AIAA JOURNAL
Vol. 55, No. 12, December 2017

Weighted Gradient-Enhanced Kriging for High-Dimensional Surrogate Modeling and Design Optimization

Zhong-Hua Han,* Yu Zhang,† Chen-Xing Song,‡ and Ke-Shi Zhang§
Northwestern Polytechnical University, 710072 Xi'an, People's Republic of China

DOI: 10.2514/1.J055842

A novel formulation of gradient-enhanced surrogate model, called weighted gradient-enhanced kriging, is proposed and used in combination with the cheap gradients obtained by the adjoint method to ameliorate the "curse of dimensionality". The core idea is to build a series of submodels with much smaller correlation matrices and then sum them up with appropriate weight coefficients, aiming to avoid the prohibitive cost associated with decomposing the large correlation matrix of a gradient-enhanced kriging. A self-contained derivation of the proposed method is presented, and then it is verified by surrogate modeling test cases. The present method is integrated into a surrogate-based optimizer and tested for design optimizations. It is further demonstrated for inverse design of a transonic wing, parameterized with a number of design variables in the range from 36 to 108, using Reynolds-averaged Navier–Stokes flow and adjoint solvers. It is observed that, for the wing design with 36 and 54 variables, the weighted and conventional gradient-enhanced kriging are comparable, and both are much more efficient than kriging without using any gradient. For the wing design with 72 and 108 variables, the cost of training a gradient-enhanced kriging increases rapidly and becomes prohibitive. In contrast, the cost of training a weighted gradient-enhanced kriging is kept in an acceptable level, which makes it more practical for higher-dimensional problems.

Nomenclature

c_l, c_d, c_m or C_L, C_D, C_M	=	lift, drag, and pitching moment coefficients
$\boldsymbol{F}$	=	regression matrix
f	=	objective function
g	=	inequality constraint function
h	=	equality constraint function
JL	=	joint logarithm likelihood function
L	=	likelihood function
Ma	=	Mach number
m	=	number of dimensions or number of design variables
n	=	number of sampling sites
R	=	correlation function
$\boldsymbol{R}$	=	correlation matrix
Re	=	Reynolds number
$\boldsymbol{r}$	=	correlation vector
$\boldsymbol{S}$	=	sampling sites
s	=	standard deviation
scf	=	spatial correlation function
w, λ	=	weight coefficients for function values and gradients
$\boldsymbol{x}$	=	independent or design variables
Y	=	random function
$\boldsymbol{y_S}$	=	vector of observed function values and gradients
y	=	unknown function
$Z(\cdot)$	=	Gaussian random processes
α	=	angle of attack
$\beta_{0,i}$, β_0	=	global trend of ith sub-gradient-enhanced kriging and WCK
$\partial\boldsymbol{R}$, $\partial^2\boldsymbol{R}$	=	matrices of the first-order and second-order derivatives of correlation function
$\partial\boldsymbol{r}$	=	vector of the first-order derivatives of correlation function
$\boldsymbol{\theta}$	=	hyperparameter vector for spatial correlation function
μ	=	Lagrange multiplier
ξ	=	weighted distance for spatial correlation function
σ^2	=	process variance

Subscripts

i	=	index $\in [1, n]$, referring to ith subsurrogate model
init	=	initial value
k, k'	=	index $\in [1, m]$, referring kth or k'th dimension
$\boldsymbol{S}$	=	sampling sites
sub	=	subsurrogate model

Superscripts

(i), (j)	=	index $\in [1, n]$, referring to the ith or jth sampling site
^	=	approximated value

Received 2 January 2017; revision received 24 May 2017; accepted for publication 1 June 2017; published online 28 July 2017.

*Professor, School of Aeronautics, National Key Laboratory of Science and Technology on Aerodynamic Design and Research, Youyi West Road No. 127, P.O. Box 754; hanzh@nwpu.edu.cn. Member AIAA.

†Ph.D. Student, School of Aeronautics, National Key Laboratory of Science and Technology on Aerodynamic Design and Research, Youyi West Road No. 127, P.O. Box 754; momozhang@126.com. Student Member AIAA.

‡M.S. Student, School of Aeronautics, National Key Laboratory of Science and Technology on Aerodynamic Design and Research, Youyi West Road No. 127, P.O. Box 754; songchenxing@163.com. Student Member AIAA.

§Associate Professor, School of Aeronautics, National Key Laboratory of Science and Technology on Aerodynamic Design and Research, Youyi West Road No. 127, P.O. Box 120; zhangkeshi@nwpu.edu.cn. Member AIAA.

I. Introduction

OVER the past two decades, surrogate-based modeling and optimization have played an increasingly important role in different areas of aerospace engineering [1–8], such as aerodynamic design optimization [9,10], structural optimization [11], and multidisciplinary design optimization of aircraft or spacecraft [12], where high-fidelity and expensive numerical simulations, such as computational fluid dynamics (CFD) or computational solid dynamics (CSD), are often employed. Surrogate model is also called "response surface model", "metamodel", "approximation model", or "emulator" by the literature in different research areas. The representative surrogate models are polynomial response surface model (PRSM) [13,14], kriging [15,16], radial-basis functions (RBFs) [17,18], artificial neural network (ANN) [19,20],

support-vector regression (SVR) [21,22], multivariate interpolation [23], polynomial chaos expansion [24,25], etc. Among them, kriging gets popularity in the fields of aerodynamic design optimization [26–30] and structural and multidisciplinary optimization [8] because it can represent nonlinear and multidimensional functions and has a unique feature of offering a mean-squared-error estimation. A surrogate model is a cheap-to-evaluate approximation model built through the sampled data, which are obtained by evaluating a limited number of sample points in the parameter space via expensive analysis code. In design optimization, a surrogate model can be used to efficiently predict the output of an expensive analysis code at any untried point. Because the computational cost of evaluating a surrogate model is usually negligible, compared to an expensive CFD or CSD simulation, the overall design efficiency could be greatly improved. Despite the continuous advance in surrogate-based modeling and optimization in recent years, the main challenge is still associated with the prohibitive computational cost of building sufficiently accurate surrogate models for high-dimensional design optimizations [8], such as aerodynamic design of complex aircraft configurations parameterized with many variables. To break or at least ameliorate the curse of dimensionality [31,32], a surrogate model enhanced by cheap auxiliary information, such as lower-fidelity data or inexpensive gradients, is of great interest.

The surrogate model incorporating auxiliary information can be divided into two categories. One is the variable-fidelity model or multifidelity model. With assistance of the data obtained by simplified, lower-fidelity physical model or the same physical model on coarser computational grid, the accuracy of a surrogate model for high-fidelity and expensive function can be significantly enhanced [33,34]. In turn, the efficiency for solving an optimization problem can be remarkably improved [35–39]. The other is the gradient-enhanced surrogate model. The inexpensive gradients, obtained by using an adjoint method [40,41] or automatic differentiation [42], can be incorporated into the construction of a surrogate model, and in turn the approximation accuracy can be dramatically improved, especially for high-dimensional problems. Here, we limit our concerns to the kriging enhanced by cheap gradient information. The use of gradient information in the construction of a kriging was first proposed by Morris et al. [43] in 1993. It was then introduced in aerospace applications by Chung and Alonso [44] in 2002, with the gradients inexpensively computed by the adjoint method [40,41]. The big potential of using cheap gradients in the construction of a kriging largely inspirited the research and development in the field of aerodynamic design optimization [45], toward breaking the curse of dimensionality [46].

Gradient-enhanced kriging (GEK) denotes the extension of kriging to models in which inexpensive gradient information is incorporated [47]. Note that GEK was also viewed as one kind of cokriging (e.g., [44,45]), although the terminology "cokriging" generally means the kriging incorporating lower-fidelity auxiliary data (see [5] or [33]). There are two ways of incorporating gradients into the construction of a kriging, which leads to indirect GEK and direct GEK [47]. The idea of an indirect GEK [44,48,49] is that the gradients are reverted to additional function values in the vicinity of the sampling sites, by means of first-order Taylor's expansion. It uses the same mathematical formulation of a kriging but augments the training data with additional function values [50]. The drawbacks of this method are that there is a chance for numerical errors to be introduced while estimating the additional function values from the gradients [44], and it is less robust because the condition number of the correlation matrix can be very large. In case of the direct GEK [5,45,47,51–53], the gradients are directly included in the kriging equations system, assuming that the predictor is defined by augmenting the weighted sum of the functional values with the weighted sum of the gradients. The optimal weights are then obtained by minimizing the mean-squared error (MSE) of the predictor, and the resulting linear equations system includes extra correlations between the function values and the gradients as well as the correlations between the gradients themselves. Theoretically, the direct GEK is more accurate because no numerical errors are introduced due to the step size of the Taylor's expansion, and practice suggests that it is more robust because the condition number of the correlation matrix is generally much smaller [54]. In addition to first-order derivatives, second-order derivatives (Hessian) can also be used to enhance the prediction of a kriging, resulting in the Hessian-enhanced kriging [55], which is beyond the scope of this paper.

Because the incorporation of inexpensive gradients can significantly improve the accuracy of a kriging, the training sample points required to build a surrogate model achieving a desired accuracy can be dramatically reduced, especially for high-dimensional problems. However, practice suggests that when the gradients are incorporated, the correlation matrix is largely expanded, from $n \times n$ to $(n + nm) \times (n + nm)$, where n is the number of sampling sites, and m is the number of dimensions (or design variables). Training such a GEK can be extremely costly, especially when m is very large. It is no doubt that how to break this "bottleneck" becomes an important issue for the community, toward using the GEK to break or at least alleviate the curse of dimensionality for applications such as in the area of aerodynamic design optimizations.

Recall that, for the aerodynamic design optimization (ADO) based on high-fidelity CFD, currently there are three kind of approaches. The first is the gradient-based method. It is very efficient when the gradients are computed by the adjoint method proposed by Jameson [40]. The computational cost of evaluating the partial derivatives of an objective or constraint function is insensitive to the number of design variables, requiring only one flow solution and one adjoint solution. This method becomes the most practical one for the design of complex aerodynamic configurations [56,57] because it can deal with the optimization problems with 100–1000 design variables [58,59] or even more. The drawback of a gradient-based method is that the solution optimality can be sensitive to the initial guesses, and it can become trapped into a local minimum [60]. Although recent study shows that, for many aerodynamic design optimization problems, the design space is not highly multimodal [60], local optima do exist, and a global optimization method is still essential to find the most efficient aerodynamic configuration [60]. The second is gradient-free method. Among the available methods, the metaheuristic optimization algorithms, such as genetic algorithms (GAs), simulated annealing, or the particle swarm algorithm, have good capability of global optimization. However, when this type of algorithm is used, a single ADO usually requires thousands of CFD simulations [61] or even more [62], and in turn, the computational cost can easily exceed the available computational budget. This situation becomes even worse when dealing with complex aircraft configurations parameterized with a large number of design variables. Therefore, the applications of this method are usually limited to either two-dimensional aerodynamic configurations [61] or three-dimensional (3-D) configurations using low-fidelity and fast CFD methods [62]. And the third approach is surrogate-based optimization (SBO) [1–9]. As previously mentioned, the surrogate models enable us to find the global optimum within a very limited number of expensive evaluations [30,38]. Practice suggests that, for the local optimization problems with number of design variables less than around 15, SBO can be as efficient as the gradient-based optimization based on the adjoint method [63], but for a global optimization, SBO can be much more efficient than any of the existing metaheuristic approaches. Although SBO is very promising, it suffers from the curse of dimensionality, especially when the number of design variables is increased up to about 30 and beyond. A remedy is to combine SBO with the adjoint method, and a few researchers have put effort on this topic [51,52,63], toward an efficient global optimization method for higher-dimensional ADO problems.

This paper is motivated by the aspiration of developing a novel formulation of gradient-enhanced surrogate model called weighted gradient-enhanced kriging (WGEK), which is to be used in combination with the adjoint method to ameliorate the curse of dimensionality for high-dimensional surrogate modeling and design optimizations. The core idea is to build a series of submodels with much smaller correlation matrices and then to sum them up with appropriate weight coefficients, aiming to avoid the prohibitive cost associated with decomposing the large correlation matrix of the conventional GEK model. A self-contained derivation about the

theory of WGEK is presented, and the method of tuning its hyperparameters is proposed. WGEK is verified by surrogate modeling test cases and validated by optimizations of analytical functions when integrated into a SBO framework. The effectiveness of WGEK is further demonstrated by aerodynamic inverse design of a transonic wing parameterized with number of design variables in the range from 36 to 108, which exhibits the superiority of the proposed WGEK for higher-dimensional design optimizations.

The remainder of this paper is organized as follows. Section II presents the formulation of the proposed WGEK. Section III is for the verification and validation of WGEK for surrogate modeling problems. Section IV describes the integration of the proposed WGEK to an SBO framework and the application to optimization of analytical functions as well as the inverse designs of a transonic wing. Section VI is for the summary and outlook.

II. Weighted Gradient-Enhanced Kriging Model: A Novel Formulation

A. Main Assumptions and Core Idea

For an m-dimensional problem, suppose that we are concerned with the prediction of an expensive-to-evaluate, high-fidelity aerodynamic function (or other kind of expensive functions) y: $\mathbb{R}^m \to \mathbb{R}$. First, we choose n sample points by using the method of design of experiments (DOE) [64,65] and run flow and adjoint solvers to get the responses of the aerodynamic function as well as its gradients with respect to all the design variables. Then, the sampled data sets $(\boldsymbol{S}, \boldsymbol{y}_S)$ are collected as

$$\begin{aligned}\boldsymbol{S} &= [\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)}, \boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)}, \ldots, \boldsymbol{x}^{(n)}]^T \in \mathbb{R}^{(n+nm)\times m} \\ \boldsymbol{y}_S &= \left[y^{(1)}, \ldots, y^{(n)}, \frac{\partial y^{(1)}}{\partial x_1}, \ldots, \frac{\partial y^{(1)}}{\partial x_m}, \ldots, \frac{\partial y^{(n)}}{\partial x_1}, \ldots, \frac{\partial y^{(n)}}{\partial x_m}\right]^T \in \mathbb{R}^{n+nm}\end{aligned} \tag{1}$$

where $\boldsymbol{S}$ is a matrix with each row vector representing one sampling site, and $\boldsymbol{y}_S$ is a column vector that contains the observed functional values and gradients at all the sampling sites. Note that the superscript (i) in this paper generally denotes the ith sampling site.

