Metamodeling Method Using Dynamic Kriging for Design Optimization

Liang Zhao, K. K. Choi, and Ikjin Lee
University of Iowa, Iowa City, Iowa 52242

DOI: 10.2514/1.J051017

Metamodeling has been widely used for design optimization by building surrogate models for computationally intensive engineering application problems. Among all the metamodeling methods, the kriging method has gained significant interest for its accuracy. However, in traditional kriging methods, the mean structure is constructed using a fixed set of polynomial basis functions, and the optimization methods used to obtain the optimal correlation parameter may not yield an accurate optimum. In this paper, a new method called the dynamic kriging method is proposed to fit the true model more accurately. In this dynamic kriging method, an optimal mean structure is obtained using the basis functions that are selected by a genetic algorithm from the candidate basis functions based on a new accuracy criterion, and a generalized pattern search algorithm is used to find an accurate optimum for the correlation parameter. The dynamic kriging method generates a more accurate surrogate model than other metamodeling methods. In addition, the dynamic kriging method is applied to the simulation-based design optimization with multiple efficiency strategies. An engineering example shows that the optimal design obtained by using the surrogate models from the dynamic kriging method can achieve the same accuracy as the one obtained by using the sensitivity-based optimization method.

Nomenclature

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>stochastic process vector in kriging model</td>
</tr>
<tr>
<td>F</td>
<td>matrix of basis functions evaluated at sample points</td>
</tr>
<tr>
<td>r</td>
<td>full set of basis functions</td>
</tr>
<tr>
<td>fEHA</td>
<td>global optimal subset of basis functions by exhaustive algorithm</td>
</tr>
<tr>
<td>fGA</td>
<td>optimal subset of basis functions by genetic algorithm</td>
</tr>
<tr>
<td>m</td>
<td>number of design variables</td>
</tr>
<tr>
<td>NTS</td>
<td>number of testing points used in local window</td>
</tr>
<tr>
<td>n</td>
<td>number of samples used for surrogate modeling</td>
</tr>
<tr>
<td>P</td>
<td>the highest order of the polynomial in basis functions</td>
</tr>
<tr>
<td>R</td>
<td>correlation matrix in kriging model</td>
</tr>
<tr>
<td>R(·)</td>
<td>correlation function in kriging model</td>
</tr>
<tr>
<td>r</td>
<td>correlation vector between the point of interest and the samples</td>
</tr>
<tr>
<td>Var(y)</td>
<td>variance of the true function values y at sample points</td>
</tr>
<tr>
<td>y(x)</td>
<td>vector of true function value at sample points</td>
</tr>
<tr>
<td>h(x)</td>
<td>prediction of response at point x</td>
</tr>
<tr>
<td>β</td>
<td>regression coefficient vector in kriging model</td>
</tr>
<tr>
<td>θ</td>
<td>correlation parameter vector in kriging model</td>
</tr>
<tr>
<td>σ²</td>
<td>process variance in kriging model</td>
</tr>
<tr>
<td>ψ(θ)</td>
<td>the objective function for correlation parameter estimation</td>
</tr>
</tbody>
</table>

I. Introduction

METAMODELING has been widely used in engineering applications when a simulation is difficult to obtain due to high computational cost. A surrogate model is used to represent the true model, with a limited number of simulations required to be evaluated. Extensive research has been carried out to investigate methods for generating the surrogate models based on limited samples. A number of methods, such as the least-squares regression, moving least-squares regression, support vector regression, and radial basis functions, have been developed over the years [1–7]. Recently, the kriging method has gained significant attention due to its capability of dealing with highly nonlinear problems [8,9]. In the kriging method, the response is modeled in two parts: the mean structure and a zero-mean stationary Gaussian stochastic process. The ordinary kriging method (OKG) assumes that this mean structure part is zero or constant on the entire domain. The universal kriging method (UKG) constructs the mean structure using the first- or second-order polynomials [10].

However, during the practical use of these methods, two problems have been discovered. The first problem is that the performance of the optimization methods used to find the optimal correlation parameter are affected by the highly nonlinear region near the origin and the large flat region in the rest area with multilocal minima of the objective function [11,12]. The popular DACE toolbox for the kriging method [11] uses the modified Hooke and Jeeves (H-J) algorithm to find the optimal correlation parameter. Martin [13] uses the Levenberg–Marquardt (L-M) method by employing a scoring method to calculate the Hessian matrix for optimization. Forrester and Keane [14] use a genetic algorithm (GA), which is a gradient-free method, to find the optimum. All these methods have their own advantages and disadvantages. The modified H-J method is efficient but unable to provide the true optimum. The L-M method is a gradient-based method, although it is efficient, it can only find a local optimum, and thus the obtained optimum is affected by the initial search point. Moreover, due to the large flat region and multiple local minima of the objective function, the L-M method often stops prematurely before converging to a true optimum. The GA method is supposed to be able to find the global optimum, but it is less efficient, and the obtained optimum varies due to the randomness within a genetic algorithm. In this paper, a generalized pattern search algorithm is used to find the optimal correlation parameter for the kriging method accurately and efficiently based on the MLE.

The second problem is that neither the OKG nor the UKG can adaptively fit the mean structure of the kriging model for highly nonlinear functions and fails to characterize the local nonlinearity of the true function in different design areas. It is shown that different basis functions may yield different surrogate models at the same sample profile. This is especially the case when a local window is used to generate the surrogate model in the design optimization.
process. That is, different basis functions may need to be used at on different local windows. Therefore, a new method that optimally selects basis functions to represent the mean structure based on current samples within the local window is desirable [15]. One method of adjusting the mean structure was proposed by Joseph et al. [16] by using a Bayesian framework to identify the mean structure for the kriging method. They use a Bayesian forward variable selection, which can be trapped into a local optimum and prevents itself from finding a global optimal subset of the basis functions. In this paper, a new method is proposed to find the pseudoglobular optimal basis functions by applying a GA for the selection procedure based on a new accuracy criterion.

Another issue is how to apply the proposed dynamic kriging method for simulation-based design optimization. Extensive works have been conducted in applying a surrogate modeling method for simulation-based design optimization [1,3,5,7,14,17]. The surrogate model generated by the dynamic kriging method needs to be efficiently and accurately applied for design optimization. In particular, because the dynamic kriging method uses the generalized pattern search for correlation parameter search and the genetic algorithm for basis functions selection, the computational time may become a concern. To overcome this difficulty, three efficiency strategies, including the local window for the surrogate model, a sequential sampling technique for new sample generation, and an adaptive initial search point for both the generalized pattern search and the genetic algorithm, are applied in this work. The proposed dynamic kriging method with the efficiency strategies is applied to solve an M1A1 Abrams road-arm design optimization example. The optimal result obtained by the proposed method can achieve the same accuracy as the result obtained by using the sensitivity-based design optimization.

