, Lisbon, Portugal

Sequential-Sampling-Based Kriging Method with Dynamic Basis Selection

Liang Zhao[1], K.K. Choi[2], Ikjin Lee[3] and David Gorsich[4]

[1,2,3]Department of Mechanical & Industrial Engineering
College of Engineering
The University of Iowa
Iowa City, IA 52242, U.S.A.
Email: liazhao@engineering.uiowa.edu
kkchoi@engineering.uiowa.edu
ilee@engineering.uiowa.edu

[4]US Army RDECOM/TARDEC
Warren, MI 48397-5000, U.S.A.
gorsichd@tacom.army.mil

1. Abstract

Over three decades, metamodeling has been widely applied to design optimization problems to build a surrogate model of computation-intensive engineering models. The Kriging method has gained significant interests for developing the surrogate model. However, traditional Kriging methods, including the ordinary Kriging and the universal Kriging, use fixed polynomials basis functions to generate the mean structure. In this paper, a new Dynamic Kriging (D-Kriging) method is proposed to fit the true model more accurately. In this D-Kriging method, the mean structure is automatically decided by applying genetic algorithm based feature selection to the candidate basis functions based on a new accuracy criterion. In addition, a new sequential sampling technique based on the prediction interval of the surrogate model is proposed and integrated into the D-Kriging method. Numerical examples show much more accurate results from D-Kriging compared with traditional Kriging methods.

2. Keywords: Response Surface Method (RSM), Kriging Method, Prediction Interval, Sequential Sampling Method, Dynamic Basis Selection, Genetic Algorithm.

3. Introduction

Metamodeling has been widely used in engineering applications when a true experiment is not feasible or is extremely hard to obtain due to high computational cost. A surrogate model is desirable for representing the true model when only a limited number of experiments can be evaluated. Researchers have been investigating various methods for generating the surrogate model based on limited samples. A number of methods, such as the least square regression, moving least square regression and radial basis functions, have been developed in recent decades [1-6]. Recently, the Kriging method has gained large interest due to its capability of dealing with highly nonlinear model [7]. In the Kriging method, the response of the model is considered as two parts: the mean structure and the residue. The ordinary Kriging (O-Kriging) assumes this mean structure part is zero or a constant among the entire domain [8]. The universal Kriging (U-Kriging) usually considers the mean structure as first- or second-order polynomials, which are obtained from a generalized least square regression [9]. However, during the practical use of these methods, a problem has been discovered: neither the ordinary Kriging nor the universal Kriging maximally uses the information from the evaluated samples due to the fixed form of the mean structure. Therefore, a new method that can automatically adjust the mean structure and maximally use the information based on current samples is needed. Joseph [10] used a Bayesian framework to identify the mean structure for the Kriging method. In this paper, we propose a new method to automatically decide the mean structure of the Kriging model by applying a feature-selection process based on a new criterion.

Another crucial issue of metamodeling is the sampling strategy. The Latin hypercube sampling method (LHS) [11,12] has been applied in metamodeling. It tries to occupy the entire design domain most evenly and gain as much information about the true model as it can. However, it is not a problem-specified method, which means that no matter what the true response is, it always give us a similar sample profile that occupies the entire domain evenly. This could be a critical problem if the distribution of the nonlinear area is aggregating in particular part of the domain. Another sampling technique, importance sampling [13], samples around the limit state area and predicts the response accurately around the limit state. This importance sampling method also only gives a good local surrogate model around the limit state area and usually does not represent the true model accurately enough in

other areas of the domain. A sequential sampling strategy is applied by Wang [14] and Jin et al [15] to identify the sample position in a sequential manner. In this paper, we propose a new sequential sampling strategy integrated with the proposed D-Kriging method. By coupling the sampling method with the D-Kriging method, the efficiency and accuracy can be significantly achieved. Mathematical examples show the promising results of the surrogate model constructed by this sequential-sampling-based Kriging method with dynamic basis selection.

4. Kriging Method with Dynamic Basis Selection

4.1 Kriging Method

The Kriging method has gained large interest for generating the response surface in recent years. In the Kriging method, the outcomes are considered as a realization of a stochastic process and the predicted values are derived later by applying stochastic process theory. Consider n sample points: $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_n]^{\mathrm{T}}$ with $\mathbf{x}_i \in \mathbf{R}^{\mathrm{m}}$, and n responses $\mathbf{Y}=[y(\mathbf{x}_1), y(\mathbf{x}_2),..., y(\mathbf{x}_n)]^{\mathrm{T}}$ with $y(\mathbf{x}_i)\in\mathbf{R}$. In the Kriging method, the response at samples is considered as a summation of two parts as