For the sampled data set in Eq. (1), the core idea of WGEK is to sum up a series of submodels with appropriate weight coefficients. We assume that each submodel is built through the observed function values at all sampling sites and the gradients at one site only. Therefore, there will be n submodels in total. Please also note that the submodel can be readily extended to include the gradients of two or more sampling sites, in which case the number of submodels will be less than n. This can make the WGEK more general but is beyond the scope of this paper.

A literature survey shows that the idea of using "submodel" techniques has been successfully applied by the methods of mixture of experts (ME) [66], moving-window kriging (MWK) [67], and locally optimized covariance kriging (LOCK) [68]. Different from ME, MWK, or LOCK, whose aim is to address the problem when a large number of sampling sites (or functional evaluations) are involved in a model where the function behaviors differ in different regions of the parameter space, the aim of WGEK is mainly to address the problem of prohibitive computational cost caused by training the surrogate model enhanced by high-dimensional gradients. The WGEK can be combined with the methods of ME, MWK, and LOCK, which would be the topic of the future work beyond this paper.

Because every submodel of WGEK uses all the sampling sites, we could reasonably assume that they are corresponding to a single random process:

$$Y(\boldsymbol{x}) = \beta_0 + Z(\boldsymbol{x}) \tag{2}$$

where $Z(\boldsymbol{x})$ is a stationary random process having zero mean and covariance of

$$\begin{aligned}&\mathrm{Cov}[Z(\boldsymbol{x}^{(i)}), Z(\boldsymbol{x}^{(j)})] = \sigma^2 R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}), \\ &\mathrm{Cov}\left[\frac{\partial Z(\boldsymbol{x}^{(i)})}{\partial x_k^{(i)}}, Z(\boldsymbol{x}^{(j)})\right] = \sigma^2 \frac{\partial R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})}{\partial x_k^{(i)}} = -\mathrm{Cov}\left[Z(\boldsymbol{x}^{(i)}), \frac{\partial Z(\boldsymbol{x}^{(j)})}{\partial x_k^{(j)}}\right], \\ &\mathrm{Cov}\left[\frac{\partial Z(\boldsymbol{x}^{(i)})}{\partial x_k^{(i)}}, \frac{\partial Z(\boldsymbol{x}^{(j)})}{\partial x_{k'}^{(j)}}\right] = \sigma^2 \frac{\partial^2 R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})}{\partial x_k^{(i)} \partial x_{k'}^{(j)}}\end{aligned} \tag{3}$$

where $\sigma^2(\boldsymbol{x}) \equiv \sigma^2$ is the process variance of $Z(\cdot)$, and R is the spatial correlation function. Note that $\partial R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})/\partial x_k^{(i)}$ is a partial derivative taken at the ith sampling site and with respect to the kth component of $\boldsymbol{x}$.

The prediction of a WGEK for an unknown expensive function at any untried $\boldsymbol{x}$ is formally defined as the weighted sum of n submodels as

$$\hat{y}(\boldsymbol{x}) = \sum_{i=1}^{n} w_i \hat{y}_i(\boldsymbol{x}) \tag{4}$$

where w_i is the weight coefficient, and $\hat{y}_i(\boldsymbol{x})$ stands for the prediction of the ith subsurrogate model. Please note that $\hat{y}(\boldsymbol{x})$ should simultaneously feature the observed function values and gradients at all the sampling sites because the preceding predictor is an interpolation model. Through reasoning, we observed that the interpolation condition can be ensured if the weight coefficient w_i satisfies the following equations:

$$\begin{aligned}&w_1 + w_2 + \ldots + w_n = 1, \qquad \forall\, \boldsymbol{x}, \\ &\begin{cases} w_1(\boldsymbol{x}^{(1)}) = 1, w_2(\boldsymbol{x}^{(1)}) = 0, \ldots, w_n(\boldsymbol{x}^{(1)}) = 0 \\ w_1(\boldsymbol{x}^{(2)}) = 0, w_2(\boldsymbol{x}^{(2)}) = 1, \ldots, w_n(\boldsymbol{x}^{(2)}) = 0 \\ \qquad\qquad \ldots \\ w_1(\boldsymbol{x}^{(n)}) = 0, w_2(\boldsymbol{x}^{(n)}) = 0, \ldots, w_n(\boldsymbol{x}^{(n)}) = 1 \end{cases}\end{aligned} \tag{5}$$

$$\begin{cases} \left.\frac{\partial(w_1\hat{y}_1)}{\partial x_j}\right|_{\boldsymbol{x}=\boldsymbol{x}^{(1)}} = \frac{\partial y^{(1)}}{\partial x_j}, \left.\frac{\partial(w_2\hat{y}_2)}{\partial x_j}\right|_{\boldsymbol{x}=\boldsymbol{x}^{(1)}} = 0.0, \ldots, \left.\frac{\partial(w_n\hat{y}_n)}{\partial x_j}\right|_{\boldsymbol{x}=\boldsymbol{x}^{(1)}} = 0.0 \\ \left.\frac{\partial(w_1\hat{y}_1)}{\partial x_j}\right|_{\boldsymbol{x}=\boldsymbol{x}^{(2)}} = 0.0, \left.\frac{\partial(w_2\hat{y}_2)}{\partial x_j}\right|_{\boldsymbol{x}=\boldsymbol{x}^{(2)}} = \frac{\partial y^{(2)}}{\partial x_j}, \ldots, \left.\frac{\partial(w_n\hat{y}_n)}{\partial x_j}\right|_{\boldsymbol{x}=\boldsymbol{x}^{(2)}} = 0.0 \\ \ldots \\ \left.\frac{\partial(w_1\hat{y}_1)}{\partial x_j}\right|_{\boldsymbol{x}=\boldsymbol{x}^{(n)}} = 0.0, \left.\frac{\partial(w_2\hat{y}_2)}{\partial x_j}\right|_{\boldsymbol{x}=\boldsymbol{x}^{(n)}} = 0.0, \ldots, \left.\frac{\partial(w_n\hat{y}_n)}{\partial x_j}\right|_{\boldsymbol{x}=\boldsymbol{x}^{(n)}} = \frac{\partial y^{(n)}}{\partial x_j} \end{cases}, \quad j = 1, \ldots, m \tag{6}$$

Although there might be various type of ways for choosing appropriate weight coefficient, we found that w_i can be obtained through building a special kriging called weight coefficient kriging (WCK), as will be presented in Sec. II.C.

B. Submodels

For building the ith submodel in Eq. (4), the subsampled data set $(\boldsymbol{S}_i, \boldsymbol{y}_{S,i})$ is collected as

$$\boldsymbol{S}_i = \left[\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)}, \underbrace{\boldsymbol{x}^{(i)}, \ldots, \boldsymbol{x}^{(i)}}_{m}\right]^T \in \mathbb{R}^{(n+m)\times m},$$

$$\boldsymbol{y}_{S,i} = \left[y^{(1)}, \ldots, y^{(n)}, \underbrace{\partial y^{(i)}/\partial x_1, \ldots, \partial y^{(i)}/\partial x_m}_{m}\right]^T \in \mathbb{R}^{n+m}, \quad i = 1, \ldots, n \tag{7}$$

This implies that the ith submodel is a GEK model using the observed functional values at all the sampling sites and the gradients at the ith sampling site only. The predictor of the ith sub-GEK is given by (see [47] for detailed derivation)

$$\hat{y}_i(\boldsymbol{x}) = \beta_0 + \boldsymbol{r}_i^T \boldsymbol{R}_i^{-1}(\boldsymbol{y}_{S,i} - \beta_0 \boldsymbol{F}_i) \tag{8}$$

where

$$\boldsymbol{F}_i = \left(\underbrace{1, \ldots, 1}_{n}, \underbrace{0, \ldots, 0}_{m}\right)^T \in \mathbb{R}^{n+m} \tag{9}$$

and $\boldsymbol{R}_i$ and $\boldsymbol{r}_i$ are the correlation matrix and correlation vector for the ith sub-GEK, respectively. They are

$$\boldsymbol{R}_i = \begin{bmatrix} \boldsymbol{R} & \partial \boldsymbol{R}_i \\ \partial \boldsymbol{R}_i^T & \partial^2 \boldsymbol{R}_i \end{bmatrix} \in \mathbb{R}^{(n+m)\times(n+m)}, \quad \boldsymbol{r}_i = \begin{bmatrix} \boldsymbol{r} \\ \partial \boldsymbol{r}_i \end{bmatrix} \in \mathbb{R}^{n+m} \tag{10}$$

where the correlation matrix $\boldsymbol{R}$, first-order derivative $\partial \boldsymbol{R}_i$, and second-order derivative of $\partial^2 \boldsymbol{R}_i$ are given by

$$\boldsymbol{R} = \begin{bmatrix} R(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(1)}) & \cdots & R(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(n)}) \\ \vdots & \ddots & \vdots \\ R(\boldsymbol{x}^{(n)}, \boldsymbol{x}^{(1)}) & \cdots & R(\boldsymbol{x}^{(n)}, \boldsymbol{x}^{(n)}) \end{bmatrix} \in \mathbb{R}^{n\times n},$$

$$\partial \boldsymbol{R}_i = \begin{bmatrix} \frac{\partial R(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(i)})}{\partial x_1^{(i)}} & \cdots & \frac{\partial R(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(i)})}{\partial x_m^{(i)}} \\ \vdots & \ddots & \vdots \\ \frac{\partial R(\boldsymbol{x}^{(n)}, \boldsymbol{x}^{(i)})}{\partial x_1^{(i)}} & \cdots & \frac{\partial R(\boldsymbol{x}^{(n)}, \boldsymbol{x}^{(i)})}{\partial x_m^{(i)}} \end{bmatrix} \in \mathbb{R}^{n\times m},$$

$$\partial^2 \boldsymbol{R}_i = \begin{bmatrix} \frac{\partial^2 R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(i)})}{\partial^2 x_1^{(i)}} & \cdots & \frac{\partial^2 R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(i)})}{\partial x_1^{(i)} \partial x_m^{(i)}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{i)})}{\partial x_m^{(i)} \partial x_1^{(i)}} & \cdots & \frac{\partial^2 R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(i)})}{\partial^2 x_m^{(i)}} \end{bmatrix} \in \mathbb{R}^{m\times m} \tag{11}$$

The correlation vector $\boldsymbol{r}$ and its first-order derivative $\partial \boldsymbol{r}_i$ are defined as

$$\boldsymbol{r} = [R(\boldsymbol{x}^{(1)}, \boldsymbol{x}), \ldots, R(\boldsymbol{x}^{(n)}, \boldsymbol{x})]^T \in \mathbb{R}^n,$$
$$\partial \boldsymbol{r}_i = \left[\frac{\partial R(\boldsymbol{x}^{(i)}, \boldsymbol{x})}{\partial x_1^{(i)}}, \ldots, \frac{\partial R(\boldsymbol{x}^{(i)}, \boldsymbol{x})}{\partial x_m^{(i)}}\right]^T \in \mathbb{R}^m \tag{12}$$

The MSE of the ith sub-GEK predictor is given by

$$\mathrm{MSE}\{\hat{y}_i(\boldsymbol{x})\} = s_i^2(\boldsymbol{x}) = \sigma^2[1.0 - \boldsymbol{r}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{r}_i + (1 - \boldsymbol{F}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{r}_i)^2 / (\boldsymbol{F}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{F}_i)] \tag{13}$$

C. Weight Coefficients and Model Predictor

We observed that the weight coefficients can be obtained from building a special kriging, WCK, in which the prediction at any untried $\boldsymbol{x}$ is defined by

$$\hat{y}(\boldsymbol{x}) = \sum_{i=1}^{n} w_i \hat{y}_i(\boldsymbol{x}) \tag{14}$$

A WCK for any untried $\boldsymbol{x}$ is built through the data set $(\boldsymbol{S}, \hat{\boldsymbol{y}}_{S,\mathrm{sub}})$ given by

$$\boldsymbol{S} = [\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)}]^T \in \mathbb{R}^{n\times m},$$
$$\hat{\boldsymbol{y}}_{S,\mathrm{sub}}(\boldsymbol{x}) = [\hat{y}_1(\boldsymbol{x}), \ldots, \hat{y}_n(\boldsymbol{x})]^T \in \mathbb{R}^n \tag{15}$$

where $\hat{y}_i(\boldsymbol{x})$ denotes the prediction of the ith sub-GEK at the untried $\boldsymbol{x}$. Note that this kriging is a special surrogate model because its sampled data $\hat{\boldsymbol{y}}_{S,\mathrm{sub}}(\boldsymbol{x})$ are changing with $\boldsymbol{x}$.