II. Dynamic Kriging Method Using Pattern Search and Basis Selection

In this section, the traditional kriging method is reviewed first. Two issues in the kriging models pointed out earlier are solved by the proposed methods, which are using the generalized pattern search for correlation parameter estimation and using the genetic algorithm for basis-function selection.

A. Kriging Method

In the kriging method, the outcomes are considered as a realization of a stochastic process. Consider n sample points, \( x = \{x_1, x_2, \ldots, x_n\} \) with \( x \in \mathbb{R}^n \), and \( n \) responses \( y = \{y(x_1), y(x_2), \ldots, y(x_n)\} \) with \( y(x) \in \mathbb{R}^1 \). In the kriging method, the response at the samples consists of a summation of two parts as

\[
y(x) = F\beta + e\tag{1}\]

The first part of the right-hand side of Eq. (1), \( F\beta \), is the mean structure of the response, where \( F = [f(x)] \) \([f(x) = f_j(x); i = 1, \ldots, n, \text{ and } k = 1, \ldots, K] \) is an \( n \times K \) model matrix, and \( f(x) \) represents user-selected basis functions, which are usually in simple polynomial form, such as \( 1, x, x^2, \ldots, \) In Eq. (1), \( \beta = [\beta_1, \beta_2, \ldots, \beta_K]^T \) is the vector of the regression coefficients, which is obtained from the generalized least-squares method. The second part of the right-hand side of Eq. (1), \( e = \{e(x_1), e(x_2), \ldots, e(x_n)\}^T \), is a realization of the stochastic process \( e(x) \) that is assumed to have the zero mean \( E[e(x)] = 0 \) and covariance structure \( E[e(x)e(x')] = \sigma^2(R(\theta, x, x')) \), where \( \sigma^2 \) is the process variance, \( \theta = [\theta_1, \theta_2, \ldots, \theta_n]^T \) is the unknown process correlation parameter vector of dimension \( m \) that has to be estimated, and \( R(\theta, x, x') \) is the correlation function of the stochastic process [11]. In most engineering problems the correlation function is set to be a Gaussian form expressed as

\[
R(\theta, x, x') = \prod_{i=1}^{m} \exp(-\theta_i(x_{i,1} - x_{j,1})^2)
\]

where \( x_{i,j} \) is the \( j \)th component of the \( i \)th vector \( x_i \). The optimal choice of \( \theta \) is defined as the maximum likelihood estimator (MLE) [18], which is the maximizer of the likelihood function \( L \), expressed as

\[
L = (2\pi)^{-\frac{n}{2}}|R|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\sigma^2(y - F\beta)^T R^{-1}(y - F\beta)\right)\tag{3}
\]

where \( R \) is the symmetric correlation matrix with \( (ij) \)-th component \( R_{ij} = R(\theta, x_i, x_j) \) \((i, j = 1, \ldots, n)\), and \( \sigma^2 = \frac{1}{n}(y - F\beta)^T R^{-1}(y - F\beta) \) and \( \beta = (F^T F)^{-1}F^T y \) are obtained from the generalized least-squares regression. By taking the logarithm of Eq. (3) with the imposed \( \sigma^2 \) value and multiplying by \(-1\), the maximization problem to obtain optimal \( \theta \) is equivalent to

\[
\minimize \psi(\theta) \ \text{where} \ \psi(\theta) = \frac{1}{2} \ell_n(|R|) + \frac{n}{2} \ell_n(\sigma^2) \tag{4}
\]

After finding the optimal \( \theta \), the prediction of kriging model which interpolates the \( n \) sample points and the derivative of the prediction with respect to \( x \) are expressed as

\[
y(x) = f^T \beta + r^T R^{-1}(y - F\beta), \quad \hat{y}(x) = J_f(x)^T \beta + J_r(x)^T r
\]

where \( f = [f_i(x)]^T \) \((k = 1, \ldots, K)\) is the basis-function values evaluated at the predicted point \( x \). \( r = [R(\theta, x, x)^T, \ldots, R(\theta, x_n, x)]^T \). \( J_f(x) \) and \( J_r(x) \) are the Jacobians of \( f \) and \( r \), respectively.

B. Correlation Parameter Estimation Using Pattern Search Method

To show how the optimal \( \theta \) affects the final accuracy of the kriging prediction in Eq. (5), consider one simple revised example based on Forrester and Keane’s work [14]. In this example, the true function is expressed as

\[
y = (6x - 2)^2 \sin(12x - 4) + 10, \quad x \in [0, 1] \tag{6}
\]

and five evenly distributed samples along the \( x \) axis are used to generate the kriging prediction. Then the accuracy of the kriging prediction is tested using different \( \theta \), and the relative root-mean-squared error (rmse) is used as the accuracy measurement. In particular, the rmse is defined as

\[
\text{rmse} = \frac{1}{\text{NTS}} \sum_{i=1}^{\text{NTS}} \left(\frac{\hat{y}(x_i) - y(x_i)}{y(x_i)}\right)^2 \tag{7}
\]

where NTS is the number of testing points, and \( \hat{y}(x_i) \) and \( y(x_i) \) are, respectively, the kriging prediction and the true response at testing point \( x_i \). In this example, NTS is 100 and all the testing points are evenly distributed along the \( x \) axis. As shown in Fig. 1a, the rmse value changes significantly as \( \theta \) changes. As \( \theta \) increases from 2.7345 to 10, the rmse decreases from its maximum (0.21544) to the minimum (0.12981), whereas in Fig. 1b, the \( \psi(\theta) \) function value only changes from 0.803734 to 0.20981. This behavior shows that it is very important to accurately solve the minimization problem of Eq. (4) and find the true optimal \( \theta \) to generate an accurate kriging prediction.

Therefore, to accurately solve Eq. (4), it is proposed to use the generalized pattern search (GPS) method. The reason for using the GPS method is that the \( \psi(\theta) \) function in Eq. (4) usually has a highly nonlinear region near the origin and a large flat region elsewhere, as shown in Fig. 1b. In addition, the \( \psi(\theta) \) usually contains multiple local minima for high-dimension problems. For such a minimization problem, a gradient-based optimization algorithm often prematurely converges to a local minimum if the initial \( \theta \) value is close to the origin, or it prematurely stops in the large flat region if the initial \( \theta \) value is close to the upper bound of the \( \theta \) domain. Among all the non-gradient-based optimization algorithms, the GA is considered time-consuming and unreliable for such a continuous optimization problem, whereas the GPS method is not affected by the initial search.
point and accurately converges to the optimal $\theta$. The global convergence of the GPS method has been proven by Lewis and Torczon [19]. It is worth mentioning that since the $\psi(\theta)$ is always highly nonlinear near the origin and large flat elsewhere, the initial search point is set to be the lower bound of the $\theta$ domain for the optimization method to converge quickly.