$$\mathbf{Y}=\mathbf{F}\boldsymbol{\beta}+\mathbf{e} \tag{1}$$

The first part of the right-hand side of Eq. (1), $\mathbf{F}\boldsymbol{\beta}$, is considered as the mean structure of the response, where $\mathbf{F}=[f_k(\mathbf{x}_i), i=1,...,n,\ k=1,...,K]$ is a $(n\times K)$ design matrix, and $f_k(\mathbf{x})$ represents user-defined basis functions, which are usually in a simple polynomial form, such as $1, x, x^2,...$. In Eq. (1), $\boldsymbol{\beta}=[\beta_1,\beta_2,...,\beta_K]^{\mathrm{T}}$ are the regression coefficients from the generalized least squares regression method. The second part of Eq. (1), $\mathbf{e}=[e(\mathbf{x}_1), e(\mathbf{x}_2),..., e(\mathbf{x}_n)]^{\mathrm{T}}$, is a realization of the stochastic process $e(\mathbf{x})$ that is assumed to have zero mean and covariance structure $\mathbf{E}[e(\mathbf{x}_i)e(\mathbf{x}_j)]=\sigma^2 R(\boldsymbol{\theta},\mathbf{x}_i,\mathbf{x}_j)$, where σ^2 is the process variance, $\boldsymbol{\theta}$ is the process parameter which has to be estimated from sample data and $R(\boldsymbol{\theta},\mathbf{x}_i,\mathbf{x}_j)$ is the correlation function of the stochastic process. For a multidimensional problem, it becomes the multiplication of the correlation functions for each dimension as $R(\boldsymbol{\theta},\mathbf{x}_i,\mathbf{x}_j)=\prod_{d=1}^{m} R_d(\theta^d, x_i^d, x_j^d)$, where x_i^d is the d^{th} component of $\mathbf{x}_i$ and θ^d is the d^{th} component of $\boldsymbol{\theta}$. For example, if the process is assumed to be a stationary Gaussian process, then $R_d(\theta^d, x_i^d, x_j^d)=\exp(-\theta^d \mid x_i^d - x_j^d \mid^2)$. Another assumption of the model in Eq. (1) is, $Cov[\mathbf{F}(\mathbf{x}), e(\mathbf{x})]=0$; that is, the residuals are uncorrelated with the basis functions. Under the general decomposition of Eq. (1), the objective is to predict the noise-free unbiased response at a new point of interest $\mathbf{x}_0$. In the Kriging method, this prediction of response is written as a linear predictor as

$$\hat{y}(\mathbf{x}_0)=\mathbf{w}_0^{\mathrm{T}}\mathbf{Y} \tag{2}$$

where $\mathbf{w}_0=[w_1(\mathbf{x}_0), w_2(\mathbf{x}_0),..., w_n(\mathbf{x}_0)]^{\mathrm{T}}$ denotes the $(n\times 1)$ weight vector for prediction at $\mathbf{x}_0$. Using Eq. (2), the unbiased prediction condition $\mathbf{E}[\hat{y}(\mathbf{x}_0)]=\mathbf{E}[y(\mathbf{x}_0)]$ is expressed as

$$\begin{aligned}\mathbf{E}[\hat{y}(\mathbf{x}_0)-y(\mathbf{x}_0)] &= \mathbf{E}[\mathbf{w}_0^{\mathrm{T}}\mathbf{Y}-y(\mathbf{x}_0)]\\ &= \mathbf{E}[\mathbf{w}_0^{\mathrm{T}}(\mathbf{F}\boldsymbol{\beta}+\mathbf{e})-(\mathbf{f}_0\boldsymbol{\beta}+e(\mathbf{x}_0))]\\ &= \mathbf{E}[\mathbf{w}_0^{\mathrm{T}}\mathbf{e}-e(\mathbf{x}_0)+(\mathbf{F}^{\mathrm{T}}\mathbf{w}_0-\mathbf{f}_0^{\mathrm{T}})^{\mathrm{T}}\boldsymbol{\beta}]\\ &= \mathbf{E}[(\mathbf{F}^{\mathrm{T}}\mathbf{w}_0-\mathbf{f}_0^{\mathrm{T}})^{\mathrm{T}}\boldsymbol{\beta}]=\mathbf{0}\end{aligned} \tag{3}$$

where $\mathbf{f}_0=[f_1(\mathbf{x}_0), f_2(\mathbf{x}_0),..., f_K(\mathbf{x}_0)]^{\mathrm{T}}$. Therefore, the unbiased condition is ensured by imposing the constraint $\mathbf{F}^{\mathrm{T}}\mathbf{w}_0=\mathbf{f}_0^{\mathrm{T}}$ on the prediction weights for each point of interest [16].

Under this constraint, $\mathbf{w}_0$ is obtained by solving

$$\begin{bmatrix}\mathbf{R} & \mathbf{F}\\ \mathbf{F}^{\mathrm{T}} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{w}_0\\ \boldsymbol{\lambda}\end{bmatrix}=\begin{bmatrix}\mathbf{r}_0\\ \mathbf{f}_0\end{bmatrix}, \tag{4}$$

which represents the Lagrangian first-order necessary conditions of minimizing the mean squared error (MSE) of the prediction [16], where $\boldsymbol{\lambda}$ represents the Lagrangian multipliers, $\mathbf{R}$ is the correlation matrix $(R)_{ij}=R(\boldsymbol{\theta},\mathbf{x}_i,\mathbf{x}_j),\ i,j=1,...,n$, and $\mathbf{r}_0=[R(\boldsymbol{\theta},\mathbf{x}_1,\mathbf{x}_0),...,R(\boldsymbol{\theta},\mathbf{x}_n,\mathbf{x}_0)]^{\mathrm{T}}$ is the correlation vector between the prediction location $\mathbf{x}_0$ and all n samples $\mathbf{x}_i$, $i=1,...,n$. The solution of Eq. (4) is given by

$$\begin{aligned}\boldsymbol{\lambda} &= (\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F})^{-1}(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{r}_0-\mathbf{f}_0)\\ \mathbf{w}_0 &= \mathbf{R}^{-1}(\mathbf{r}_0-\mathbf{F}\boldsymbol{\lambda})\end{aligned} \tag{5}$$