Taking the ordinary-type kriging as an example, the optimal weight coefficients are found by solving the following linear equations:

$$\begin{bmatrix} \boldsymbol{R} & \boldsymbol{F} \\ \boldsymbol{F}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{w} \\ \mu \end{bmatrix} = \begin{bmatrix} \boldsymbol{r} \\ 1 \end{bmatrix} \tag{16}$$

where μ is the Lagrange multiplier. $\boldsymbol{R}$ and $\boldsymbol{r}$ are given by Eqs. (11) and (12), and

$$\boldsymbol{w} = [w_1, \ldots, w_n]^T \quad \boldsymbol{F} = [1, \ldots, 1]^T \in \mathbb{R}^n \tag{17}$$

After determining the submodels and weight coefficients, the resulting WGEK predictor is of the form

$$\hat{y}(\boldsymbol{x}) = \beta_0 + \boldsymbol{r}^T(\boldsymbol{x}) \boldsymbol{R}^{-1} [\hat{\boldsymbol{y}}_{S,\mathrm{sub}}(\boldsymbol{x}) - \beta_0 \boldsymbol{F}] \tag{18}$$

Because the prediction of the submodels can be generally considered to be normal distributed $\hat{Y}_i \sim [\hat{y}_i, s_i^2]$, $i = 1, \ldots, n$, the mean value of the WGEK prediction is given by Eq. (18), and the MSE can be derived according to the theory of combining a group of normal distributions with different mean values and variances:

$$\begin{aligned} \mathrm{MSE}[\hat{y}(\boldsymbol{x})]) = s^2(\boldsymbol{x}) &= \mathrm{var}\left(\sum_{i=1}^{n} w_i \hat{Y}_i\right) = \sum_{i=1}^{n} w_i^2 \mathrm{var}(\hat{Y}_i) \\ &\quad + \sum_{i,j=1, i\neq j}^{n} 2 w_i w_j \mathrm{cov}(\hat{Y}_i, \hat{Y}_j) \\ &= \sum_{i=1}^{n} w_i^2 s_i^2 + \sum_{i,j=1, i\neq j}^{n} 2 w_i w_j s_i s_j \\ &= \left(\sum_{i=1}^{n} w_i s_i\right)^2 \end{aligned} \tag{19}$$

in which

$$\mathrm{cov}(\hat{Y}_i, \hat{Y}_j) = s_i s_j R(\boldsymbol{x}, \boldsymbol{x}) = s_i s_j \tag{20}$$

because the random variables $\hat{Y}_i$, $\hat{Y}_j$ at the same location $\boldsymbol{x}$ are fully correlated.

D. Correlation Functions and Hyperparameter Tuning

Here, we focus on a family of correlation functions that are of the form

$$R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) = \prod_{k=1}^{m} \mathrm{scf}\left(\theta_k, \left|x_k^{(i)} - x_k^{(j)}\right|\right) \tag{21}$$

where "scf" denotes the spatial correlation function. Note that, for the success of a GEK, the correlation function has to be at least second-order differentiable; otherwise, the use of gradients will not enhance the prediction. Thus far, several types of correlation functions can be used by GEK, such as "Gaussian function [47]", "cubic spline function", "Wendland C4 function", "Hermite function", etc. Because the Gaussian function can lead to a correlation matrix with higher condition number [69], the cubic spline function [70] is employed in this paper:

$$\mathrm{scf}_k \equiv \mathrm{scf}\left(\theta_k, \left|x_k^{(i)} - x_k^{(j)}\right|\right) = \begin{cases} 1\text{–}15\xi_k^2 + 30\xi_k^3 & \text{if } 0 \le \xi_k \le 0.2 \\ 1.25(1-\xi_k)^3 & \text{if } 0.2 < \xi_k < 1 \\ 0 & \text{otherwise} \end{cases} \quad (22)$$

where

$$\xi_k = \theta_k \left|x_k^{(i)} - x_k^{(j)}\right|, \qquad k = 1, \ldots, m \quad (23)$$

The first-order partial derivative of is given by

$$\frac{\partial R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})}{\partial x_k^{(i)}} = \frac{\partial (scf_k)}{\partial x_k^{(i)}} \prod_{l=1, l\neq k}^{m} \mathrm{scf}_l \quad (24)$$

where

$$\frac{\partial(\mathrm{scf}_k)}{\partial x_k^{(i)}} = \begin{cases} 30\theta_k\xi_k(3\xi_k - 1)\mathrm{sign}(x_k^{(i)} - x_k^{(j)}) & \text{if } 0 \le \xi_k \le 0.2 \\ -3.75\theta_k(\xi_k - 1)^2\mathrm{sign}(x_k^{(i)} - x_k^{(j)}) & \text{if } 0.2 < \xi_k < 1 \\ 0 & \text{otherwise} \end{cases} \quad (25)$$

and "sign" denotes a sign function of

$$\mathrm{sign}(x) = \begin{cases} -1 & \text{if} \quad x < 0 \\ 0 & \text{if} \quad x = 0 \\ 1 & \text{if} \quad x > 0 \end{cases} \quad (26)$$

The second-order partial derivative of R is given by

$$\frac{\partial^2 R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})}{\partial x_k^{(i)} \partial x_{k'}^{(j)}} = \begin{cases} \dfrac{\partial^2(\mathrm{scf}_k)}{\partial x_k^{(i)} \partial x_k^{(j)}} \displaystyle\prod_{l=1, l\neq k}^{m} \mathrm{scf}_l \text{ if } k = k' \\ -\dfrac{\partial(\mathrm{scf}_k)}{\partial x_k^{(i)}} \dfrac{\partial(\mathrm{scf}_{k'})}{\partial x_{k'}^{(i)}} \displaystyle\prod_{l=1, l\neq k, k'}^{m} \mathrm{scf}_l \text{ otherwise} \end{cases} \quad (27)$$

where

$$\frac{\partial^2(\mathrm{scf}_k)}{\partial x_k^{(i)} \partial x_k^{(j)}} = \begin{cases} -30\theta_k^2(6\xi_k - 1) & \text{if} \quad 0 \le \xi_k \le 0.2 \\ 7.5\theta_k^2(\xi_k - 1) & \text{if} \quad 0.2 < \xi_k < 1 \\ 0 & \text{otherwise} \end{cases} \quad (28)$$

In Eq. (27), $\partial(\mathrm{scf}_k)/\partial x_k^{(i)}$ and $\partial(\mathrm{scf}_{k'})/\partial x_{k'}^{(i)}$ can be calculated according to (25). This paper might be the first time to present the detailed expressions of the first- and second-order partial derivatives of the cubic spline function, which is expected to be helpful for the readers to follow this method.

Then, we propose a method of tuning the hyperparameters of WGEK. Because we already assume that all the submodels correspond to the same random process, we can reasonably assume that all the submodels share the same $\boldsymbol{\theta}$, σ^2, and β_0. For the ith submodel, the likelihood function is

$$L_i(\beta_0, \sigma^2, \boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi\sigma^2)^{n+m}|\boldsymbol{R}_i|}} \exp\left(-\frac{1}{2}\frac{(\boldsymbol{y}_{S,i} - \beta_0 \boldsymbol{F}_i)^{\mathrm{T}} \boldsymbol{R}_i^{-1} (\boldsymbol{y}_{S,i} - \beta_0 \boldsymbol{F}_i)}{\sigma^2}\right) \quad (29)$$

Taking the logarithm, one can get

$$\ell_n L_i = -\frac{(n+m)}{2}\ell_n(\sigma^2) - \frac{1}{2}\ell_n|\boldsymbol{R}_i| - \frac{1}{2}\frac{(\boldsymbol{y}_{S,i} - \beta_0 \boldsymbol{F}_i)^T \boldsymbol{R}_i^{-1} (\boldsymbol{y}_{S,i} - \beta_0 \boldsymbol{F}_i)}{\sigma^2} - \frac{(n+m)}{2}\ell_n(2\pi) \quad (30)$$

Similar to [67], a uniform aggregation is used to get the joint logarithm likelihood function as

$$\begin{aligned} &\max. \quad \mathrm{JL} = (\ell_n L_1 + \ell_n L_2 + \cdots + \ell_n L_n)/n \\ &\text{w.r.t.} \quad \beta_0,\ \sigma^2,\ \boldsymbol{\theta} \end{aligned} \quad (31)$$

However, other kinds of aggregation strategies can also be used to consider the underlying function behavior. Taking the partial derivatives of JL with respect to β_0 and σ^2, and forcing them to zero, we get the analytical optima

$$\begin{aligned} \beta_0(\boldsymbol{\theta}) &= \sum_i^n (\boldsymbol{F}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{y}_{S,i}) / \sum_i^n (\boldsymbol{F}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{F}_i) \\ \sigma^2(\beta_0, \boldsymbol{\theta}) &= \frac{1}{n(n+m)} \sum_{i=1}^{n} (\boldsymbol{y}_{S,i} - \beta_0 \boldsymbol{F}_i)^T \boldsymbol{R}_i^{-1} (\boldsymbol{y}_{\mathrm{S}.i} - \beta_0 \boldsymbol{F}_i) \end{aligned} \quad (32)$$

Because there is no analytical solution for the optimal $\boldsymbol{\theta}$, we have to use numerical optimization algorithm [71] to maximize the concentrated joint logarithm likelihood function:

$$\boldsymbol{\theta} = \arg\max_{\boldsymbol{\theta}>0} \left[-(n+m)\,\ell_n(\sigma^2) - \frac{1}{n}\sum_{i=1}^{n} \ell_n |\boldsymbol{R}_i|\right] \quad (33)$$

In this paper, we use an improved version of Hooke and Jeeves pattern search method (by using multistarts search and a trust region method) to solve the preceding optimization problem. In addition, we normalize the design variables $\boldsymbol{x}$ to the range [0.0, 1.0] and limit the searching of optimal $\boldsymbol{\theta}$ to the range [0.001, 1.0] for cubic spline function, according to our past numerical experiments, to make the model more robust. For the WCK model, we temporarily assume that it shares the same hypermeters with the submodels, to make the resulting surrogate model continuous and smooth.

E. Discussion About the Benefit of Present Method

The main benefit is that the computational cost of training a gradient-enhanced surrogate model can be greatly reduced, especially for high-dimensional problems. For an m-dimensional problem, the cost of training a conventional GEK with n sampling sites is mainly associated with decomposing the correlation matrix as large as $(n + nm) \times (n + nm)$. The floating-point operations per second (FLOPS) of decomposing such a matrix by Cholesky method is around $O(n^3(1+m)^3)$. In contrast, when training a WGEK, we only need to decompose n much smaller correlation matrices $\boldsymbol{R}_i$, whose size is as small as $(n + m) \times (n + m)$, and the total FLOPs of decomposing all of the matrices is around $O(n(n+m)^3)$. As a result, training a WGEK is about $n^2(1+m)^3/(n+m)^3$ times faster than training a GEK. The comparison of FLOPs for different numbers of dimensions and sampling sites is plotted in Fig. 1 and listed in Table 1. It can be seen that, along with the increase of numbers of dimensions and sampling points, the cost of training a GEK raises very quickly; in contrast, the cost of training a WGEK rises slowly and is kept in an acceptable level. Besides, the tuning and prediction of the submodels can be easily parallelized, which can makes it even more efficient, when running on a computer with multiple cores or on a cluster.

III. Verification and Validation of Weighted Gradient-Enhanced Kriging Model

A. One-Dimensional Analytical Test Case

In this subsection, the following one-dimensional test function is used to explain the typical procedure of building a WGEK and to verify its correctness:

$$f(x) = e^{-x} + \sin(x) + \cos(3x) + 0.2x + 1.0, \qquad x \in [0.2, 6.0] \quad (34)$$

Three sampling sites (blue circles, $\boldsymbol{x} = \{1.65, 3.1, 4.55\}$) are chosen, and the functional values as well as the gradients are computed analytically. The procedure of building the WGEK is

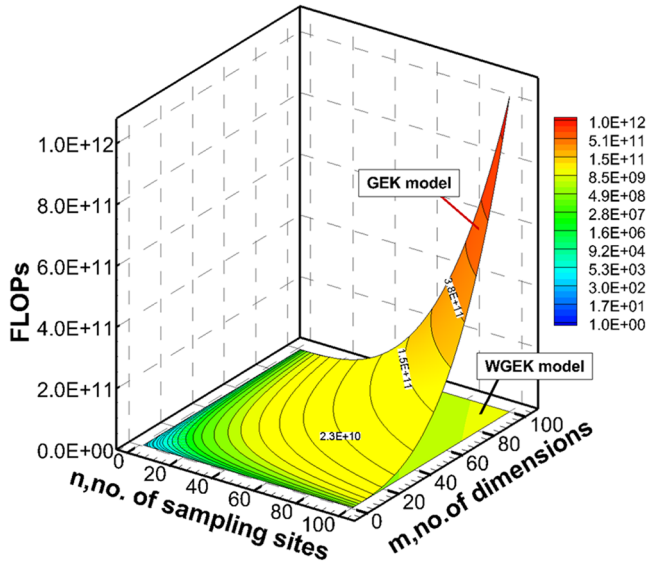


Fig. 1 Comparison of FLOPs of decomposing the correlation matrices of GEK and proposed WGEK.

sketched in Fig. 2. The first submodel, as sketched in Fig. 2a, is built using the functional values at all the sampling sites and the gradient at the first sampling site only. One can see that this model does pass through all the sample points and features the observed gradient at the first sample site. The second and third submodels, as shown in Figs. 2b and 2c, are built in a similar way, but using the gradients at second and third sites, respectively. Summing up all the three submodels with appropriate weight coefficients defined by the WCK model (see Fig. 2d), the WGEK is obtained and plotted in Fig. 2e. It is shown that the resulting WGEK precisely features the observed functional values and gradients at all the sample sites, which verifies the correctness of the proposed method. The results of the conventional GEK are also plotted in Fig. 2e, which shows that the WGEK model is comparable to that of the GEK model. Note that, in Fig. 2, the square root of the MSE (RMSE) is the model predicted error, not the real error.

B. Two-Dimensional Analytical Test Case

Then, a two-dimensional six-hump camelback function taken from [72] is employed to validate the WGEK:

$$f(x_1, x_2) = \left(4 - 2.1x_1^2 + \frac{x_1^4}{3}\right)x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2$$
$$x_1 \in [-2, 2], \qquad x_2 \in [-1, 1] \tag{35}$$

Fifteen sampling sites are chosen randomly, and the functional values as well as the corresponding gradients (30 partial derivatives) are obtained analytically. For the purpose of comparison, the conventional GEK and kriging are also built, as shown in Figs. 3 and 4. All the surrogate models are compared with the exact function by plotting the isolines (see Fig. 3), which shows that both the WGEK and GEK are significantly more accurate than the kriging. One can also observe that the WGEK is slightly worse than the GEK. This is not a surprise because the correlations between the gradients at different sampling sites, for a WGEK, are not explicitly taken into account. In addition, in Fig. 4, it is shown that both the predicted RMSEs of the WGEK and GEK models are smaller than that of the kriging model, and the RMSE of the WGEK model is slightly larger than that of the GEK.