To better demonstrate the challenge in this minimization problem and how the GPS method performs, the Branin–Hoo problem

$$f(x_1, x_2) = \left(x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(x_1) + 10$$

$x_1 \in [-5, 10], \ x_2 \in [0, 15]$ (8)

is used. The 20-Latin-hypercube sample (LHS) profile and the true function contour are first shown in Fig. 2a. The associated $\psi(\theta)$ plot is shown in Fig. 2b, where one can see that $\psi(\theta)$ indeed has a sharp corner region near the origin and a large flat region in the rest of the $\theta$ domain. The initial search point is set to be the origin of the $\theta$ domain. To have a fair comparison, it is worth mentioning the computer codes used in this example first. The modified H-J algorithm is applied by using the DACE MATLAB package. The L-M method is applied by following Martin’s [13] work. The GA method and the GPS method are applied by using the MATLAB Genetic Algorithm and Direct Search Toolbox R14, respectively. The stopping criteria set for all methods are the same:

1) The change in $\theta$ value is less than $1E-6$.
2) The change in objective function $\psi(\theta)$ value is less than $1E-6$.

To compare the accuracy of the kriging predictions based on the optimal $\theta$ obtained from different optimization methods, 100 grid testing points are used to calculate the rmse values for each method. Table 1 shows that the GPS method finds the best optimal $\theta$ value with the smallest $\psi(\theta)$ value, and the kriging prediction based on the optimal $\theta$ obtained by the GPS method achieves the best accuracy as well.

To demonstrate performance of the GPS for finding a better optimum of the $\psi(\theta)$ function compared with the other three optimization methods in a general way, a statistical study is conducted using the Branin–Hoo example again. In this statistical study, 100 randomly generated sets of 20-LHS samples are used. For each sample set, the ordinary kriging is applied to generate the prediction. The four optimization methods discussed above are applied to solve Eq. (4), and the $\psi(\theta)$ function values at the global optimum of $\theta$ are ranked from the smallest to the largest order. After the 100 trials, the frequency of the rank for four methods is shown in Table 2, where it shows that the GPS found the best optimal $\theta$ 92 times out of 100. Even though it is hard to claim that one optimization algorithm performance is better than the others all the time, to solve this
particular bounded constrained problem of Eq. (4), it is clearly shown that the GPS method can obtain the best results in finding the accurate optimal $\theta$, while the L-M method and the GA method have comparable performance thereafter.

To show the accuracy of the kriging models based on different optimal $\theta$ values from four optimization methods, $100 \times 100$ grid testing points are evaluated to calculate the rmrse values, and the rank of rmrse values associated with each optimization method from the smallest to the largest is shown in Table 3. It shows that the kriging model using the GPS method achieves a better accuracy than the other three methods. Note that the difference among the four methods in Table 3 is not as significant as the one in Table 2.

As the dimension of the design variables increases, the difference of using four optimization methods to solve Eq. (4) is becoming more significant. Consider a 12-D mathematical example, expressed as

$$y = (x_i - 1)^2 + \sum_{i=2}^{12} (2x_i^2 - x_{i-1})^2, \quad -10 \leq x_i \leq 10 \quad (9)$$

This function is called the Dixon–Price function. With 60 samples generated using the Latin hypercube sampling method, the ordinary kriging prediction using four different optimization methods for finding optimal $\theta$ is generated. Table 4 shows the optimal $\theta$ obtained using four different optimization methods and the associated objective function values and the rmrse values. The GPS method finds the best optimal $\theta$ with the smallest $\psi(\theta_{opt})$ value and generates the most accurate kriging prediction.

Like the previous example, to exclude the effect from the sample position and show the general performance of four optimization methods, 100 randomly generated sets of 60-LHS samples are used. The four optimizations are applied to find the optimal $\theta$ values to generate kriging prediction. The rmrse values are calculated based on a fixed set of 1000-LHS samples. Table 5 shows that the GPS method finds the best optimal $\theta$ value in 83 times, and the L-M method finds the best optimal $\theta$ in 17 times. The GA method and H-J method fails to find the best optimal $\theta$ value. At the same time, the associated rmrse values are ranked as well, as shown in Table 6. The kriging model with the GPS method generates the most accurate surrogate model in 80 times, followed by the one with the L-M method of 18 times. Tables 5 and 6 indeed show that the GPS method outperforms the other three optimization methods for this high-dimension problem.

### C. Dynamic Basis-Function Selection Using Genetic Algorithm

For the UKG method, the basis functions $f$ of $F$ used in Eq. (1) are fixed during the entire metamodeling process, and it usually takes up to the second-order polynomial functions. However, it is clear that higher-order terms can predict nonlinear mean structure, which may vary for different problems. Hence, in general, for highly nonlinear problems, fixed lower-order basis functions may not be suitable to describe the nonlinearity of the mean structure. On the other hand, Martin and Simpson [20] pointed out that, in some cases, the accuracy of the surrogate model may not be enhanced by using higher-order terms. That is, the surrogate model may become even worse when some particular higher-order terms are used.

The impact of selection of basis functions can be shown using the following illustrative example:

$$f(x_1, x_2) = \frac{0.0234 x_1^2 + 0.1072 x_2 + 0.0825 x_1 x_2 + 0.0755 x_1^2 x_2}{100}, \quad x_1, x_2 \in [0, 10] \quad (10)$$

where the true function plot and the samples are shown in Fig. 3.

The true function is highly nonlinear in the $x_1$ direction and linear in the $x_2$ direction. The kriging method with different basis functions is applied to this problem using the 14 samples obtained from LHS, as shown in Fig. 3, and the rmrse values that are calculated from $100 \times 100$ grid testing points are compared. From Table 7 one can