Hence the prediction is expressed as

$$\begin{aligned}\hat{y}(\mathbf{x}_0) &= \mathbf{w}_0^{\mathrm{T}}\mathbf{Y} = (\mathbf{r}_0 - \mathbf{F}\boldsymbol{\lambda})^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y} \\ &= \mathbf{r}_0^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y} - (\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{r}_0 - \mathbf{f}_0)^{\mathrm{T}}(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y} \\ &= \mathbf{f}_0^{\mathrm{T}}\boldsymbol{\beta}^* + \mathbf{r}_0^{\mathrm{T}}\boldsymbol{\gamma}^*\end{aligned} \tag{6}$$

where $\boldsymbol{\beta}^* = (\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y}$, $\boldsymbol{\gamma}^* = \mathbf{R}^{-1}(\mathbf{Y} - \mathbf{F}\boldsymbol{\beta}^*)$, and the prediction error variance [17] is expressed as

$$\sigma_p^2(\mathbf{x}_0) = \mathbf{Var}[\hat{y}(\mathbf{x}_0) - y(\mathbf{x}_0)] = \sigma^2(1 + \mathbf{w}_0^{\mathrm{T}}\mathbf{R}\mathbf{w}_0 - 2\mathbf{w}_0^{\mathrm{T}}\mathbf{r}_0) \tag{7}$$

Also, the sensitivity of the predicted response is given by

$$\hat{y}'(\mathbf{x}_0) = \left[\frac{\partial y}{\partial x_1}, ..., \frac{\partial y}{\partial x_n}\right]^{\mathrm{T}} = J_f(\mathbf{x}_0)^{\mathrm{T}}\boldsymbol{\beta}^* + J_r(\mathbf{x}_0)^{\mathrm{T}}\boldsymbol{\gamma}^* \tag{8}$$

where $(J_f(\mathbf{x}_0))_{ij} = \dfrac{\partial f_i(\mathbf{x}_0)}{\delta x_j}$ and $(J_r(\mathbf{x}_0))_{ij} = \dfrac{\partial R(\boldsymbol{\theta}, \mathbf{x}_i, \mathbf{x}_0)}{\partial x_j}$ are the Jacobians of $\mathbf{f}_0$ and $\mathbf{r}_0$, respectively [17].

According to Eqs. (6) and (8), the function value and sensitivity of response are obtained as long as the correlation matrix $\mathbf{R}(\boldsymbol{\theta}, \mathbf{x}_i, \mathbf{x}_j)$ is available. In this paper, $e(\mathbf{x})$ is assumed to be a Gaussian process, hence $R_d(\theta^d, x_i^d, x_j^d) = \exp(-\theta^d \mid x_i^d - x_j^d \mid^2)$. The optimum value of process parameter $\boldsymbol{\theta}$ is defined as the maximum likelihood estimator (MLE), which is the maximizer of $-\frac{1}{2}(n\ln\sigma^2 + \ln|\mathbf{R}(\boldsymbol{\theta})|)$, and it is equivalent with the formulation as

$$\boldsymbol{\theta} \text{ is the minimizer of } \quad \psi(\boldsymbol{\theta}) = |\mathbf{R}(\boldsymbol{\theta})|^{\frac{1}{n}} \cdot \sigma^2 \tag{9}$$

where the process variance σ^2 is estimated by $\sigma^2 = \dfrac{1}{n}(\mathbf{Y} - \mathbf{F}\boldsymbol{\beta}^*)^{\mathrm{T}}(\mathbf{Y} - \mathbf{F}\boldsymbol{\beta}^*)$ [2]. Under the assumption of Gaussian process, the α-level prediction interval of response is written as

$$\hat{y}(\mathbf{x}_0) - Z_{(1+\alpha)/2}\sigma_p(\mathbf{x}_0) \le y(\mathbf{x}_0) \le y(\mathbf{x}_0) + Z_{(1+\alpha)/2}\sigma_p(\mathbf{x}_0) \tag{10}$$

where $Z_{(1+\alpha)/2}$ is the α-level quantile of standard normal distribution. Therefore, the bandwidth of the prediction interval at point of interest $\mathbf{x}_0$ is

$$d(\mathbf{x}_0) = 2Z_{(1+\alpha)/2}\sigma_p(\mathbf{x}_0) \tag{11}$$

4.2 Feature Selection: Dynamic Regression Basis Selections

In the universal Kriging method, the basis function $f_k(\mathbf{x})$ used in Eq. (1) is fixed along the entire metamodeling process, and usually it takes up to second order, i.e., x^2. However, it is obvious that higher-order terms can catch up more nonlinear trend in the mean structure. Hence in many highly nonlinear cases, fixed-order basis functions are not enough to describe the nonlinearity of the mean structure. On the other hand, researchers also observe that in some cases the accuracy of the surrogate model may not be necessarily enhanced by introducing higher-order terms; more straightforwardly, the surrogate model may become even worse. Therefore, the problem becomes how to find the optimal set of these basis functions such that the surrogate model would be most accurate. It is clear that the global optimal subset of these candidate basis functions can be obtained only by applying the exhaustive algorithm. The number of total tests would be 2^M, where *M* is the number of candidate basis functions. Consequently the computational expense may increase to become unaffordable. A common way of deciding which candidate basis functions should be included is called sequential feature selection [18]. However, we found out this sequential feature selection can easily hit the local optimum and therefore may not perform best for selection purpose. Therefore, we used another way of carrying out this selection process which is based on genetic algorithm for feature selection which was also used by Broadhurst for variable selection [19].