Table 1 Speed-up ratios of proposed WGEK against GEK for typical numbers of dimensions and samples

Number of dimensions (m)	Number of sample points (n)	Speed-up ratio $n^2(1+m)^3/(n+m)^3$
1	5	0.9
2	5	2.0
5	10	6.4
10	10	16.6
10	20	19.7
30	30	124.1
30	60	147.1
50	50	331.0
50	100	393.0
100	100	1,287.9
100	200	1526.4
500	500	31,437.9
500	1000	37,259.7

C. High-Dimensional Analytical Test Case

In this subsection, we adopt the following sum-square function to evaluate the accuracy loss of the proposed WGEK model against GEK for higher-dimensional problems:

$$f(\boldsymbol{x}) = \sum_{i=1}^{m} i x_i^2, \qquad x \in [-5, 5], \qquad i = 1, \ldots, m \tag{36}$$

where m is the number of dimensions. A correlation coefficient [73] r^2 is introduced to assess the accuracy of the kriging, GEK, and WGEK models for the sum-square functions of varying dimensions:

$$r^2 = \left(\frac{\sigma_{f\hat{f}}}{\sqrt{\sigma_f \sigma_{\hat{f}}}}\right)^2 = \left(\frac{N\sum f\hat{f} - \sum f \sum \hat{f}}{\sqrt{[N\sum f^2 - (\sum f)^2][N\sum \hat{f}^2 - (\sum \hat{f})^2]}}\right)^2 \tag{37}$$

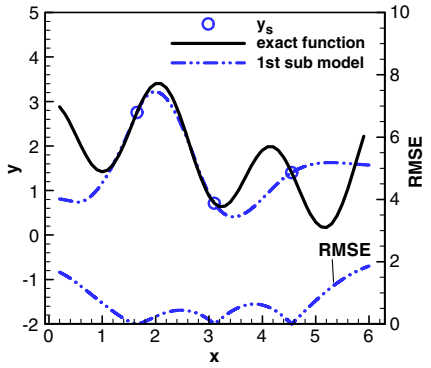
where N is the number of test points, and f and $\hat{f}$ denote the exact functional value and predicted value by surrogate model, respectively. When $r^2 = 1$, the surrogate model is perfectly accurate, whereas $r^2 = 0$ indicates a bad approximation.

For 5, 10, and 20-dimensional sum-square functions, the surrogate models are built using increasing numbers of sample sites, which are chosen by Latin hypercube sampling (LHS) [74] method. Additional 200 test points, generated by uniform design (UD) [75], are used to evaluate the accuracy of the surrogate models. The growth of the correlation coefficient r^2 along with the increase of number of sample sites is plotted in Fig. 5. Because both the functional values and gradients are used in GEK and WGEK, the number of sample points for constructing a GEK and WGEK model is doubled, assuming that the computational cost of a gradient evaluation is similar to that of a functional evaluation like the adjoint method. It is shown that the conventional GEK model is more accurate than the WGEK model. Again, this is not a surprise because the core idea of the WGEK model is to greatly reduce the cost of tuning the surrogate model, and the decreased accuracy is a price to pay. Nevertheless, we can see that the WGEK model is still much more accurate than the kriging, which is beneficial when used for optimization.

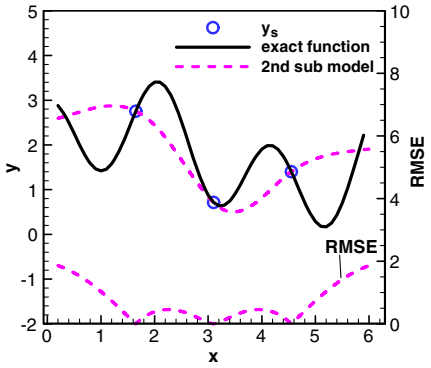
IV. Application to Design Optimizations

A. Integration into a Surrogate-Based Optimization Code

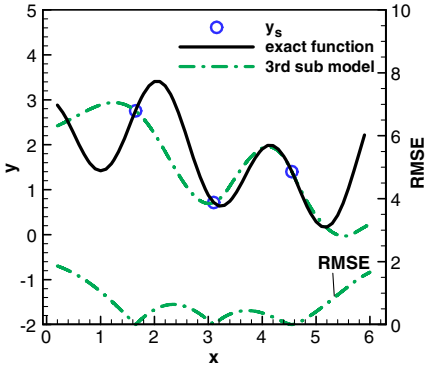
The WGEK, as a novel gradient-enhanced surrogate model, was integrated into an in-house code called "SurroOpt" [6,76]. SurroOpt, a surrogate-based optimization code, was developed mainly for academic research and engineering designs driven by high-fidelity, expensive numerical simulations. SurroOpt can be used to solve arbitrary single and multi-objective, unconstrained and constrained optimization problems with continuous and smooth design space. It has built-in modern DOE methods suited for deterministic computer experiments, such as LHS, UD, and Monte Carlo design. A variety of surrogate models, such as PRSM, kriging and its variants (GEK [47], cokriging [33], HK [34]), RBFs, ANN, SVR [22], etc., were implemented. A couple of infill-sampling criteria and dedicated constraint handling methods were implemented, such as minimizing



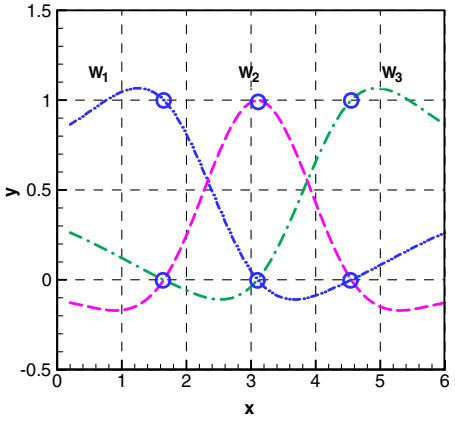
a) First submodel (three function values + one gradient at first sampling site)



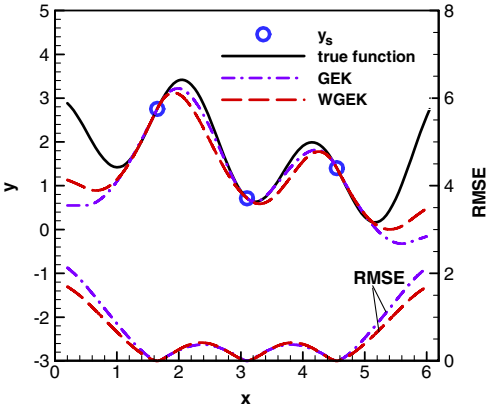
b) Second submodel (three function values + one gradient at second sampling site)



c) Third submodel (three function values + one gradient at third sampling site)

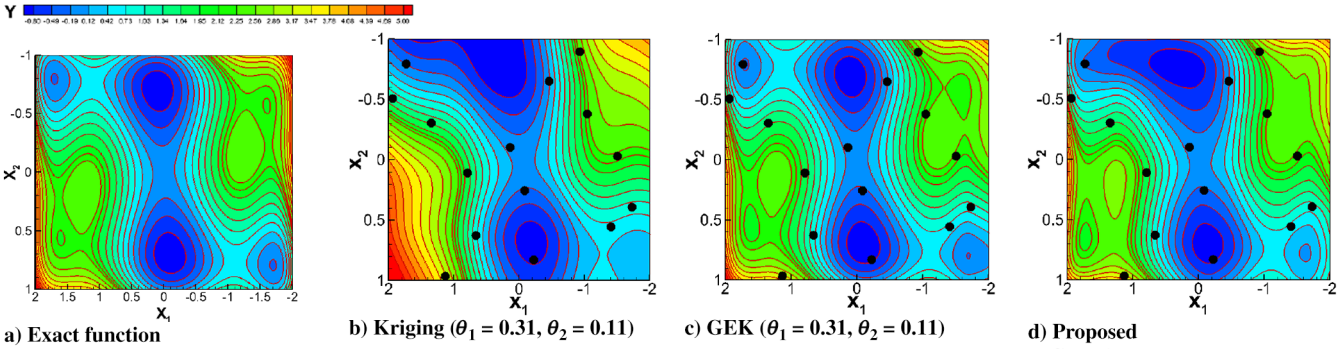


d) Weight coefficients for summing up three sub-GEK models (interpolation condition ensured)



e) Resulting WGEK model and comparison with GEK model ($\theta_{GEK} = \theta_{WGEK} = 1.0$)

Fig. 2 Verification and validation of proposed WGEK for a one-dimensional test function.



a) Exact function

b) Kriging ($\theta_1 = 0.31$, $\theta_2 = 0.11$)

c) GEK ($\theta_1 = 0.31$, $\theta_2 = 0.11$)

d) Proposed WGEK ($\theta_1 = 0.31$, $\theta_2 = 0.11$)

Fig. 3 Comparison of response surface by kriging, GEK, and WGEK models for a six-hump camelback test function.

surrogate predictor (MSP) [76], expected improvement [77,78], probability of improvement [5], mean-squared error (MSE) [79,80], lower-confidence bounding [51,81], target searching [74], and parallel infilling [30]. Some well-accepted and highly matured optimization algorithms, such as Hooke and Jeeves pattern search, Simplex, BFGS quasi-Newton's method, sequential quadratic programming (SQP), and single/multi objective genetic algorithms (GAs) [82], are employed to solve the suboptimization(s), in which the cost function(s) and constraint function(s) are evaluated by the cheap surrogate models. SurroOpt has been paralleled by message passing interface [30], which allows the user to run the code with multiple cores to speed up the optimization process.

For an m-dimensional problem, SurroOpt is capable of solving the following generic constrained optimization problem within a continuous (and smooth) design space:

$$\begin{aligned} \text{min.} \quad & f_{\text{obj},1}(\boldsymbol{x}), f_{\text{obj},2}(\boldsymbol{x}), \ldots, f_{\text{obj},N}(\boldsymbol{x}) \\ \text{w.r.t.} \quad & \boldsymbol{x}_l \le \boldsymbol{x} \le \boldsymbol{x}_u \\ \text{s.t.} \quad & h_i(\boldsymbol{x}) = 0, \qquad i = 1, \ldots, N_E \\ & g_j(\boldsymbol{x}) \le 0, \qquad j = 1, \ldots, N_C \end{aligned} \tag{38}$$

where $f_{\text{obj},1}(\boldsymbol{x}), \ldots, f_{\text{obj},N}(\boldsymbol{x})$ are the objective functions; $\boldsymbol{x}_l$ and $\boldsymbol{x}_u$ are the lower and upper bounds of the design variables $\boldsymbol{x} \in \mathbb{R}^m$, respectively; $h_i(\boldsymbol{x})$ and $g_j(\boldsymbol{x})$ denote the ith equality constraint function and jth inequality constraint function, respectively; and N_E and N_C are the numbers of equality and inequality constraints, respectively. The optimization framework of SurroOpt can be found in [76].

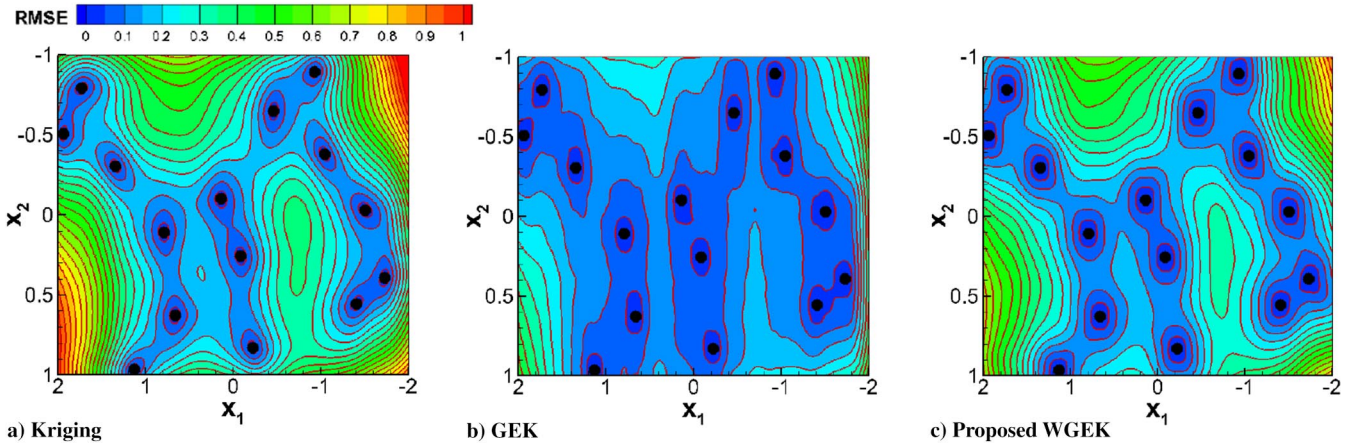


Fig. 4 Comparison of RMSEs predicted by kriging, GEK, and WGEK models for a six-hump camelback test function.

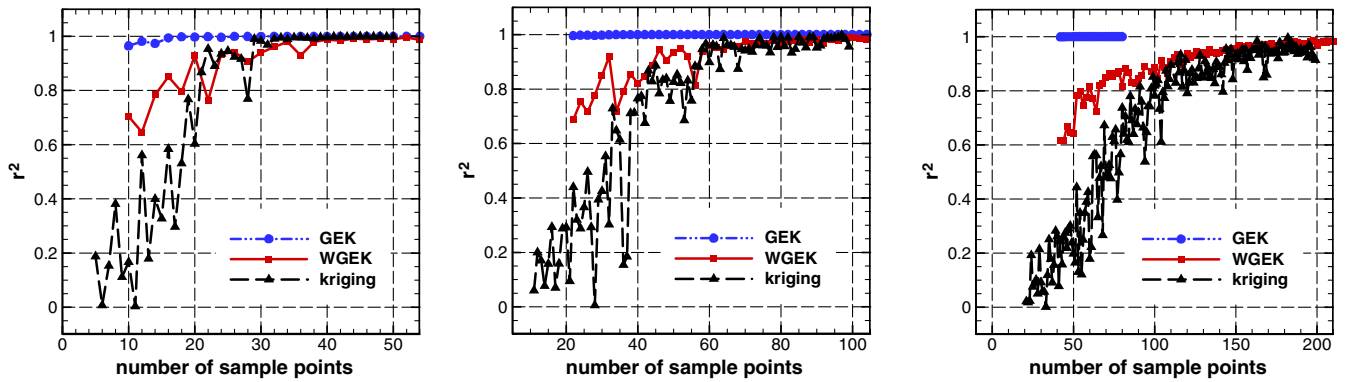


Fig. 5 Evaluation of kriging, GEK, and WGEK models for higher-dimensional test functions.