<table>
<thead>
<tr>
<th>Table 1 Comparison between four optimization methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-J</td>
</tr>
<tr>
<td>(1.7679, 0.2628)</td>
</tr>
<tr>
<td>$\psi(\theta_{opt})$</td>
</tr>
<tr>
<td>0.1169</td>
</tr>
<tr>
<td>RMSE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2 Frequency of rank of $\psi(\theta)$ function value at optimal $\theta$ by different optimization methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
</tr>
<tr>
<td>First</td>
</tr>
<tr>
<td>Second</td>
</tr>
<tr>
<td>Third</td>
</tr>
<tr>
<td>Fourth</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3 Frequency of rank of rmrse values by different optimization methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
</tr>
<tr>
<td>First</td>
</tr>
<tr>
<td>Second</td>
</tr>
<tr>
<td>Third</td>
</tr>
<tr>
<td>Fourth</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4 Comparison between four optimization methods (12-D problem)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-J</td>
</tr>
<tr>
<td>Optimal $\theta_{opt}$</td>
</tr>
<tr>
<td>0.0100</td>
</tr>
<tr>
<td>0.0100</td>
</tr>
<tr>
<td>0.0133</td>
</tr>
<tr>
<td>0.0100</td>
</tr>
<tr>
<td>0.0157</td>
</tr>
<tr>
<td>0.0100</td>
</tr>
<tr>
<td>0.0115</td>
</tr>
<tr>
<td>0.0100</td>
</tr>
<tr>
<td>0.0170</td>
</tr>
<tr>
<td>0.0124</td>
</tr>
<tr>
<td>0.0234</td>
</tr>
<tr>
<td>0.0189</td>
</tr>
<tr>
<td>RMSE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5 Frequency of rank of rmrse values by different optimization methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
</tr>
<tr>
<td>First</td>
</tr>
<tr>
<td>Second</td>
</tr>
<tr>
<td>Third</td>
</tr>
<tr>
<td>Fourth</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6 Frequency of rank of rmrse values by different optimization methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
</tr>
<tr>
<td>First</td>
</tr>
<tr>
<td>Second</td>
</tr>
<tr>
<td>Third</td>
</tr>
<tr>
<td>Fourth</td>
</tr>
</tbody>
</table>
see that the rmse value decreases from zeroth-order to first-order polynomials. However, it increases from first-order to third-order polynomials, which indicates that increasing order does not necessarily improve the accuracy of the surrogate model. Moreover, if several unnecessary basis functions are excluded to obtain the customized-order kriging, the kriging prediction becomes more accurate.

Therefore, the problem is how to find the optimal subset of the basis functions such that the obtained kriging prediction would have the best accuracy: that is, to find a subset of the basis functions such that the obtained kriging prediction can have the smallest rmse value. However, since the rmse value is not available unless the true function is explicitly known, it is proposed to use the kriging process variance \( \sigma^2 \) as the estimator of the accuracy for the kriging prediction in this paper. Therefore, the formulation for this problem becomes to find a subset of the basis functions to minimize

\[
\sigma^2 = \frac{1}{n} (y - F\beta)^T R^{-1} (y - F\beta) \tag{11}
\]

Note that different types of the candidate of basis functions, such as Hermit polynomials, trigonometric functions, and exponential functions, have been tested in this work, and it is found that the simple polynomial forms perform efficiently and effectively without losing accuracy. Thus, in this paper, all the candidate basis functions are assumed to be polynomials and in the form of their multiplications \( x_1^{p_1} x_2^{p_2} \cdots x_m^{p_m} \) where \( m \) is the number of design variables, \( p_i \in [0, P] \) is an integer power of \( x_i \),

\[
\sum_{i=1}^{m} p_i \leq P
\]

and \( P \) is the highest order of the mean structure in the kriging model. The total number of possible candidate basis functions is \( C_{m+P}^P \). Therefore, the full set \( \mathbf{f} \) becomes

\[
\mathbf{f} = \{1, x_1, x_2, \ldots, x_m, x_1^2, x_2^2, \ldots, x_m^2, x_1 x_2, \ldots, x_1 x_m, \ldots, x_1^{p_1} x_2^{p_2} \cdots x_m^{p_m} \} \tag{12}
\]

In Eq. (11), one constraint needs to be satisfied first. That is, the total number of possible candidate basis functions cannot be larger than the number of samples to generate the kriging prediction. Therefore, by finding the largest \( P \) such that \( C_{m+P}^P \leq n - 1 \), the highest-order \( P \) in Eq. (11) is determined. The reason for using \( n - 1 \) instead of \( n \) is that it is known that when the number of basis functions equals the number of samples, it sometimes causes an overfitting problem in the kriging prediction. Therefore, to make the kriging method work robustly, it is recommended to use \( n - 1 \) instead of \( n \) to find the highest-order \( P \). After \( P \) is determined according to the number of samples, Eq. (10) becomes a classic variable selection problem, expressed as follows:

Find the subset of \( \mathbf{f} \) to minimize

\[
\sigma^2 = \frac{1}{n} (y - F\beta)^T R^{-1} (y - F\beta) \tag{13}
\]

It is obvious that the global optimal subset of these candidate basis functions can be guaranteed only by applying the exhaustive algorithm (EHA), which evaluates all possible \( 2^m \) subsets of the basis functions, where \( M \) is the number of candidate basis functions. Consequently, the computational expense of EHA increases rapidly and becomes unaffordable when \( M \) is large. Therefore, an alternative method to solve Eq. (11) needs to be applied such that the kriging prediction based on this alternative optimal subset is accurate enough and close to the result obtained using the true optimal subset with less computational expense.

As discussed in section I, many research works have been carried out for the variable selection problems. In an area related to kriging modeling, the blind kriging method uses a Bayesian forward variable selection to find the significant coefficients in \( \beta \) by using the cross-validation error as the objective function. In this blind kriging framework, the forward-selection scheme can make the optimization process prematurely converge to a local optimal selection. In this work, the GA is applied to find the optimal selection for basis functions. A main concern of using the GA method is the number of iterations and convergence time [21]. This is true for using the GA to solve a continuous problem. However, in this particular basis-function selection problem there are several reasons that the GA method can be efficient and attractive. The main reason is that the GA intends to find the global optimum instead of the local optimum, which leads to a more accurate kriging prediction, compared with the result using a forward-selection scheme. The second reason is that it is a discrete minimization problem in Eq. (12) with limited \( C_{m+P}^P \) possible basis functions. Unlike the encoding or decoding computation for the solution in a continuous problem, the selection of the basis function itself can be directly expressed in genetic form, where 1 means selected and 0 means nonselected. The third reason is that with selection of complementary (i.e., opposite) subsets for the initial generation, the GA can converge quickly. The fourth reason is that to restrict the total computational time, one can set the maximum number of iterations and modifies the highest order of \( P \) for the GA method for complex engineering application and yet obtains a satisfactory result. Efficiency strategies for how to apply the GA method for solving Eq. (10) are discussed in detail in the following sections.