In this paper, all the candidate basis functions are simple polynomials and their multiplications, in the form of $x_1^{k_1}x_2^{k_2}...x_i^{k_i}$, $i = 1, 2, ..., D$, where *D* is the number of input variables, $k_i = 0, 1, .., P$ is the power of x_i, $\sum_{i=1}^{n} k_i \le P$, and *P* is the highest order of the mean structure in the current surrogate model. The total number of possible candidate basis functions is C_{D+P}^{P}. Therefore, the full set of row vector of **F** in Eq. (1) becomes

$$[1, x_1, x_2, ..., x_D, x_1^2, ..., x_D^2, x_1x_2, ..., x_{D-1}x_D, x_1^3, ..., x_D^3, x_1^2x_2, ..., x_1^P, x_1^{P-1}x_2, ...]_{1\times C_{D+P}^{P}} \tag{12}$$

Before applying the genetic algorithm based feature selection, a constraint must be satisfied. That is **F** cannot be underdetermined; i.e., the total number of possible candidate basis functions cannot be larger than the number of samples. This constraint also determines the highest order *P* by

$$\begin{aligned} &\max\{P\} \\ &s.t\ C_{D+P}^{P} \le n \\ &\textit{where n is the number of samples} \end{aligned} \tag{13}$$

After finding the highest order P, $\mathbf{F}$ becomes an $N \times C_{D+P}^{P}$ matrix. Then we apply the genetic algorithm for feature selection to $\mathbf{F}$ and decide which column (basis function) should be kept for the surrogate model. To carry out the genetic algorithm for feature selection, first we need an objective function to decide if the current subset is a good one or not.

An objective function, called the criterion, which the method seeks to minimize over all feasible feature subsets. Usually this criterion is mean-squared-error (MSE) for regression models from a generalized least square (GLM) regression. However we found that the MSE from GLM is not suitable for the Kriging method. It may lead to a wrong selection in many cases, which means the surrogate model from the Kriging method becomes worse. To make the feature selection procedure compatible with the Kriging method, we propose a new criterion in this paper, which is defined as the average value of the ratio between the prediction interval bandwidth and the predicted response, given by

$$C = \frac{1}{K}\sum_{i=1}^{K}\frac{d(\mathbf{x}_i)}{|\hat{y}(\mathbf{x}_i)|} \tag{14}$$

where K is the number of checking points $\mathbf{x}_i$. Usually the testing points are chosen from a $\sqrt{K}\times\sqrt{K}$ mesh of the entire design domain. To calculate this C criterion during the genetic generation process, we calculate the Kriging response based on the current selected subset by applying Eqs. (1) – (11). It is obvious that this C criterion can also be used as the accuracy measure of the surrogate model in the design domain. Therefore, choosing C as the criterion for feature selection can make the feature selection result consistent with Kriging method for the surrogate model. Therefore, with this feature selection process, we can find a good subset, even if it may not be the best, for the Kriging method based on the current sample site.

4.3 Sequential Sampling Technique

We propose a sequential sampling technique using the bandwidth of the prediction interval determined by Eq. (11) to integrate it into the D-Kriging method for metamodeling purposes. First, the insertion criterion IC is defined as

$$IC(\mathbf{x}) = \frac{d(\mathbf{x})}{|\hat{y}(\mathbf{x})|} \tag{15}$$

The next sample point $\mathbf{x}_{new}$ is identified by

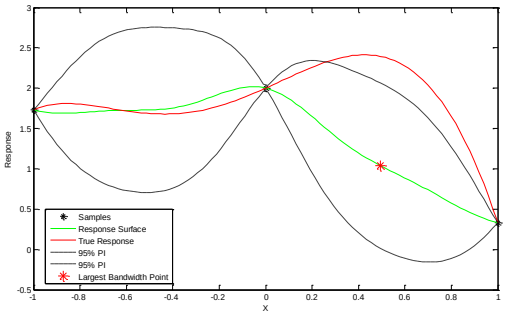
$$\max(IC(\mathbf{x}))\ \ \mathbf{x}\in\mathbf{\Omega},\ \ \mathbf{\Omega}: \textit{Design Domain} \tag{16}$$

By defining IC as Eq. (15) and enforcing $\mathbf{x}_{new}$ to be the maximizer of Eq. (16), we are trying to find the "weakest point" in the domain where we have the least confidence on the prediction.

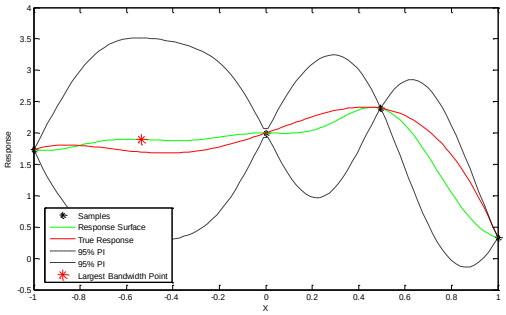
To demonstrate the process of this sequential sampling method, consider a *1-D* problem expressed as

$$y(x) = 0.5x^5 - 1.5x^4 - 2.5x^3 + 0.53x^2 + 1.3x + 2.0 \tag{17}$$

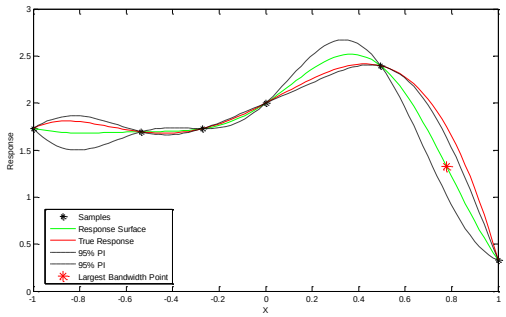
where $x \in [-1, 1]$. The sequential sampling method is initiated with three evenly distributed samples shown as Fig. 1(a). In the figures, the solid red line, the solid green line, and the two dashed black lines are the true response given by Eq. (17), the predicted surrogate model by Eq. (6), and the 95% prediction interval by Eq. (10), respectively. The black stars are the initial grid samples and the red star is the identified next inserted point within the domain. From Figs. 1(a)-(d), we can see that the bandwidth of the prediction interval keeps decreasing in fast speed and finally converges to the true response. A position where it has a large discrepancy between the true model and the surrogate model is identified during the iterations by applying the IC criterion.