B. Optimization of a Two-Dimensional Benchmark Test Function with Constraint

A modified Branin–Hoo function representative of engineering design problems, taken from [5], is employed, and the mathematical model is

$$\text{min.} \quad f(\boldsymbol{x}) = \left[15x_2 - \frac{5.1}{4\pi^2}(15x_1 - 5)^2 + \frac{5}{\pi}(15x_1 - 5) - 6\right]^2 + 10\left[\left(1 - \frac{1}{8\pi}\right)\cos(15x_1 - 5) + 1\right] + 5x_1$$

$$\text{w.r.t.} \quad x_1, x_2 \in [0, 1]$$

$$\text{s.t.} \quad g(\boldsymbol{x}) = 0.2 - x_1 x_2 < 0.0 \tag{39}$$

The design space is sketched in Fig. 6; the region above the red line represents the feasible region, and the red cross symbol denotes the global optimum. There are three extrema, and the global optimum is located at $(x_1^*, x_2^*) = (0.967586, 0.206700)$ with an optimal functional value of $f(x_1^*, x_2^*) = 5.575664$. Figure 7 shows the optimization convergence histories. The surrogate optimizations are repeated 20 times to study the influence of the randomness in the initial sampling and suboptimization algorithms (when GA is used). Please note that the convergence histories are plotted in a manner that the lines show the variation of the best objective function value observed thus far versus the number of functional evaluations (excluding initial sampling), and the vertical bars represent the standard deviation of the objective function caused by repeating the optimizations. Four initial sample points chosen by LHS are used to build the initial surrogate models for the objective and constraint functions; then, the suboptimization defined by the infill-sampling criterion of EI is solved by a GA

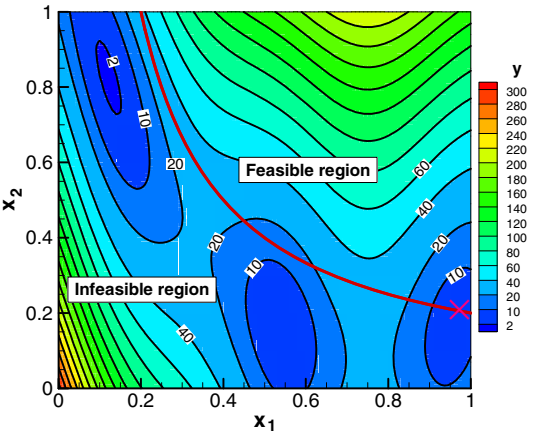
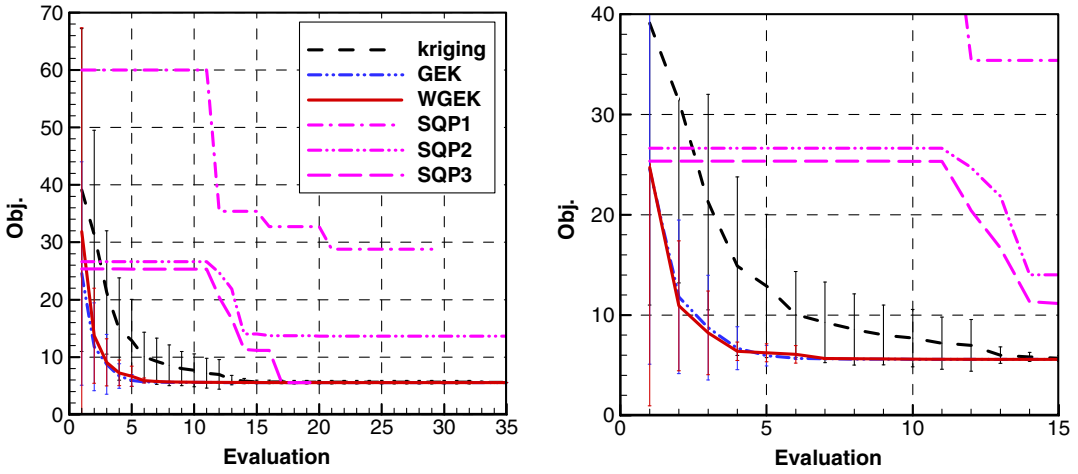


Fig. 6 Design space of a modified Branin–Hoo test function [5] (× denotes the constrained global optimum).

followed by a SQP optimization to suggest a new sample point. As we can see in Fig. 7, the WGEK-based optimization is very close to the GEK-based method, and both apparently outperform the kriging-based optimization. It is also shown that the surrogate-based methods are all better than the direct optimization using SQP, which is sensitive to the starting point. From Table 2, one can see that all the surrogate-based optimizations accurately converge to the global optimum, but the gradient-based optimization can be trapped by a local optimum, depending on the starting point. The example shows that the theory and implementation of the proposed WGEK is correct and matured enough for design optimization.



a) Overall view b) Detailed view

Fig. 7 Convergence histories of minimizing the Branin–Hoo function based on kriging, GEK, and WGEK.

C. Optimization of High-Dimensional Analytical Functions

The Rosenbrock functions with varying dimensions are used to examine the performance of the proposed WGEK for high-dimensional design optimizations. The mathematical model of optimization is

$$\begin{aligned} &\text{min.} \quad f(\boldsymbol{x}) = \sum_{i=1}^{m-1} [(1.0 - x_i)^2 + 100.0(x_{i+1} - x_i^2)^2] \\ &\text{w.r.t.} \quad x_i \in [-5, 5], \qquad i = 1, \ldots, m \end{aligned} \tag{40}$$

where m is the number of design variables. The initial sample points, whose number is equal to $0.5m$, are chosen by LHS. The suboptimization defined by the infill-sampling criterion of MSP is solved by a combination of Hooke and Jeeves pattern search and BFGS optimization. Note that the optimizations were run on a personal computer, a Lenovo ThinkCentre M8200T with Intel Core i7 CPU (4.0 GHz). The optimization of each method is repeated a number of times to study the influence of the randomness in the initial sampling and suboptimization algorithms (when GA is used). Figure 8 shows the convergence histories of the objective function in a statistic manner, with the number of design variables in the range from 30 to 100. Regarding functional evaluations of the optimizations, the GEK and WGEK are comparable, and both are significantly faster than the kriging. However, when the CPU time elapsed for optimization is examined, the GEK gets less efficient and even slower than the kriging, which is due to the fact that training a GEK is too expensive. Especially for the 100-dimensional case, the cost of training the initial GEK is more than one month, which makes it impractical for such a high-dimensional problem. In contrast, when WGEK is used, the optimization efficiency is greatly improved, although the approximation accuracy slightly deteriorates. With the increase of number of design variables, the WGEK becomes more beneficial. The results of directly using the BFGS method are also plotted in Fig. 8, which shows that, for such a particular type of optimization problem, the WGEK-based optimization is slowed down after it quickly figures out the area where the optimum is located because the response surface is located in a very long and flat valley.

D. Aerodynamic Inverse Design of Transonic Wing

To demonstrate the capability of WGEK for high-dimensional engineering design problems, here we are concerned with the inverse design of the ONERA M6 wing parameterized with number of design variables in the range from 36 to 108. Please note that the WGEK, as a generic surrogate model, can be used for constrained drag minimization or other kinds of aerodynamic design problems as well, which are beyond the scope of this paper.

1. *Problem Statement of Wing Inverse Design*

To parameterize the wing, several control sections are used to determine the aerodynamic shape, with the planar shape being fixed. The class function/shape function transformation (CST) method proposed by Kulfan [83] is then used to parameterize each of the control sections. Because an eighth-order CST method is used, there are 18 design variables for each control section. To set up a series of design optimization problems, we adopt two, three, four, and six control sections, which raises up the problems with 36, 54, 72, and 108 design variables, respectively (see Fig. 9).

The objective of the inverse design is to find the aerodynamic shape that best matches the target pressure prescribed by the designer. Here, the target pressure distribution is obtained by simulating the flow over the ONERA M6 wing at a freestream condition of $Ma = 0.8395$, $Re = 1.17 \times 10^7$, $\alpha = 3.06$ deg. If the target pressure is exactly matched, the optimal wing shape should exactly match the shape of the ONERA M6 wing, which enables us to validate the optimization method. The mathematical model is

$$\begin{aligned} &\text{min.} \quad f(\boldsymbol{x}) = \int_B [(p - p_{\text{target}})^2] \, \mathrm{d}s \\ &\text{w.r.t.} \quad \boldsymbol{x}_l \le \boldsymbol{x} \le \boldsymbol{x}_u, \qquad \boldsymbol{x} \in \mathbb{R}^m \end{aligned} \tag{41}$$

where p and p_{target} are normalized surface pressure and target pressure evaluated by CFD analysis, respectively. The design space is determined by expanding and shrinking the CST coefficients of each section of the target shape by 20%, respectively. The initial sample shapes whose number is equal to that of the design variables are chosen in the design space by LHS method. An in-house Reynolds-averaged Navier–Stokes (RANS) solver and the corresponding

Table 2 Comparison of optima obtained by using kriging, GEK, WGEK and gradient-based methods

Optimization methods	Best optimal design variables $\boldsymbol{x}$	Best objective function $f(\boldsymbol{x})$	Constraint function $g(\boldsymbol{x})$
Kriging_Opt	$(x_1^*, x_2^*) = (0.9675893, 0.2066993)$	5.575664087	−0.35582093E − 11
GEK_Opt	$(x_1^*, x_2^*) = (0.9675858, 0.2067000)$	5.575664094	−0.14570759E − 08
WGEK_Opt	$(x_1^*, x_2^*) = (0.9675857, 0.2067000)$	5.575664111	−0.24572787E − 08
SQP_Opt1$(x_1, x_2)_{\text{ini}} = (0.3, 0.3)$	$(x_1^*, x_2^*) = (0.9675856, 0.2067001)$	5.575663800	0.4635181E − 08
SQP_Opt2$(x_1, x_2)_{\text{ini}} = (0.5, 0.5)$	$(x_1^*, x_2^*) = (0.5354652, 0.3735070)$	13.63412800	−0.1289802E − 11
SQP_Opt3$(x_1, x_2)_{\text{ini}} = (0.6, 0.6)$	$(x_1^*, x_2^*) = (0.2521341, 0.7932288)$	28.77858700	0.9796886E − 12