1. Initial Generation

The GA procedure starts with an initial generation, called the zeroth generation. In this paper, the zeroth generation includes both the single basis-function subsets and almost-full basis-function subsets. The \( C_{m+p}^{p} \) single basis-function subsets are defined as

\[
[1, 0, 0, \ldots, 0], [0, 1, 0, \ldots, 0], \ldots, [0, 0, \ldots, 0, 1]
\]

which indicate that each single basis-function subset is included in the initial generation. The other \( C_{m+P}^{P} \) almost-full basis-function subsets

\[
[0, 1, 1, \ldots, 1], [1, 0, 1, \ldots, 1], \ldots, [1, 1, \ldots, 1, 0]
\]
which are complementary to the single basis-function subsets, are included in the initial generation. Other than these subsets, the full basis-function subsets for each order (from first order to Pth order) are also included, which adds another P subsets in the initial generation. Altogether, there are $2^{\text{R}_{m,s}^f + P}$ subsets in the zeroth generation. The single basis-function subsets and almost-full basis-function subsets are used to avoid reaching the local optimum for the basis selection. Based on numerous examples tested during this study, it is found that if the GA starts only at one side, either the single basis-function subsets or the almost-full basis-function subsets, there is a good chance that the GA procedure will prematurely converge to a local optimum.

2. Convergence Criteria

As discussed earlier, the convergence criterion of the GA needs to be carefully set to have the GA converge efficiently. In this work, the convergence conditions are chosen as follows:

1) Condition 1 is that the number of stalled iterations is equal to 2.

2) Condition 2 is

$$\frac{\sigma^2_{i-1} - \sigma^2_i}{\sigma^2_{i-1}} \leq 1\% \quad \text{and} \quad \frac{\sigma^2_{i-2} - \sigma^2_{i-1}}{\sigma^2_{i-2}} \leq 1\%$$

(14)

3) Condition 3 is that the maximum number of iterations is $n^f_{\text{EHA}}$.

Condition 1 means the GA stops if the process variances $\sigma^2$ in any two consecutive iterations are the same. Condition 2 means the GA stops if the absolute relative change of $\sigma^2$ between two consecutive iterations is less than 1%. Condition 3 means the GA stops if the maximum number of iteration reaches the number of the candidate basis functions. The total stopping criterion is that the GA stops if any of these three conditions is satisfied. After testing a number of different problems, it is found that condition 1 is the most frequent, which indeed indicates that the GA stops very quickly.

3. Additional Efficiency Strategy in Basis-Function Selection

When evaluating the process variance $\sigma^2$ using different subsets, the optimization problem for $\theta$ search in Eq. (4) needs to be solved every time, which results in a significant computational time if the GPS method is used. With a number of testing problems, it is found that the optimal $\theta$ does not significantly affect the result of the basis-function selection. Therefore, in this paper, the $\theta$ search in evaluating each subset during the GA process is removed; instead, the ordinary kriging model is generated every time, which results in a significant computational time spent by the EHA to find a solution for this problem. Specifically, the clock time spent on the GA-based selection is 604 ms, whereas the clock time spent on the EHA-based selection is 152,370 ms, and the clock time spent on the blind kriging is 374 ms on the Intel P8700 CPU computer.

To verify whether the GA-based selection algorithm is robust, a robustness study is carried out as follows. First, a performance is defined as

$$\text{performance} = \frac{\sum_{i=1}^{n} I (\text{rrmse}_{\text{GA}} < \text{rrmse}_i) \times 100\%}{n}$$

(15)

where $I(\cdot)$ is the indicator function, and rmse, is calculated using the kriging prediction with the $i$th subset out of the total $2^{\text{R}_{m,s}^f + P}$ subsets during the EHA selection process. The performance in Eq. (14) indicates the percentile of the accuracy of the subset of the basis functions obtained by the GA method among all possible $2^{\text{R}_{m,s}^f + P}$ subsets. Since the sample position has an influence on the result, 10 consecutive trials with different sample sets from LHS are carried out to see if the GA-based selection method is robust. In Fig. 5, the solid line is the rmse values obtained by using the GA method for basis selection, the dashed line is the rmse values obtained by the optimal basis selection using the EHA method, and the dotted line is the performance as defined in Eq. (14) for each trial. It shows that the GA-based selection process can find a very good subset of basis functions, which is better than about 97% of other subsets in the EHA process, while only using 0.5% of the computational time used by the EHA method.

The stepwise selection method was also tested for selecting the basis functions. It is found that the performance by the stepwise selection method was not as accurate as the proposed GA method. In Appendix A, a comparison study is carried out to demonstrate the difference between the two methods.

With the optimal $f_{\text{GA}}$ obtained using the GA method by solving Eq. (12) and the optimal $\theta$ obtained using the GPS method by solving Eq. (4), the dynamic kriging (DKG) method is formed and compared with other existing surrogate modeling methods in the following sections.

III. Comparison Study Between Dynamic Kriging and Other Metamodeling Methods

A. Comparison Procedure

To compare the performance of the DKG method against other metamodeling methods, we selected the four most widely used metamodeling methods [22], which are the UKG method, the polynomial response surface (PRS) method, the radial basis-function (RBF) method, and the blind kriging (BKG) method. To make a fair comparison, we first need to specify how these methods are optimally used in this paper.

For the UKG method, the mean structure is set to be second-order polynomials. For the PRS method, the response $y$ is considered as the linear combination of regression basis functions. The predicted response is expressed as

$$y = a_0 + \sum_{i=1}^{m} a_i x_i + \sum_{i=1}^{m} \sum_{i=1}^{m} a_{i,j} x_i x_j + \cdots$$

$$+ \sum_{i=1}^{m} \cdots \sum_{i=1}^{m} a_{i_1 \cdots i_k} x_{i_1} x_{i_2} \cdots x_{i_k}$$

(16)

To accurately apply the PRS method, the highest-order $P$ of the polynomials is decided by finding the best $P$ such that the prediction would have the smallest cross-validation error. For the RBF method,
the response $y$ is considered as a linear combination of basis functions, expressed as

$$y = w^T \psi = \sum_{i=1}^{n} w_i \psi(||x - c_i||)$$

where $c_i$ is the center of the $i$th basis function. In this paper, $\psi(r) = e^{-r/\sigma^2}$ is used. The $\sigma$ value is also determined by minimizing the cross-validation error.

The SURROGATES toolbox is used to test the above three methods with the modification of using cross-validation to find the best values of the model parameters for each method. For the blind kriging method, the original code from Joseph et al. [16] is used.

The comparison procedure is carried out as follows. First, $n$ samples are generated by the LHS method. Secondly, five surrogate models are generated using the given samples. After constructing the surrogate models using the five methods, the function values from the surrogate model at $S$ evenly distributed testing points are predicted and the rmse values are calculated as the accuracy measurement. Then a rank is determined for these five methods in terms of the accuracy of the generated surrogate model based on the rmse values from each method. To eliminate the effect of the specific sample profile, the comparison is conducted for 50 trials with 50 different sample sets, and the frequency of being identified as the best surrogate model is counted to find the method that performs the best.