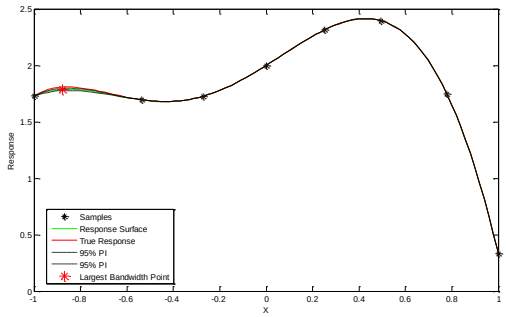
(a) Initial Stage: 3 Samples



(b) After 1 More Samples Inserted



(c) After 3 More Samples Inserted



(d) After 5 More Samples Inserted

Figure 1: Surrogate Model Using Sequential Sampling Method

To assess the accuracy of the surrogate model, the error definition is given by the average of relative error within the domain as

$$Err = \frac{1}{K}\sum_{i=1}^{K}\left|\frac{\sigma(\mathbf{x}_i)}{\hat{y}(\mathbf{x}_i)}\right| \tag{18}$$

where K is the number of checking points over the entire domain. Usually it takes a larger enough number, 10^4 in this paper. With this accuracy of the surrogate model, we can eventually decide how many samples are needed to generate an accurate surrogate model. The sequential sampling process will be continued until it meets the accuracy tolerance *Tol*, given as

$$RErr = \left|\frac{Err_{k-1} - Err_k}{Err_k}\right| < Tol \tag{19}$$

where Err_k is the error at the k^{th} iteration and $RErr$ is the relative error. Overall, the entire process of conducting this sequential-sampling-based D-Kriging (SS D-Kriging) is shown as in Fig. (2).

5. Mathematical Examples

5.1 Convergence Study

The proposed sequential sampling method can achieve the accuracy faster than other sampling methods, such as grid sampling and Latin hypercube sampling. A *2-D* profile example, which was used by Lee et al [20] as a highly nonlinear function for reliability-based design optimization, is given to demonstrate the fast convergence of this SS D-Kriging method. The true function is expressed as

$$\begin{aligned} y(x_1, x_2) = &-1 + (0.9063x_1 + 0.4226x_2 - 6)^2 + (0.9063x_1 + 0.4226x_2 - 6)^3 \\ &- 0.6 \times (0.9063x_1 + 0.4226x_2 - 6)^4 - (-0.4226x_1 + 0.9063x_2) \\ &\text{where } x_1 \in [5,9]\ x_2 \in [1,5.5] \end{aligned} \tag{20}$$

The SS D-Kriging method is initiated with five grid samples which are the red star points in Fig. (3). In the figures, the solid red line and the solid green line are the contour of the surrogate model and the contour of true response at $Y(\mathbf{X}) = 0$ given by Eq. (20), respectively. The green star point is the identified next inserting point. The blue star points are the sequentially inserted samples. The numbers next to the stars are the sequence of the inserted samples.

To assess the accuracy of the surrogate model, the MSE of the entire domain is calculated as the error measure, which is defined as

$$MSE = \frac{1}{K}\sum_{i=1}^{K}(y(x_i) - \hat{y}(x_i))^2 \tag{21}$$

and K is the number of the checking points which are grid samples evenly distributed over the entire domain. In this case, K equals to $100\times100 = 10^4$.

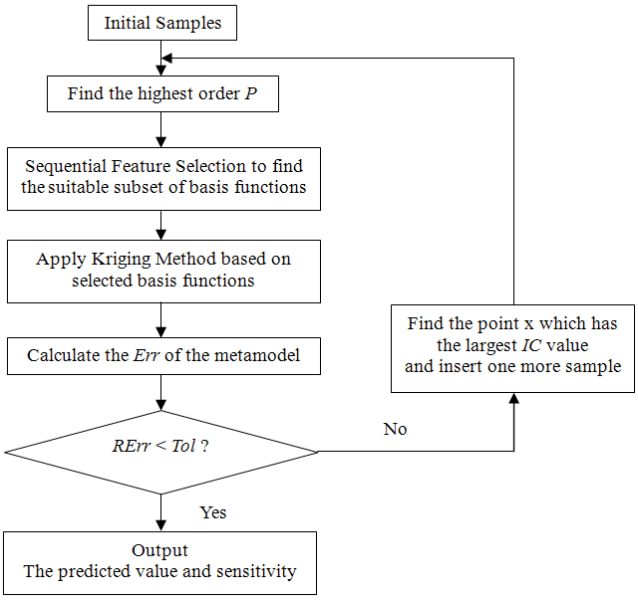
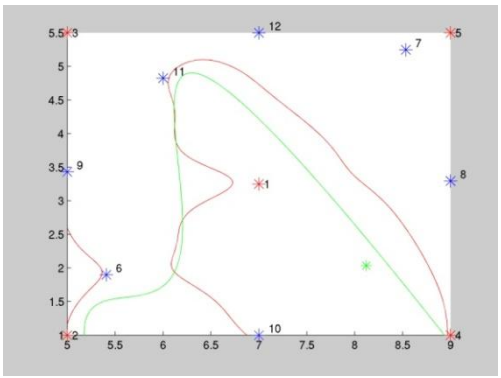
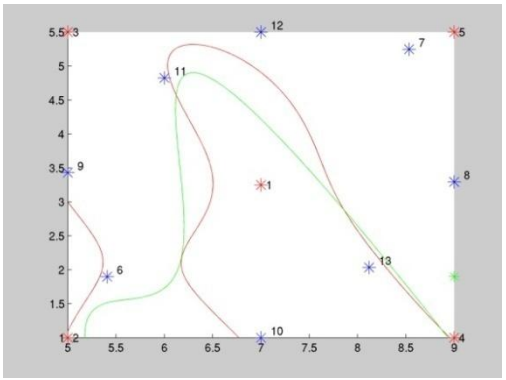