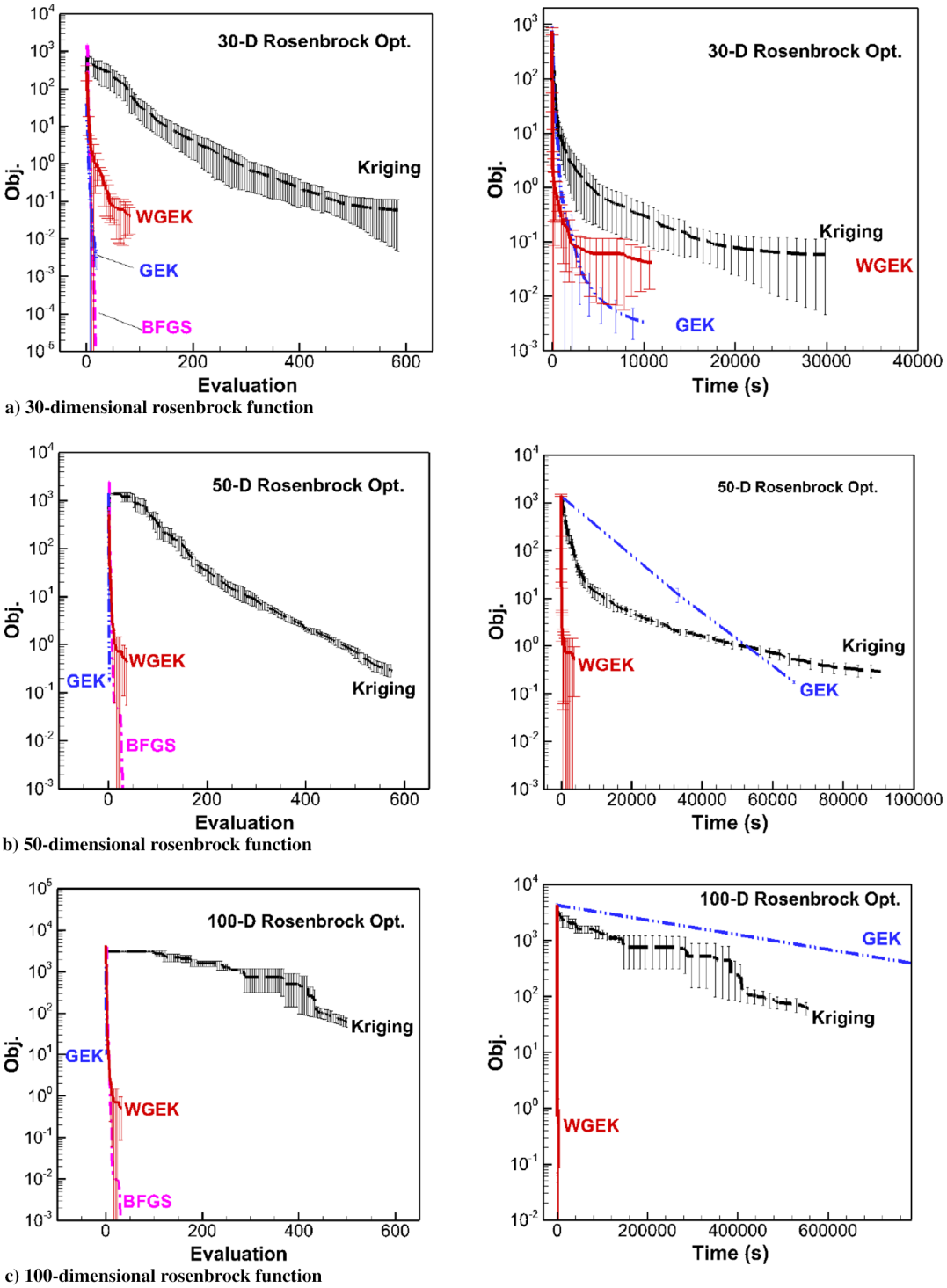
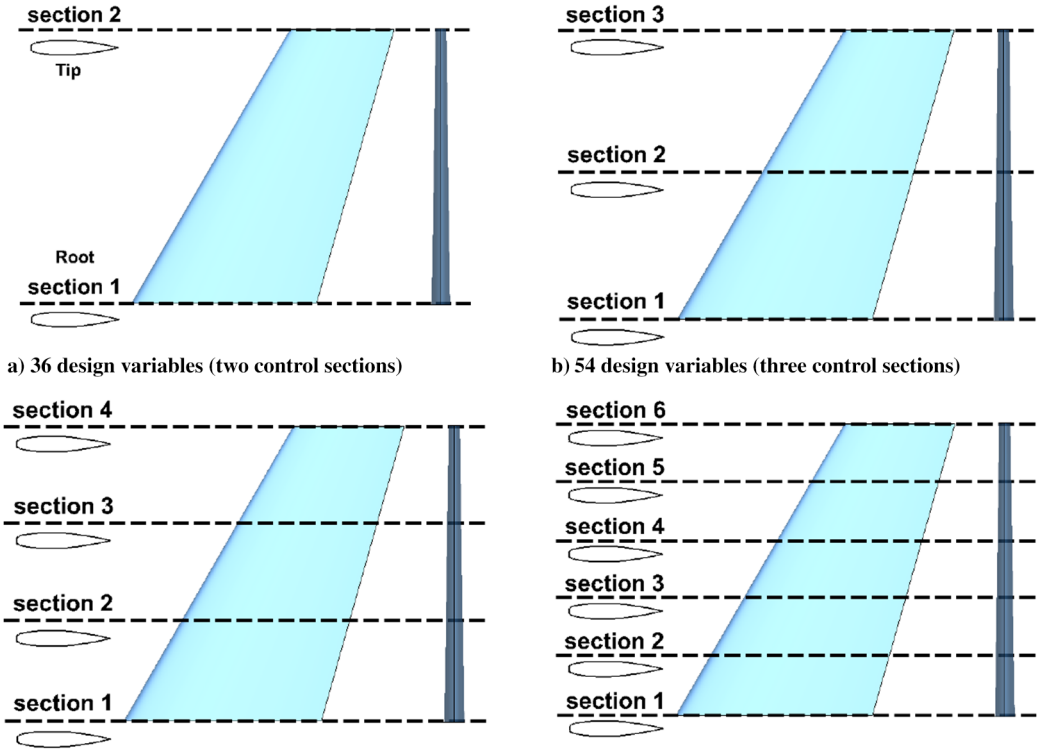


Fig. 8 Convergence histories of objective function versus number of function evaluations and CPU time for minimizing Rosenbrock functions with number of design variables in the range from 30 to 100.

adjoint solver are used to evaluate the sample shapes and get the response of the objective function as well as its gradients, respectively. Then, an initial surrogate model is built, and the suboptimization defined by the infill-sampling criterion of MSP is solved by the combination of GA, Hooke and Jeeves pattern search, and BFGS method. The surrogate model is repetitively updated until the optimal shape is reached. The success of the inverse design is validated by comparing the optimal wing with the target wing (i.e., the ONERA M6 wing).

The flow analyses are performed with an in-house code called PMNS3D [84], which solves the Reynolds-averaged Navier–Stokes (RANS) equations to simulate the flows around a 3-D configuration. The governing equations are solved on structured grids by using a cell-centered finite volume method. The second-order Jameson central scheme is used for spatial discretization, and the Baldwin–Lomax turbulence model is used for turbulence closure. Implicit residual smoothing, local time stepping, and multigrid techniques are used to accelerate the solution to converge to the steady state. Figure 10 shows the C-H type grids used in this paper, which are generated automatically by our in-house code, with the grid distribution of 160 (chordwisedirection) × 40 (normaldirection) × 60 (spanwise direction). Figure 11 shows the representative convergence histories of the averaged residual of the continuous equation and force coefficients. Figure 12 shows the comparison of



a) 36 design variables (two control sections)

b) 54 design variables (three control sections)

c) 72 design variables (four control sections)

d) 108 design variables (six control sections)

Fig. 9 Parameterization of ONERA M6 wing with number of design variables in the range from 36 to 108.

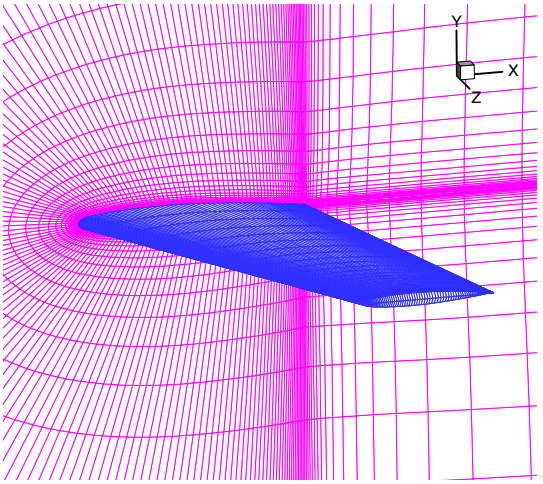


Fig. 10 Sketch of C-H type computational grid for ONERA M6 wing.

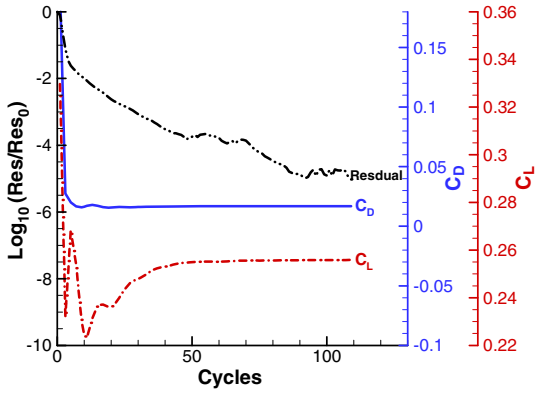


Fig. 11 Representative convergence history of CFD simulation for ONERA M6 wing.

computed pressure distributions with the wind-tunnel experimental data at a freestream condition of $Ma = 0.8395$, $Re = 1.17 \times 10^7$, $\alpha = 3.06$ deg. As one can see, the CFD results are in reasonably good agreement with the experimental data.

Furthermore, the gradients of the objective function with respect to the design variables are obtained by a continuous adjoint method, which was implemented in the PMNS3D code. The adjoint solver is validated by the comparison of computed gradients with that by the finite difference method (central difference), as shown in Fig. 13.

2. *Results of Optimizations with Varying Number of Design Variables*

The inverse design problems are solved based on kriging, GEK, and WGEK surrogate models. The convergence histories of the objective function, versus the number of evaluations as well as the elapsed CPU time, are examined to compare the performance of different types of surrogate models. It is worth noting that the number of iterations is actually the same as the number of evaluations excluding initial sampling because, in this study, only one sample is added in each updating cycle. For counting the CPU time, it starts from the beginning of first modeling, after the initial sample points have been evaluated.

Figures 14 and 15 show the convergence histories for the problems with 36 and 54 design variables, respectively. In Fig. 14a, it is shown that, for the 36-dimensional problem, the GEK performs the best; although the proposed WGEK is slower than the GEK, it is still much faster than the kriging. The same conclusion applies to the convergence of the objective function versus the CPU time, as shown in Fig. 14b. However, when the number of design variables is increased up to 54, the performance of the GEK, regarding the elapsed CPU time, becomes comparable to that of the WGEK. The reason is explained by the fact that, with increase of number of design variables, the correlation matrix of a GEK becomes larger and larger, and the model training becomes more and more expensive.

Figures 16 and 17 illustrate the convergence histories for the problems with 72 and 108 design variables, respectively. For such high-dimensional design problems, the correlation matrix of a GEK is huge, and the cost associated with training the GEK becomes dominated or even prohibitive. For the 72-dimensional problem, the GEK-based method costs more than one month but only decreases

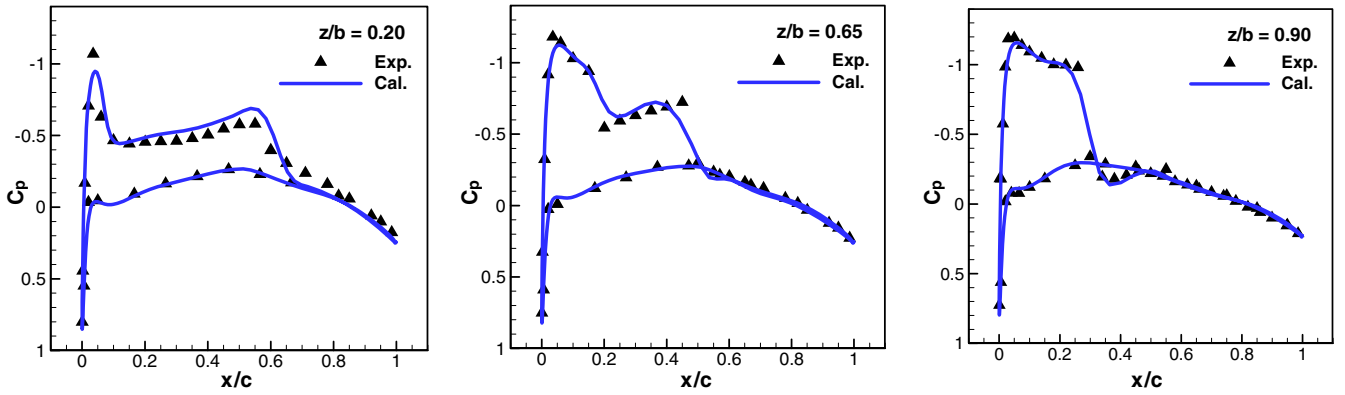


Fig. 12 Validation of flow solver by comparison of the calculated pressure distribution and experimental data at three spanwise locations of ONERA M6 wing ($Ma = 0.8395, \alpha = 3.06$ deg, $Re = 1.17 \times 10^7$).

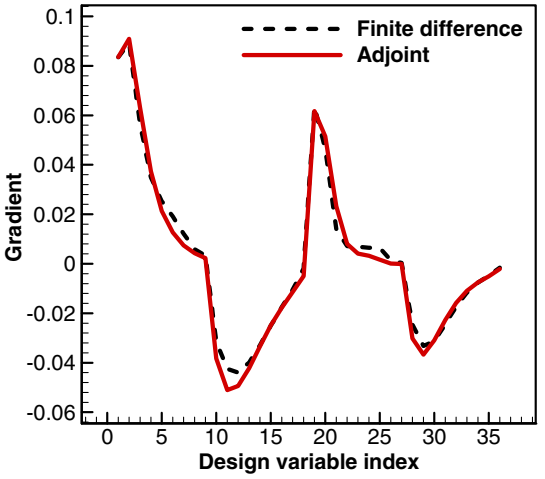


Fig. 13 Validation of adjoint solver by comparison of computed gradients and that obtained by finite difference method.

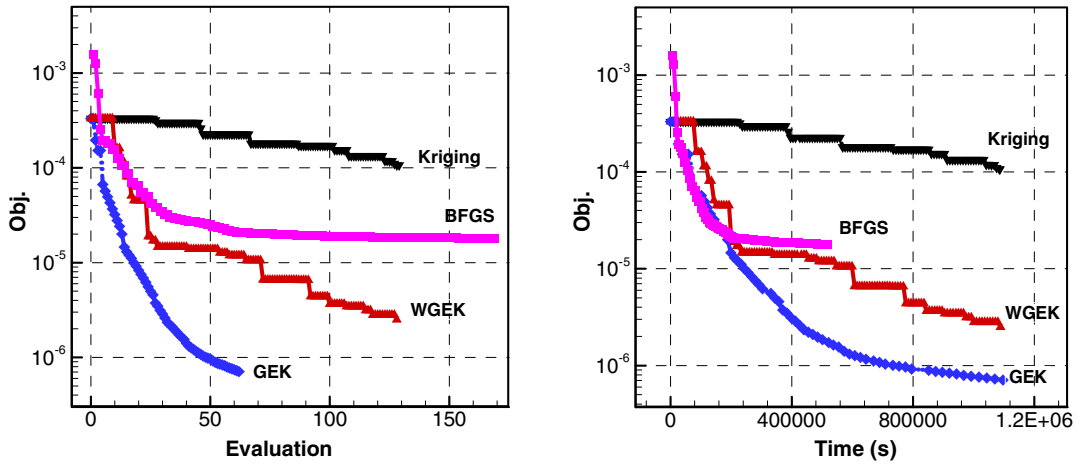
the objective function around one order in magnitude, whereas the WGEK-based method, with the same level of CPU time, decreases the objective function by two orders in magnitude. For the 108-dimensional problem, the GEK-based method becomes impractical, but the WGEK-based method is still working.

The convergence histories of a gradient-based method, BFGS, are also shown in these figures. It is observed that, for the problems with number of design variables in the range from 36 to 72, the WGEK-based optimization is comparable to or slightly better than the BFGS method. However, when the number of design variables reaches 108 (or beyond), the WGEK-based optimization is less efficient than the BFGS method, when the elapsed CPU time for optimization is considered. This implies that the model training of the WGEK model still costs much time for the problems with more than 100 design variables, which is an open issue to be addressed in the future.