For comparison of these methods, one important point is the level of accuracy at which these surrogate models should be compared. That is, comparing performance of metamodeling methods when none of the surrogate models achieved an appropriate level of accuracy for the purpose of applications is meaningless. Therefore, one first needs to set the level of accuracy at which the surrogate model will be used for the comparison study. In this paper, the coefficient of determination $R^2$ is used as the normalized accuracy measurement to check if the surrogate model is acceptable or not.

![Graphs](Image of graphs showing optimization histories and contour predictions)

**Fig. 4** Contours from different basis selection results.

**Fig. 5** Comparison between GA and EHA selections.
The surrogate model is defined to be accurate when the median of the $R^2$ value is larger than 0.99 for 50 trials. The rank of the performance of each metamodeling method is compared at the sample size when at least one method can generate a surrogate model with $R^2$ larger than 0.99.

### B. Benchmark Problems for Comparison Study

The first problem chosen for comparison is the Branin–Hoo function given in Eq. (8). In this problem the true function is a combination of polynomial and cosine functions. Therefore, it is not in favor of any of the five methods and can be viewed as an unbiased problem for all methods. As shown in Tables 8 and 9, the comparison started at the 16-sample case, and the DKG method achieved the acceptable surrogate model first at the 18-sample case, where the DKG method has been identified as the best for 43 times, as shown in Table 9. Table 10 shows that the mean of rrmse of the five methods with the DKG method performed the best.

The second problem used for comparison study is an engineering application of a M1A1 tracked-vehicle road-arm problem. The road arm is modeled using 1572 eight-node isoparametric finite elements (SOLID45) and four beam elements (BEAM44) of ANSYS, as shown in Fig. 6, and is made of S4340 steel with Young’s modulus $E = 3.0 \times 10^7$ psi and Poisson’s ratio $v = 0.3$. The durability analysis of the road arm is carried out using the Durability and Reliability Analysis Workspace (DRAW) [24] to obtain the fatigue life. The fatigue lives at the 13 critical nodes shown in Fig. 7 are chosen as the design constraints. In Fig. 8, the shape design variables consist of four cross-sectional shapes of the road arm, where the widths ($x_1$ direction) of the cross-sectional shapes are defined as design variables $d_1$, $d_2$, $d_3$, and $d_4$ at intersections 1, 2, 3, and 4, respectively, and the heights ($x_3$ direction) of the cross-sectional shapes are defined as design variables $d_5$, $d_6$, $d_7$, and $d_8$. Since the finite element analysis and fatigue analysis are time-consuming, the surrogate model is needed when carrying out design optimization. For comparison of the DKG and other metamodeling methods, the normalized fatigue life at the first critical node is used as the response, and the surrogate model is to be generated for

$$G(d) = 1 - \frac{L(d)}{L_t}$$

(18)
where \( L(d) \) is the crack-initiation fatigue life at the first critical node, and \( L_t \) is the crack-initiation target fatigue life, which is 5 years for this example. The domain for generating the surrogate model is defined as a hypersphere with a radius of \( 5\% \times \|d_0\| \), where

\[
d_0 = [1.750 \ 3.250 \ 1.750 \ 3.170 \ 1.756 \ 3.038 \ 1.752 \ 2.908]
\]

The same 50-trial statistical study as conducted in the previous example is carried out. In each trial, 50 LHS samples within the hypersphere are randomly generated. The surrogate models are generated by each of the five metamodeling methods, 2000 LHS samples are first evaluated using the finite element analysis and used to calculate the rmse value for each surrogate model. After 50 trials, the rank of the rmse values of the surrogate models using different metamodeling methods is calculated and shown in Table 11. Again, the DKG method performs the best in 28 times for generating the most accurate surrogate model, followed by the blind kriging method. Table 12 shows the mean rmse values for each of the surrogate modeling methods, where the rmse from the DKG method is smallest and followed by the blind kriging method.

IV. Design Optimization Using Dynamic Kriging Method

In simulation-based design optimization, surrogate models are widely used. In this section, a detailed practical use of the DKG method for design optimization is discussed. After the explanation of several efficiency strategies of how to use the DKG method, an engineering design optimization problem is used to demonstrate the overall performance of the DKG method for the simulation-based design optimization.

A. Local Window for Surrogate Modeling

Since the DKG method selects the best basis functions subset according to the nonlinearity of the response, it is better to generate the surrogate model on a local window than to generate a global surrogate model on the entire design domain. When the candidate design point moves at each iteration, the sample set within the local window changes; therefore, the DKG method will choose different basis functions subsets according to the local nonlinearity of the response to generate the most accurate surrogate model locally. This local window concept is visualized in Fig. 9. The hypersphere used to define the local window is expressed as

\[
\sum_{i=1}^{m} (x_i - d_i)^2 \leq R^2
\]

where \( d = [d_1, d_2, \ldots, d_m] \) is the current design point and \( R \) is the radius. In this paper, the \( R \) value is set as \( 5\% \times \|d\| \).

To see how the DKG method works effectively in the local window, consider a 2-D highly nonlinear polynomial function expressed as

\[
\text{Table 11 Frequency of rank of five methods (road arm, 50 points, 50 trials)}
\]

<table>
<thead>
<tr>
<th>Rank</th>
<th>UKG</th>
<th>RBF</th>
<th>PRS</th>
<th>BKG</th>
<th>DKG</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>28</td>
</tr>
<tr>
<td>Second</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>21</td>
<td>22</td>
</tr>
<tr>
<td>Third</td>
<td>38</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Fourth</td>
<td>6</td>
<td>11</td>
<td>32</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Fifth</td>
<td>0</td>
<td>36</td>
<td>14</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[
\text{Table 12 Mean rmse values for each method (road arm, 50 points, 50 trials)}
\]

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean rmse</th>
</tr>
</thead>
<tbody>
<tr>
<td>UKG</td>
<td>1.4461</td>
</tr>
<tr>
<td>RBF</td>
<td>3.1603</td>
</tr>
<tr>
<td>PRS</td>
<td>2.2046</td>
</tr>
<tr>
<td>BKG</td>
<td>0.6518</td>
</tr>
<tr>
<td>DKG</td>
<td>0.5845</td>
</tr>
</tbody>
</table>
for this work, the surrogate model can be used for $	ext{fl}$ is chosen as the one that has the largest $j_{2043}$ is chosen as $\text{fl}$ are selected as the best basis functions at local $^\text{N}_{\text{max}}$ among all constraints, expressed as for $i$ initial samples are generated on the local window can be $\text{fl}$ method at each local window, $\text{fl}$ are selected as the best basis terms are not selected shows that it is not true function. The fact that different basis-function sets are selected at $\text{fl}$ sampling points and local windows at each point are shown in Fig. 10.