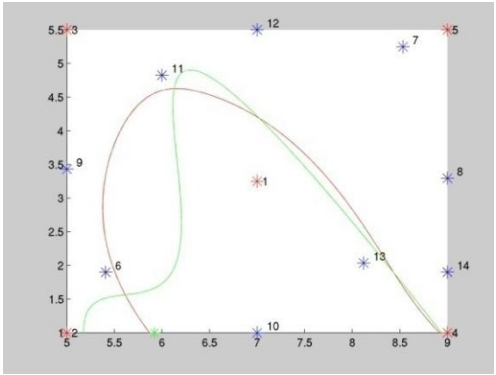
Figure 2: Entire Process of Sequential-Sampling-Based D-Kriging for Metamodeling



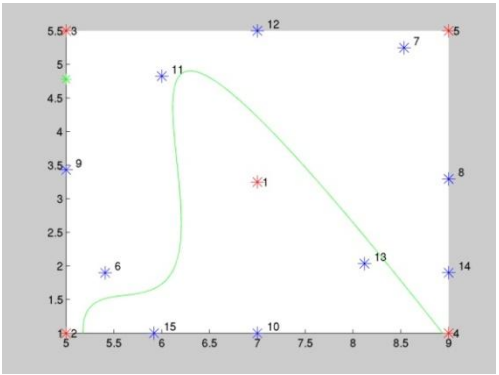
(a) 12 Samples MSE = 23.25



(b) 13 Samples MSE = 3.65



(c) 14 Samples MSE = 1.13



(d) 15 Samples MSE = 4.25E-27

Figure 3: Demonstration of Fast Convergence of SS D-Kriging Method

To show the comparison between the sequential sampling and Latin hypercube sampling with different Kriging methods, a convergence study is conducted. Since Latin hypercube sampling provides a different sample profile each time, 500 trials of Latin hypercube sampling are conducted, and the max, min, mean and median value of MSE from the surrogate models are calculated for comparison purposes. The overall convergence speed comparison is shown in Fig. (4). The SS D-Kriging converges faster than U-Kriging or D-Kriging with Latin hypercube sampling in the statistical sense. In details, we can see from Table.1 that, when the sample size is small (less than 10), the Latin hypercube gives a better surrogate model because it evenly distributes the samples over the entire domain and can have a relatively accurate result, but none of these results is acceptable since MSE is still larger than 100. When the sample size increases, the sequential sampling starts to work better because it identifies the "weak" region in the domain and increases the fidelity of the surrogate model by inserting samples in the weak region. Finally it converges to a very accurate result in faster speed than the other two methods. Moreover, when the sample size increases to 15, the number of cases in which MSE from SS D-Kriging is smaller than MSE of LHS with U-Kriging is 373 out of 500 trials; the number of cases in which MSE from SS D-Kriging is smaller than MSE of LHS with D-Kriging is 322 out of 500 trials. Both results show that sequential sampling works better than Latin hypercube sampling no matter which Kriging method is used.

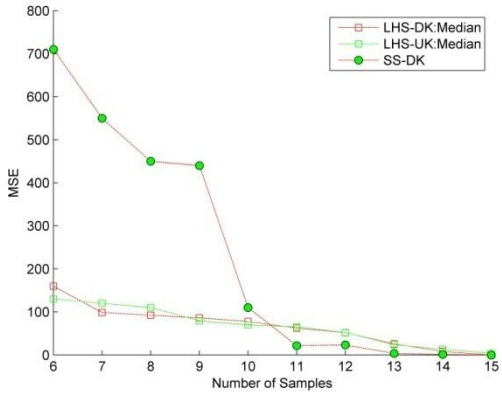


Figure 4: Convergence Study of SS D-Kriging Method

Table 1: Convergence Study of Sequential Sampling Method (MSE)

Methods			Number of Samples									
			6	7	8	9	10	11	12	13	14	15
LHS	U-Kriging	Min	55.1	49.9	44.8	37.1	11.3	5.9	2.2	1.6	0.9	0.45
		Max	1.4E6	9.4E3	1.7E3	2.3E2	1.8E2	2.5E2	1.6E2	1.6E2	1.5E2	1.5E2
		Mean	2.0E4	3.2E2	1.3E2	1.1E2	79.1	68.3	55.4	38.4	31.1	17.6
		Median	1.3E2	1.2E2	1.1E2	79.4	69.7	65.1	52.2	23.9	13.4	3.8
	D-Kriging	Min	52.1	48.7	42.6	27.9	21.3	5.8	2.6	1.5	0.3	4.3E-28
		Max	7.4E5	4.9E3	1.3E4	2.3E3	1.2E3	2.4E2	1.5E2	1.3E2	3.9E2	1.4E-10
		Mean	1.4E4	3.1E2	4.8E2	95.2	90.5	72.1	39.6	22.1	18.4	5.4E-12
		Median	1.6E2	98.6	92.2	86.1	77.7	62.3	31.3	11.2	9.5	5.6E-26
SS D-Kriging			7.1E2	5.5E2	4.5E2	4.4E2	1.1E2	21.6	23.3	3.7	1.1	4.3E-27