To figure out the reason why the efficiency of the GEK-based optimization deteriorates with increase of number of design variables, a study is carried out for the breakdown of the total CPU time, which mainly consists of the time spent on CFD and the time for training surrogate model. The CPU time for each flow solution and gradient evaluation is almost unchanged (around 3000 to 4000 s), whereas the cost for training the surrogate models increases dramatically with increase of number of design variables. Because the GEK and kriging are not successful for the 72- and 108-dimensional problems, it would make no sense to make a breakdown for the total CPU time. To enable a reasonable comparison, we made a breakdown for the first iteration; see Fig. 18. It is confirmed that the rapidly increased cost associated with training the GEK leads to the deterioration of its optimization efficiency. The problem of the GEK is successfully avoided by using the proposed WGEK, and the cost of model training is kept in an acceptable level, which makes it practical for the problems with number of design variables as large as 100.

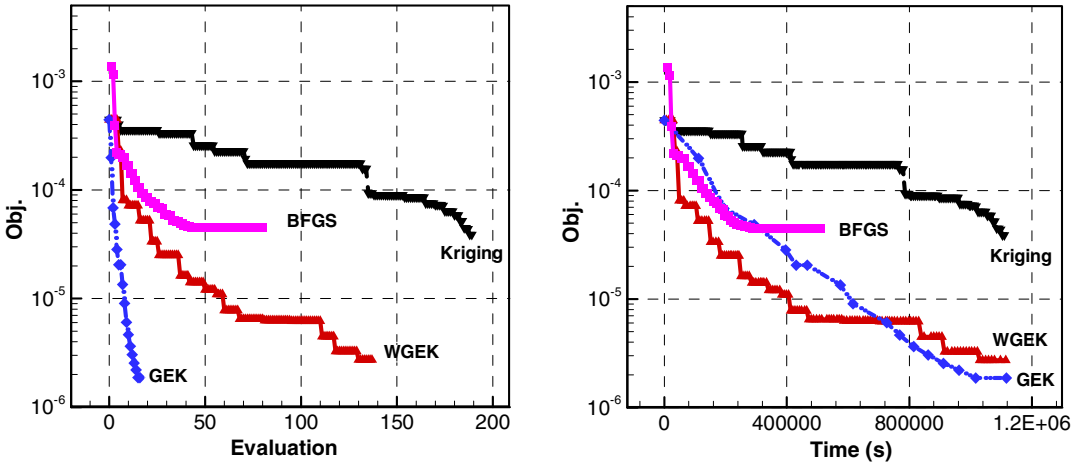
3. *Summarizing the Performance of Different Methods*

Three indicators are defined to evaluate the performance of optimizations based on different surrogate models:



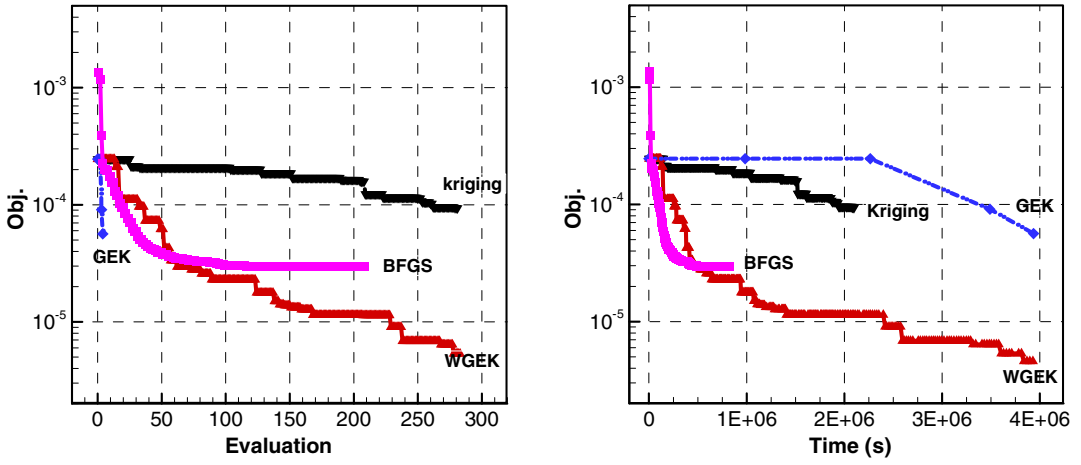
a) Objective function versus number of CFD evaluations **b) Objective function versus elapsed CPU time**

Fig. 14 Convergence histories of inverse design for a transonic wing with 36 design variables ($Ma = 0.8395, \alpha = 3.06$ deg, $Re = 1.17 \times 10^7$).



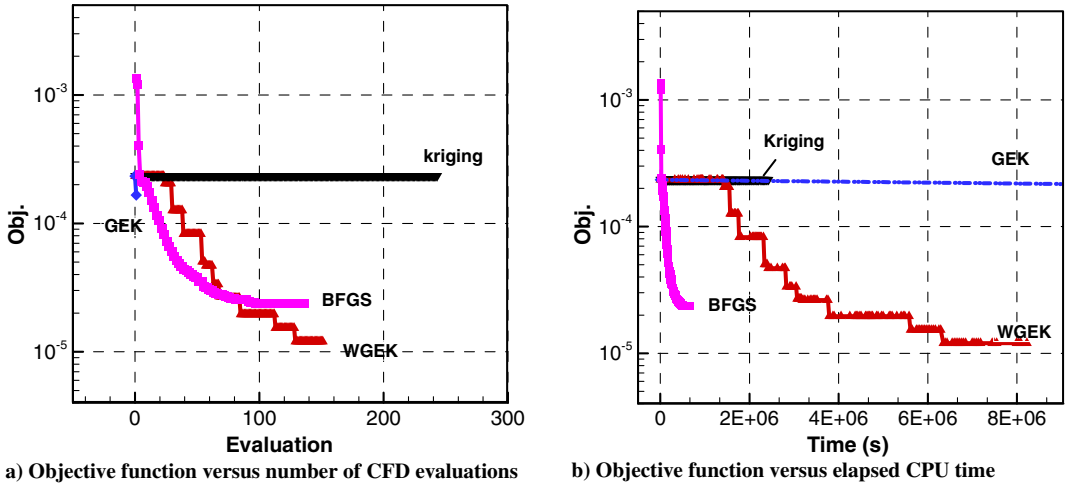
a) Objective function versus number of CFD evaluations b) Objective function versus elapsed CPU time

Fig. 15 Convergence histories of inverse design of a transonic wing with 54 design variables ($Ma = 0.8395, \alpha = 3.06$ deg, $Re = 1.17 \times 10^7$).



a) Objective function versus number of CFD evaluations b) Objective function versus elapsed CPU time

Fig. 16 Convergence histories of inverse design for a transonic wing with 72 design variables ($Ma = 0.8395$, $Re = 1.17 \times 10^7$, $\alpha = 3.06$ deg).

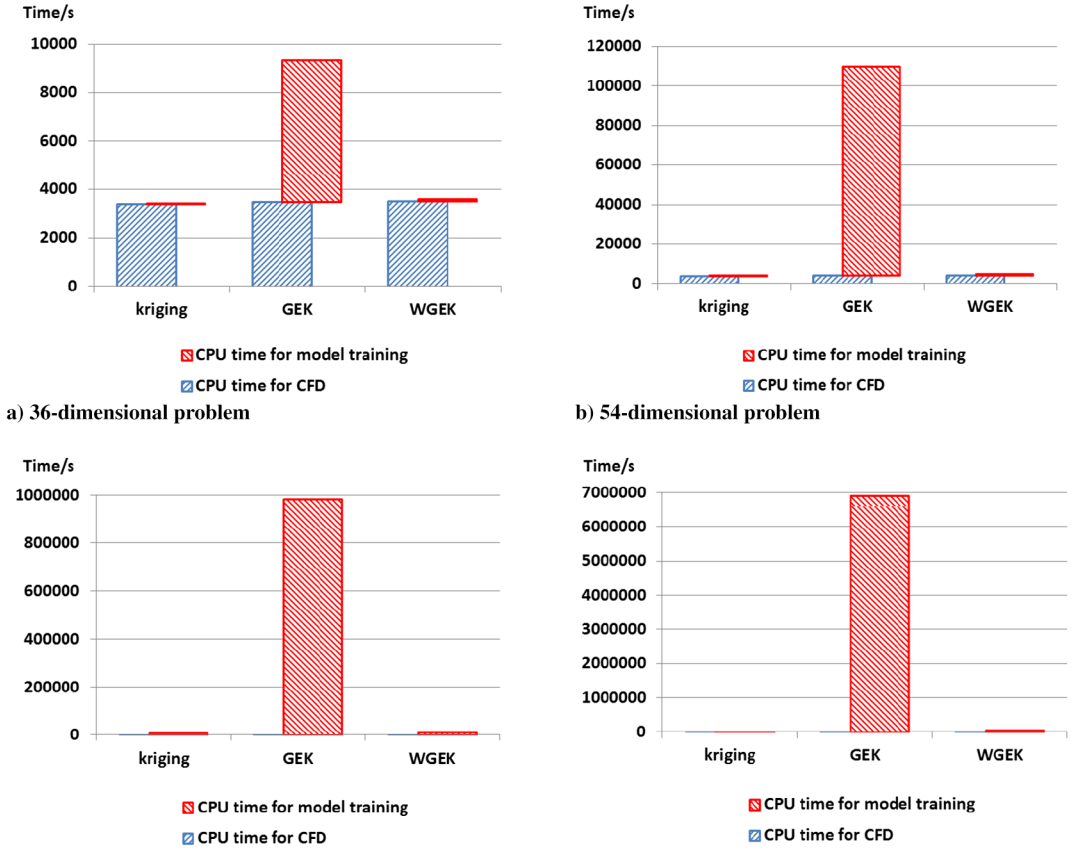


a) Objective function versus number of CFD evaluations b) Objective function versus elapsed CPU time

Fig. 17 Convergence histories of inverse design for a transonic wing with 108 design variables ($Ma = 0.8395$, $Re = 1.17 \times 10^7$, $\alpha = 3.06$ deg).

$$I_1 = \log\left(\frac{\log(\mathrm{Obj}_{\mathrm{init}}) - \log(\mathrm{Obj}_{\mathrm{opt}})}{N_{\mathrm{evaluation}}}\right)$$
$$I_2 = \log\left(\frac{\log(\mathrm{Obj}_{\mathrm{init}}) - \log(\mathrm{Obj}_{\mathrm{opt}})}{t_{\mathrm{total}}}\right)$$
$$I_3 = \log\left(\frac{t_{\mathrm{CFD}}}{t_{\mathrm{total}}}\right) \tag{42}$$

The three indicators are normalized and plotted for the inverse design problems of varying number of design variables, as the radar diagrams shown in Fig. 19. The upper and bottom-left indicators show the convergence speed regarding the number of evaluations and the total CPU time, denoted by the normalized I_1 and I_2, respectively. Large values of both indicators are generally preferred. The right-bottom indicator, denoted by the normalized I_3, stands for the ratio of



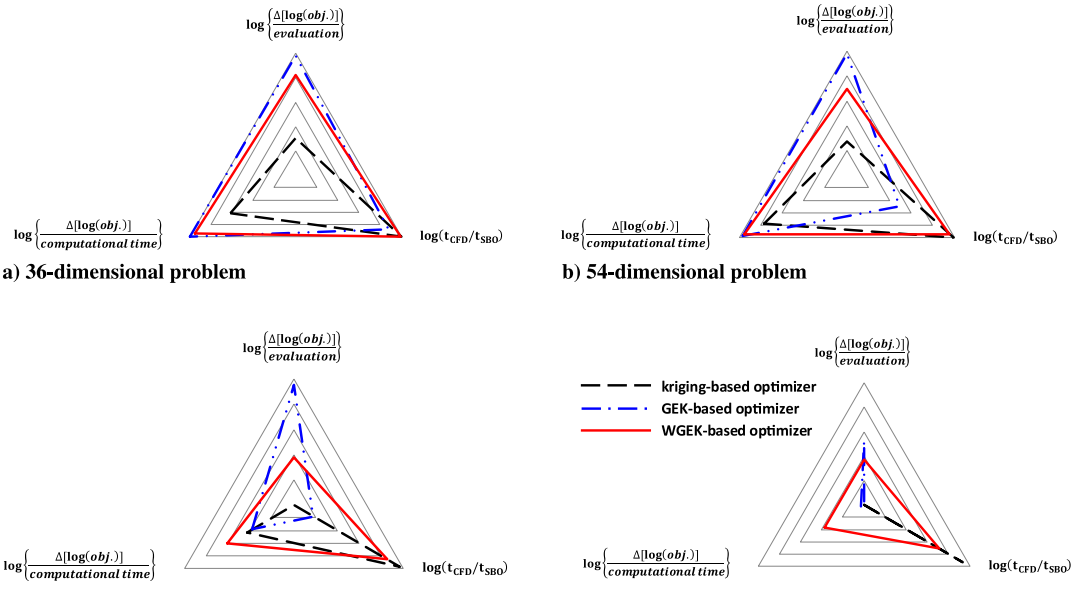
a) 36-dimensional problem

b) 54-dimensional problem

c) 72-dimensional problem

d) 108-dimensional problem

Fig. 18 Breakdown of the CPU time for inverse design of a transonic wing ($Ma = 0.8395$, $Re = 1.17 \times 10^7$, $\alpha = 3.06$ deg).



a) 36-dimensional problem

b) 54-dimensional problem

c) 72-dimensional problem

d) 108-dimensional problem

Fig. 19 Summary of performance of kriging-, GEK-, and WGEK-based optimizations for high-dimensional aerodynamic design.

the CPU time spent on CFD simulations to the total CPU time. Because we prefer to spend more computational time on CFD, the larger value of the right-bottom indicator is also preferred.