Nine initial samples are randomly generated within each local window, and basis functions up to the second-order polynomial function are used for the DKG method. After applying the DKG method at each local window, $\{x_1, x_2, x_3, x_4\}, \{x_1, x_2, x_3, x_4, x_5\}$, and $\{x_1, x_2, x_3, x_4, x_5\}$ are selected as the best basis functions at local windows at $\text{fl}$, $\text{fl}$, and $\text{fl}$, respectively, to accurately describe the true function. The fact that different basis-function sets are selected at different design points means that one basis-function set cannot best describe the local nonlinearity of the true function, since the local nonlinearity changes as the design point moves. Furthermore, the fact that all six possible basis terms are not selected shows that it is not necessarily good to use all available terms for the generation of surrogate models; this indeed shows the effectiveness of the DKG method.

### B. Sampling Strategy

After deciding the local window for the surrogate model generation, $N$, initial samples are generated on the local window using the Latin centroidal Voronoi tessellations (LCVT) \[22\] for evenly distributed samples, and then surrogate models are generated on the local window. The minimum number of the initial samples is decided by $\left(\binom{n}{k}\right)$. For example, for an 8-D example with up to second-order polynomial basis functions, the minimum number of $N$ will be $\binom{8}{2} = 45$. However, for high-dimensional problems, the minimum number of initial samples may not be sufficient to generate accurate surrogate models; more samples may be needed, depending on the accuracy of the surrogate model \[25\]-\[29\]. The accuracy of the surrogate model generated with the initial samples $N$, can be estimated using

$$\eta = \frac{\text{mean}(\text{mse}(x))}{\text{Var}(y(x))}$$

for $i = 1$–NTS, $j = 1$–$n$ \[21\]

where $\text{Var}(y(x))$ is the variance of $n$ true responses at the sample points and is used to normalize the accuracy measure. NTS is the total number of testing points generated using LCVT, and $\text{mse}$ is the predicted mean square error (mse) from the DKG model. The physical meaning of the accuracy measure in Eq. (20) is related to the prediction variance of the kriging model. Hence, the smaller the prediction variance, the more accurate the surrogate model. If the accuracy of a surrogate model is satisfactory, which is defined as $\eta \leq 1\%$ in this work, the surrogate model can be used for optimization. However, if the accuracy does not satisfy the target, more samples are sequentially inserted within the local window until the surrogate model satisfies the target accuracy condition. The new inserting point $x_{\text{new}}$ is chosen as the one that has the largest true value among the testing points in the local window. For a typical design optimization problem, multiple constraints are usually involved. In this case, the accuracy measure in Eq. (20) needs to be modified to reflect the effect of multiple surrogate models. In this paper, the maximum value of accuracy measures for each surrogate model is used as the overall accuracy measure for multiple surrogate models, and thus the accuracy measure is given by

$$\eta_{\text{max}} = \max \left\{ \frac{\text{mean}(\text{mse}_i(x))}{\text{Var}(y_i(x))} \right\}, \text{ for } i = 1$–NTS, \hspace{1cm} j = 1$–$n, \hspace{1cm} k = 1$–$n_c$$

where $y_i(x_j)$ and $\text{mse}_i$ are the variance of $n$ true responses and the $\text{mse}$ for the $k$th surrogate model, and $n_c$ is the number of surrogate models. Correspondingly, the new inserting point $x_{\text{new}}$ is chosen as the point that has the largest $\text{mse}$ among all constraints, expressed as $\arg \max \{\text{mse}_i(x_i)\}$, for $i = 1$–NTS, $j = 1$–$n$, and $k = 1$–$n_c$.

### C. Adaptive Initial Point for Pattern Search and Basis-Function Selection

When applying the DKG method to a complex engineering design optimization problem, the number of variables used for surrogate modeling is usually large. In such cases, the pattern search algorithm to find the optimal correlation parameter $\theta$ in Eq. (4) and the genetic algorithm to find the optimal basis-function subset in Eq. (12) may become computationally expensive. It is known that the computational efficiency of the pattern search algorithm and the genetic algorithm is strongly affected by the initial search point. If the initial search starts from the neighboring area of the true optimum, it can find the optimal $\theta$ within a remarkably shorter time than if it starts from a point far away from the optimum. Moreover, along the optimization history, if the design movement is small, which means that the current design is near the optimal design, the surrogate model generated at the current design will be very similar to the one generated at the previous design. This means that two optimal $\theta$ will be close. Therefore, we can adaptively use the optimal $\theta$ obtained in the previous iteration as the initial point for the pattern search of the current iteration instead of using arbitrary initial point. Similarly, the optimal basis-function subset $\text{opt}$ found in the previous iteration will be included in the initial generation in GA for the current iteration. This will save computational time for the DKG method, in particular, when the current design is in the neighborhood of the optimal design.

The overall flowchart of using the DKG method for the design optimization is shown in Fig. 11.
D. Numerical Example

This section uses the M1A1 tracked-vehicle road-arm problem to carry out the design optimization using the DKG method with the efficiency strategies discussed in previous sections. The finite element model, the critical points, and the design variables definition are the same, as shown in Figs. 6–8. Table 13 shows the initial design and the design domain for each variable.

The design optimization for the M1A1 tracked-vehicle road arm is formulated to

\[
\begin{align*}
\text{Minimize} \quad & \text{cost}(\mathbf{d}) \\
\text{subject to} \quad & G_j(\mathbf{d}) < 0, \quad j = 1, \ldots, 13, \\
& \mathbf{d}^L \leq \mathbf{d} \leq \mathbf{d}^U, \quad \mathbf{d} \in \mathbb{R}^8
\end{align*}
\]  

(23)

where \( \text{cost}(\mathbf{d}) \) is the weight of the road arm,

\[
G_j(\mathbf{d}) = 1 - \frac{L(\mathbf{d})}{L_t}, \quad j = 1-13
\]  

(24)

\( L(\mathbf{d}) \) is the crack-initiation fatigue life, and \( L_t \) is the crack-initiation target fatigue life (5 years).