5.2 Coupling Effect of the Sequential Sampling Technique with the Kriging Methods

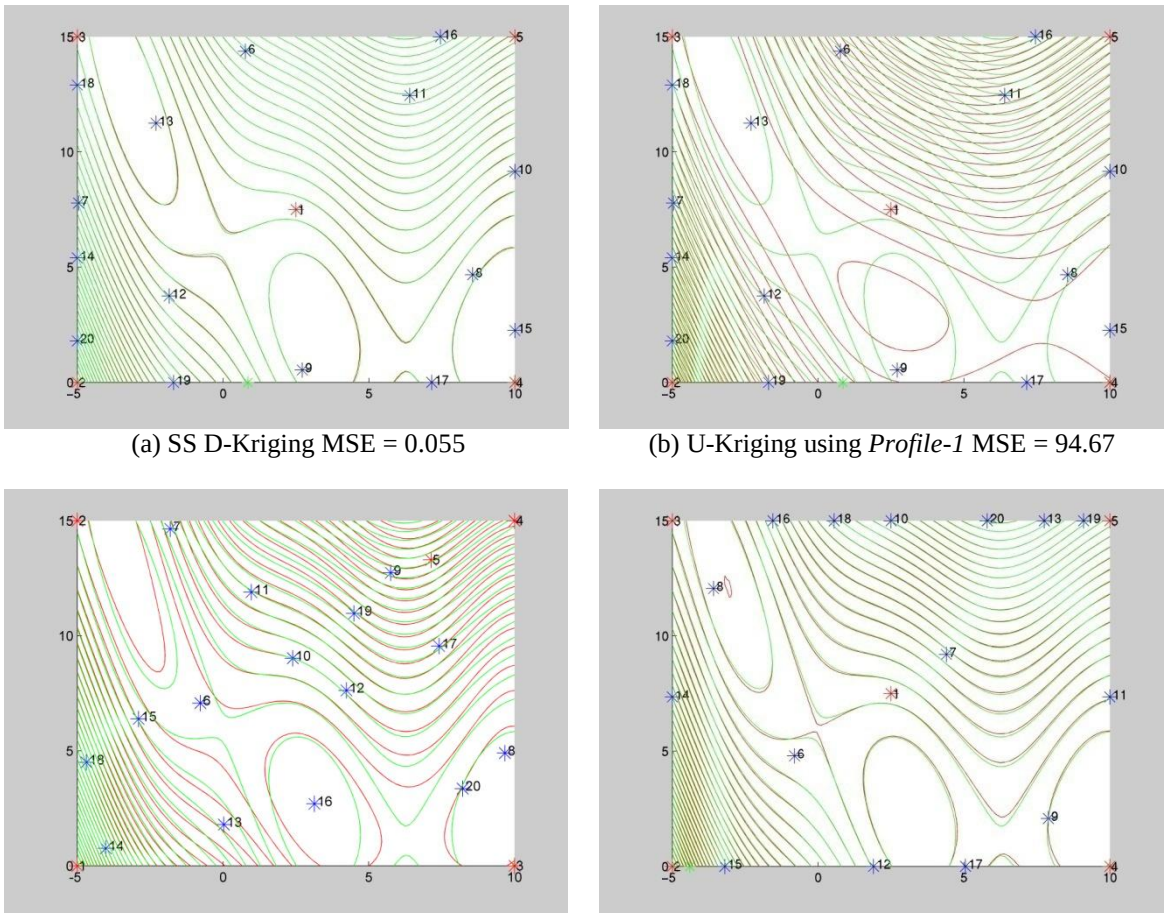
In practical cases, we also noticed that the improvement of the accuracy of the surrogate model is affected by the coupling effect from the combination of the sampling method and the metamodeling methods. Different Kriging methods will provide different sample profiles in the end and essentially will give quite different surrogate models. Combining the sequential sampling with the D-Kriging method provides the best final result. To demonstrate this effect, a more complex *2-D* mathematical example is given to compare the accuracy of D-Kriging and U-Kriging when different sampling strategies are applied. This example is so called Brian function which has been widely used as a benchmark problem for surrogate modeling purposes [21]. The Brian function is given as

$$f(x_1, x_2) = (x_2 - \frac{5.1}{4\pi^2} x_1^2 + \frac{5}{\pi} x_1 - 6)^2 + 10(1 - \frac{1}{8\pi})\cos(x_1) + 10 \qquad -5 \le x_1 \le 10, \quad 0 \le x_2 \le 15 \tag{22}$$

The total sample size is 20 with 5 (2×2+1) initial grid samples. Four tests are given as follows:

Case.1 The SS D-Kriging is first applied to generate the final contours of the surrogate model and the 20-sample profile, defined as *Profile-1*.
Case.2 Then the U-Kriging is also applied to the sample *Profile-1*.
Case.3 The U-Kriging is applied again to the sample profiles from Latin hypercube sampling (100 trials) and the best result is chosen.
Case.4 Finally, the sequential-sampling-based universal Kriging is applied with five initial grid samples.

Figures 5(a)-(d) show the comparison results.



(a) SS D-Kriging MSE = 0.055 (b) U-Kriging using *Profile-1* MSE = 94.67

(c) U-Kriging w/LHS MSE = 6.82 (best of 100 trials) (d) SS U-Kriging MSE = 0.41

Figure 5: Comparison between SS D-Kriging, U-Kriging and SS U-Kriging

From Figs. 5(a) and (b) we can see that if the metamodeling process is driven by SS D-Kriging from the very beginning, it gives us a very promising result in the end, showing that the MSE is as low as 0.055. If U-Kriging is applied to the samples from SS D-Kriging (*Profile-1* in this case), even if the samples are the same, the final result could be totally different (MSE=94.67), or significantly wrong. A more common case is using U-Kriging with Latin hypercube sampling, shown in Fig. 5(c); the best result out of 100 trials is better than case 2. Finally, from Fig. 5(d), if the metamodeling process is initiated with SS U-Kriging from the beginning, it could still give a good result in the end (the MSE is 0.41). Overall, both the sequential sampling method and the D-Kriging method contribute to the improvement of the accuracy of the surrogate model and the best result is achieved by combining them. In addition, we also need to be aware that it might be dangerous to apply U-Kriging to the sample profile from SS D-Kriging.