The constraint value \( G_j(\mathbf{d}) \) and the sensitivity of \( G_j(\mathbf{d}) \) with respect to the design variables are predicted by \( \tilde{G}_j(\mathbf{d}) \) and

<table>
<thead>
<tr>
<th>Design variables</th>
<th>Lower bound ( \mathbf{d}^L )</th>
<th>Initial design ( \mathbf{d}^e )</th>
<th>Upper bound ( \mathbf{d}^U )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_1 )</td>
<td>1.350</td>
<td>1.750</td>
<td>2.150</td>
</tr>
<tr>
<td>( d_2 )</td>
<td>2.650</td>
<td>3.250</td>
<td>3.750</td>
</tr>
<tr>
<td>( d_3 )</td>
<td>1.350</td>
<td>1.750</td>
<td>2.150</td>
</tr>
<tr>
<td>( d_4 )</td>
<td>2.570</td>
<td>3.170</td>
<td>3.670</td>
</tr>
<tr>
<td>( d_5 )</td>
<td>1.356</td>
<td>1.756</td>
<td>2.156</td>
</tr>
<tr>
<td>( d_6 )</td>
<td>2.438</td>
<td>3.038</td>
<td>3.538</td>
</tr>
<tr>
<td>( d_7 )</td>
<td>1.352</td>
<td>1.752</td>
<td>2.152</td>
</tr>
<tr>
<td>( d_8 )</td>
<td>2.508</td>
<td>2.908</td>
<td>3.408</td>
</tr>
</tbody>
</table>

Number of function evaluations \( -- \) \( 11 + 11 \times 8 \) \( 146 \)

Active constraints \( -- \) \( 1,3,5,8,12 \) \( 1,3,5,8,12 \)

Cost \( 515.09 \) \( 466.80 \) \( 466.93 \)
$\frac{\partial G_i(d)}{\partial d_i}$ from Eq. (5) where $\hat{G}_i(d)$ is the prediction generated by the DKG method. The sequential quadratic programming algorithm is used to solve the optimization problem of Eq. (22). In each iteration, 15 samples are used as the initial samples in the local window and new samples are sequentially inserted if the accuracy of the surrogate model is not achieved. After 11 iterations, the optimization process converged to the optimal design, using 146 samples altogether. To verify the accuracy of the optimal result, the sensitivity-based design optimization is carried out again.

In the sensitivity-based design optimization, the sensitivity information $\frac{\partial G_i}{\partial d_j}, i = 1, 13$ and $j = 1, 8$ is available through the design sensitivity analysis method [30] and used to solve Eq. (22); therefore, no surrogate model is needed in this procedure. Table 14 shows the comparison results for the two approaches. The optimal design obtained using the surrogate-based design optimization is almost identical with the optimal design obtained using the sensitivity-based design optimization except the small difference in $d_{10}$ and $d_{14}$. The sensitivity-based design optimization requires 11 function and 11 sensitivity evaluations. One sensitivity evaluation includes sensitivity calculations for all design variables, so it requires eight sensitivity calculations in this example, whereas the sampling-based design optimization requires a total of 146 samples for the surrogate model generation using the DKG method.

In Fig. 12, the average computational time to generate each surrogate model is shown as the solid line with squares. The average computational time to generate each surrogate model without applying the efficiency strategy is also shown as the solid line with stars. It is shown that with applying the efficiency strategy, the average computational time is reduced from 4 to 1 s. In addition, the most reduction occurs at the second iteration due to the improvement of the initial theta point and basis-function selection.

V. Conclusions

When applying the metamodeling method to generate the surrogate model based on limited number of samples, the kriging method is often used. The traditional UKG method has some limitations because of the fixed order of regression basis functions for the mean structure and the optimization method that are used to obtain the optimal correlation parameter. The DKG method is proposed to find an accurate optimal correlation parameter by using the generalized pattern search method and determining the best subset of the basis functions dynamically by applying the genetic algorithm. Comprehensive comparison studies show that the DKG method can generate accurate surrogate models compared with the traditional metamodeling methods such as the UKG method, the radial basis-function method, the polynomial response surface method, and the blind kriging method. With the use of sequential sampling and efficiency strategies, the DKG method is applied to a simulation-based design optimization problem to efficiently obtain an accurate optimal design. The numerical example shows that the optimal design obtained using the DKG method with efficiency strategies achieves the same accuracy as the optimal design obtained using the sensitivity-based design optimization.

Appendix: Comparison Between Stepwise Selection and Genetic Algorithm Selection for Basis Functions

To compare the performance between the stepwise selection and the genetic algorithm selection methods for solving Eq. (12), the Branin–Hoo example with the 20-run LHS samples is used. Table A1 shows results of selected basis functions from two methods and the rrmse values of the surrogate models. The genetic algorithm found the optimal subset as $[1, x_1, x_2, x_1^2, x_2^2, x_1 x_2]$, and the associated rrmse value for the surrogate model is 0.07, whereas the stepwise selection found the optimal subset as $[1, x_1, x_2, x_1 x_2, x_1^2, x_2^2, x_1^2 x_2, x_1 x_2^2]$ and the associated rrmse value as 0.33.

A simulation study is also carried out to compare the performance between the genetic algorithm basis selection and the stepwise basis selection. In this study, 50 sets of 20-LHS samples are randomly generated. For each sample set, the genetic algorithm basis selection and the stepwise basis selection are applied to find the optimal basis functions for the kriging model. After finding the basis functions, the surrogate models are constructed based on the selected basis functions and 100 × 100 grid testing points are used to calculate the rrmse values of the surrogate models. Table A2 shows the comparison result. Again, the genetic algorithm outperforms the stepwise selection.

Acknowledgments

Research is primarily supported by the U.S. Army Research Office project W911NF-09-1-0250. This research was also partially supported by the World Class University Program through a National Research Foundation of Korea grant funded by the Ministry of Education, Science and Technology (grant number R32-2008-000-10161-0 in 2009). This support is greatly appreciated. We would like to thank Roshan Joseph and Ying Hung for providing the original code for the blind kriging method and their valuable suggestions. We would also like to thank Felipe Viana for providing the SURROGATES toolbox used in the comparison study and for his generous suggestion for the development of the dynamic kriging method.

References


doi:10.1214/ss/1177012413

doi:10.2514/1.12366

Table A1 Comparison of basis-function selection and stepwise methods and rrmse values

<table>
<thead>
<tr>
<th>Selection methods</th>
<th>Selected basis function</th>
<th>RRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stepwise selection</td>
<td>$1, x_1, x_2, x_1 x_2, x_1^2, x_2^2$</td>
<td>0.33</td>
</tr>
<tr>
<td>Genetic algorithm selection</td>
<td>$1, x_1, x_2, x_1 x_2, x_1^2, x_2^2$</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Table A2 Comparison of GA selection and stepwise selection (50 trials)

<table>
<thead>
<tr>
<th>Selection method</th>
<th>Branin–Hoo 20 points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median of rrmse based on the genetic algorithm selection</td>
<td>0.085</td>
</tr>
<tr>
<td>Median of rrmse based on stepwise selection</td>
<td>0.356</td>
</tr>
<tr>
<td>No. of cases (rrmse_{genetic algorithm} &lt; rrmse_{stepwise})</td>
<td>49</td>
</tr>
</tbody>
</table>