6. Conclusion

Metamodeling is widely applied for engineering application to represent the computation-intensive model. The traditional universal Kriging method has some limitations because of the fixed order of regression basis functions for the mean structure. We proposed a new method to automatically decide the regression basis functions by applying an genetic algorithm based feature selection algorithm to the candidate basis functions. An adaptive higher-order regression basis subset is obtained and leads to a more accurate surrogate model in the end. The sampling method is also a crucial issue in metamodeling. To be integrated with the D-Kriging method, a

prediction-bandwidth-based sequential sampling method is proposed. It identifies the position that has the least confidence of the prediction accuracy and inserts a new sample there accordingly. Mathematical examples show a faster convergence speed of the sequential sampling method compared with Latin hypercube sampling in a statistical sense. Moreover, the coupling effect reveals that this sequential sampling method is strongly connected with the Kriging method. Different Kriging methods will lead to different sample profiles. The sequential-sampling-based D-Kriging method yields the best result of the surrogate model in terms of accuracy and efficiency.

7. Acknowledgement

Research is supported by the Automotive Research Center which is sponsored by the U.S. Army Tank Automotive Research, Development and Engineering Center (TARDEC).

8. References

[1] T.W. Simpson, V. Toropov, V. Balabanov and F.A. Viana, Design and analysis of computer experiments in multidisciplinary design optimization: a review of how we have come—or not," *12th AIAA/ISSMO multidisciplinary analysis and optimization conference*, Victoria, British Colombia, 2008.

[2] J. Sacks, W.J. Welch, T.J. Mitchell and H. Wynn, Design and analysis of computer experiments, *Statistical Science*, 4(4), 409-23, 1989.

[3] C. Kim, S. Wang and K.K. Choi, Efficient response surface modeling by using moving least-squares method and sensitivity, *AIAA Journal*, 43(11), 2404-11, 2005.

[4] P. Lancaster and K. Salkauskas, Surfaces generated by moving least squares methods, *Mathematics of Computation*, 37(155), 141-58, 1981.

[5] A.J. Keane, Design search and optimization using radial basis functions with regression capabilities, *Proceedings of the conference on adaptive computing in design and manufacture*, Berlin, 2004

[6] S.M. Clarke, J.H. Griebsch and T.W. Simpson , Analysis of support vector regression for approximation of complex engineering analyses, *Journal of Mechanical Design*, 127(6), 1077-1087, 2005

[7] T. Goel, R. Haftka and W. Shyy, Comparing error estimation measures for polynomial and Kriging approximation of noise-free functions, *Structural and Multidisciplinary Optimization*, DOI 10.1007/s00158-008-0290-z, 2008

[8] D.G. Krige, A statistical approach to some basic mine valuation problems on the Witwatersrand, *Journal of the Chemical, Metallurgical and Mining Engineering Society of South Africa*, 52(6), 119–39, 1951

[9] P.C. Kyriakidis and M.F. Goodchild, On the prediction error variance of three common spatial interpolation schemes, *International Journal of Geographical Information Science*, 20(8), 823-855, 2006

[10] V.R. Joseph, Y. Hung, and A. Sudjianto, Blind Kriging: a new method for developing metamodels, *ASME Journal of Mechanical Design*, 130(3), 2008.

[11] J.R. Walker, Practical application of variance reduction techniques in probabilistic assessments, *the Second International Conference on Radioactive Waste Management*, Winnipeg, Canada, 517-521, 1986.

[12] L. Gu, R.J. Yang and C.H. Tho, Optimization and robustness for crashworthiness of side impact, *International Journal of Vehicle Design*, 26(4), 348-360, 2001

[13] A. Dey and S. Mahadevan, Ductile structural system reliability analysis using importance sampling, *Structure & Safety*, 20(2), 137-154, 1998.

[14] G.G. Wang, Adaptive Response surface method using inherited Latin hypercube design points, *Journal of Mechanical Design*, 125(2), 210-221, 2003

[15] R. Jin and W. Chen, On sequential sampling for global metamodeling in engineering design, Proceedings of *DETC'02 ASME 2002 Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Montreal, Canada, 2002

[16] J.P. Chiles and P. Delfiner, *Geostatistics: Modeling Spatial Uncertainty*, Wiley, New York, 1999.

[17] H.B. Nielsen, DACE: A matlab Kriging toolbox, http://*www2.imm.dtu.dk/~hbn/dace/*, 2007.

[18] D. Fisher, H.J. Lenz, *Learning from Data: Artificial Intelligence and Statistics V (Lecture Notes in Statistics)*, Springer, New York, 1996

[19] D. Broadhurst, R. Goodacre, A. Jones, J. Rowland and D. Kell, Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression, with applications to pyrolysis mass spectrometry, *Analytica Chimica Acta*, 348(1-3), 71-86, 1997

[20] I. Lee, K.K. Choi, L. Du and D. Gorsich, Inverse analysis method using MPP-based dimension

reduction for reliability-based design optimization of nonlinear and multi-dimensional systems, *Computer Methods in Applied Mechanics and Engineering*, 198(1), 14-27, 2008

[21] A. Forrester and A. Keane, Recent advances in surrogate-based optimization, *Aerospace Sciences*, 45(1-3), 50-79, 2009