For the 36-dimensional problem, as shown in Fig. 19a, the GEK-based method performs better than the WGEK-based method for the upper and bottom-left indicators and slightly worse for the bottom-right indicator. When the number of design variables is increased to 54, the GEK is still working well, but we do see the potential crisis associated with the rapidly increased cost spent on model training. For the 72- and 108-dimensional problems, the convergence speed regarding the computational time of the GEK is apparently worse than that of the WGEK. It can be concluded that

1) in view of the overall performance, the proposed WGEK is the most efficient, especially for high-dimensional problems ($m \geq 72$); 2) the GEK is the most accurate but is only practical for the problem with medium size of design variables ($m < 72$); and 3) the kriging without using any gradient is only practical for relatively low-dimensional problems ($m < 54$).

V. Conclusions

In this paper, a novel formulation of gradient-enhanced kriging, called weighted gradient-enhanced kriging (WGEK), was proposed to address the problem associated with the large cost of training a gradient-enhanced kriging (GEK) for high-dimensional problems. The core idea is to build a series of submodels with smaller correlation matrices and then to sum them up with appropriate weight coefficients. In contrast to the conventional GEK, the cost of training a WGEK increases slowly with increasing number of dimensions and is kept in an acceptable level. The efficiency improvement of a WGEK against a GEK was analyzed, which shows that the analytical speed-up ratio is $n^2(1+m)^3/(n+m)^3$, where m and n are the numbers of design variables and sampling sites, respectively. The theory of the proposed WGEK was verified and validated by analytical test cases. It was noted that, compared to a GEK, the accuracy of WGEK is somehow decreased, as a price paid for the large efficiency improvement.

The WGEK was integrated into a surrogate-based optimization code called SurroOpt and compared with both the conventional GEK and kriging for benchmark optimizations of analytical functions, with the number of design variables in the range from 30 to 100. As the number of design variables is increased up to 50 and 100, it was observed that the WGEK significantly outperforms the GEK.

At last, the proposed WGEK was applied to the inverse designs of an ONERA M6 wing parameterized with number of design variables in the range from 36 and 108. It was observed that, for medium size of design variables (up to 54), the overall design efficiency of using GEK and WGEK is comparable, and both are much more efficient than that of kriging without using any gradient. When the number of design variables is increased up to 72 and 108, the cost of training a GEK increases rapidly and soon becomes prohibitive. In contrast, the WGEK performs well, and the cost for model training is kept in an acceptable level, which shows that the proposed WGEK has great potential for applications to higher-dimensional engineering design problems.

Besides aerodynamic applications, the proposed WGEK could be suited for applications in any other research areas, where a large number of design variables are involved and cheap gradients are available. The future work will focus on the further extension of WGEK to model in which each submodel can incorporate the gradients at arbitrary number of sample sites and application of WGEK to efficient global optimization of complex aircraft configuration.

Acknowledgments

This work was supported by the national Natural Science Foundation of China under grant number 11272265 and the Aeronautical Science Foundation of China under grant 2016ZA53011. The authors would like to thank Wenping Song, Jun Liu, and Shaoqiang Han in our group, for the useful discussion and technical support. The authors also would like to thank Nicolas Gauger, Richard Dwight, Ralf Zimmermann, and Stefan Göertz for their valuable suggestions and useful discussions.

References

[1] Simpson, T. W., Peplinski, J., Koch, P. N., and Allen, J. K., "Metamodels for Computer-Based Engineering Design: Survey and Recommendations," *Engineering with Computers*, Vol. 17, No. 2, 2001, pp. 129–150. doi:10.1007/PL00007198

[2] Simpson, T. W., Booker, A. J., Ghosh, D., Giunta, A. A., Koch, P. N., and Yang, R.-J., "Approximation Methods in Multidisciplinary Analysis and Optimization: A Panel Discussion," *Structural and Multidisciplinary Optimization*, Vol. 27, No. 5, 2004, pp. 302–313. doi:10.1007/s00158-004-0389-9

[3] Queipo, N. V., Haftka, R. T., Shyy, W., Goela, T., Vaidyanathan, R., and Tucker, P. K., "Surrogate-Based Analysis and Optimization," *Progress in Aerospace Sciences*, Vol. 41, No. 1, 2005, pp. 1–28.

[4] Wand, G. G., and Shan, S., "Review of Metamodeling Techniques on Support of Engineering Design Optimization," *Journal of Mechanical Design*, Vol. 129, No. 4, 2007, pp. 370–380. doi:10.1115/1.2429697

[5] Forrester, A. I. J., and Keane, A. J., "Recent Advances in Surrogate-Based Optimization," *Progress in Aerospace Sciences*, Vol. 45, Nos. 1–3, 2009, pp. 50–79.

[6] Han, Z.-H., and Zhang, K.-S., "Surrogate-Based Optimization," *Real-World Applications of Genetic Algorithms*, edited by O. Roeva, InTech, Rijeka, Croatia, 2012, pp. 343–362, http://www.intechopen.com/books/real-world-applications-of-genetic-algorithms/surrogate-based-optimization [retrieved 07 March 2012].

[7] Koziel, S., Leifsson, L., and Yang, X. S., "Surrogate-Based Optimization," *Simulation-Driven Design Optimization and Modeling for Microwave Engineering*, edited by S. Koziel, X. S. Yang, and Q. J. Zhang, Imperial College Press, London, 2013, pp. 53–86.

[8] Viana, F. A. C., Simpson, T. W., Balabanov, V., and Toropov, V., "Metamodeling in Multidisciplinary Design Optimization: How Far Have We Really Come?" *AIAA Journal*, Vol. 52, No.4, 2014, pp. 670–690.

[9] Ahmed, M. Y. M., and Qin, N., "Surrogate-Based Multi-Objective Aerothermodynamic Design Optimization of Hypersonic Spiked Bodies," *AIAA Journal*, Vol. 50, No. 4, April 2015, pp. 797–810.

[10] Robinson, T. D., Eldred, M. S., Willcox, K. E., and Haimes, R., "Surrogate-Based Optimization Using Multifidelity Models with Variable Parameterization and Corrected Space Mapping," *AIAA Journal*, Vol. 46, No.11, 2008, pp. 2814–2822.

[11] Barthelemy, J.-F. M., and Haftka, R. T., "Approximation Concepts for Optimum Structural Design—A Review," *Structural Optimization*, Vol. 5, No. 3, 1993, pp. 129–144. doi:10.1007/BF01743349

[12] Sobieszczanski-Sobieski, J., and Haftka, R. T., "Multidisciplinary Aerospace Design Optimization: Survey of Recent Developments," *Structural Optimization*, Vol. 14, No. 1, 1997, pp. 1–23. doi:10.1007/BF01197554

[13] Schmit, L. A., and Farshi, B., "Some Approximation Concepts for Structural Synthesis," *AIAA Journal*, Vol. 12, No. 5, 1974, pp. 692–699.

[14] Box, G. E. P., and Drapper, N. R., "Empirical Model Building and Response Surfaces," *Journal of the Royal Statistical Society*, Vol. 30, No. 2, 1987, pp. 229–231.

[15] Krige, D. G., "A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand," *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, Vol. 94, No. 3, 1951, pp. 95–111.

[16] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., "Design and Analysis of Computer Experiments," *Statistical Science*, Vol. 4, No. 4, 1989, pp. 409–423.

[17] Powell, M. J. D., *Algorithms for Approximation*, Oxford Univ. Press, New York, 1987, pp. 141–167.

[18] Mullur, A. A., and Messac, A., "Extended Radial Basis Functions: More Flexible and Effective Metamodeling," *AIAA Journal*, Vol. 43, No. 6, 2005, pp. 1306–1315.

[19] Park, J., and Sandberg, I. W., "Universal Approximation Using Radial-Basis-Function Networks," *Neural Computation*, Vol. 3, No. 2, 1991, pp. 246–257.

[20] Elanayar, S. V. T., and Shin, Y. C., "Radial Basis Function Neural Network for Approximation and Estimation of Nonlinear Stochastic Dynamic Systems," *IEEE Transactions on Neural Networks*, Vol. 5, No. 4, 1994, pp. 594–603.

[21] Smola, A. J., and Schölkopf, B. A., "Tutorial on Support Vector Regression," *Statistics and Computing*, Vol. 14, No. 3, 2004, pp. 199–222.

[22] Zhang, K.-S., and Han, Z.-H., "Support Vector Regression-Based Multidisciplinary Design Optimization in Aircraft Conceptual Design," *51st AIAA Aerospace Sciences Meeting*, AIAA Paper 2013-1160, Jan. 2013.

[23] Wang, Q., Moin, P., and Iaccarino, G., "A High-Order Multi-Variate Approximation Scheme for Arbitrary Data Sets," *Journal of Computational Physics*, Vol. 229, No. 18, 2010, pp. 6343–6361.

[24] Wiener, N., "The Homogeneous Chaos," *American Journal of Mathematics*, Vol. 60, No. 4, 1938, pp. 897–936.

[25] Xiu, D., *Numerical Methods for Stochastic Computations: A Spectral Method Approach*, Princeton Univ. Press, Princeton, NJ, 2010, p. 152.

[26] Jeong, S., Murayama, M., and Yamamoto, K., "Efficient Optimization Design Method Using Kriging Model," *Journal of Aircraft*, Vol. 42, No. 2, 2005, pp. 413–420.

[27] Vavalle, A., and Qin, N., "Iterative Response Surface Based Optimization Scheme for Transonic Airfoil Design," *Journal of Aircraft*, Vol. 44, No. 2, 2007, pp. 365–376.

[28] Kanazaki, M., Tanaka, K., Jeong, S., and Yamamoto, K., "Multi-Objective Aerodynamic Exploration of Elements' Setting for High-Lift Airfoil Using Kriging Model," *Journal of Aircraft*, Vol. 44, No. 3, 2007, pp. 858–864.

[29] Han, Z.-H., Liu, J., Song, W.-P., and Zhang, K.-S., "Surrogate-Based Aerodynamic Shape Optimization with Application to Wind Turbine Airfoils," *51st AIAA Aerospace Sciences Meeting*, AIAA Paper 2013-1108, Jan. 2013.

[30] Liu, J., Song, W.-P., Han, Z.-H., and Zhang, Y., "Efficient Aerodynamic Shape Optimization of Transonic Wings Using a Parallel Infilling Strategy and Surrogate Models," *Structural and Multidisciplinary Optimization*, Vol. 55, No. 3, 2016, pp. 925–943. doi:10.1007/s00158-016-1546-7

[31] Shan, S., and Wang, G. G., "Survey of Modeling and Optimization Strategies to Solve High-Dimensional Design Problems with Computationally-Expensive Black-Box Functions," *Structural and Multidisciplinary Optimization*, Vol. 41, No. 2, March 2010, pp. 219–241.

[32] Koch, P. N., Simpson, T. W., Allen, J. K., and Mistree, F., "Statistical Approximations for Multidisciplinary Design Optimization: The Problem of the Size," *Journal of Aircraft*, Vol. 36, No. 1, 1999, pp. 275–286.

[33] Han, Z.-H., Zimmermann, R., and Goertz, S., "An Alternative Cokriging Model for Variable-Fidelity Surrogate Modeling," *AIAA Journal*, Vol. 50, No. 5, 2012, pp. 1205–1210.

[34] Han, Z.-H., and Goertz, S., "Hierarchical Kriging Model for Variable-Fidelity Surrogate Modeling," *AIAA Journal*, Vol. 50, No. 9, 2012, pp. 1285–1296.

[35] Choi, S., Alonso, J. J., Kim, S., Kroo, I., and Wintzer, M., "Two-Level Multi-Fidelity Design Optimization Studies for Supersonic Jets," *43rd Aerospace Sciences Meeting and Exhibit*, AIAA Paper 2005-0531, Jan. 2005.

[36] Gano, S. E., Renaud, J. E., Martin, J. D., and Simpson, T. W., "Update Strategies for Kriging Models for Using in Variable Fidelity Optimization," *46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, AIAA Paper 2005-2057, April 2005.

[37] Leifsson, L., Koziel, S., and Tesfahunegn, Y. A., "Multiobjective Aerodynamic Optimization by Variable-Fidelity Models and Response Surface Surrogates," *AIAA Journal*, Vol. 54, No. 2, 2016, pp. 531–541.

[38] Leifsson, L., and Koziel, S., "Aerodynamic Shape Optimization by Variable-Fidelity Computational Fluid Dynamics Models: A Review of Recent Progress," *Journal of Computational Science*, Vol. 10, Sept. 2015, pp. 45–54.

[39] March, A., and Willcox, K., "Convergent Multifidelity Optimization Using Bayesian Model Calibration," *13th AIAA/ISSMO Multidisciplinary Analysis Optimization Conference*, AIAA Paper 2010-9198, Sept. 2010.

[40] Jameson, A., "Aerodynamic Design via Control Theory," *Journal of Scientific Computing*, Vol. 3, No. 3, Sept. 1988, pp. 233–260. doi:10.1007/BF01061285

[41] Anderson, W. K., and Venkatakrishnan, V., "Aerodynamic Design Optimization on Unstructured Grids with a Continuous Adjoint Formulation," *Computers and Fluids*, Vol. 28, No. 4, 1999, pp. 443–480. doi:10.1016/S0045-7930(98)00041-3

[42] Neidinger, R. D., "Introduction to Automatic Differentiation and MATLAB Object-Oriented Programming," *SIAM Review*, Vol. 52, No. 3, 2010, pp. 545–563.

[43] Morris, M. D., Mitchell, T. J., and Ylvisaker, D., "Bayesian Design and Analysis of Computer Experiments: Use of Derivatives in Surface Prediction," *Technometrics*, Vol. 35, No. 3, 1993, pp.243–255.

[44] Chung, H. S., and Alonso, J. J., "Using Gradients to Construct Cokriging Approximation Models for High-Dimensional Design Optimization Problems," *40th AIAA Aerospace Sciences Meeting & Exhibit*, AIAA Paper 2002-0317, 2002.

[45] Laurenceau, J., and Sagaut, P., "Building Efficient Response Surfaces of Aerodynamic Functions with Kriging and Cokriging," *AIAA Journal*, Vol. 46, No. 2, 2008, pp. 498–507.

[46] Dalbey, K. R., "Efficient and Robust Gradient Enhanced Kriging Emulators," Sandia National Lab. Rept. SAND2013-7022, Albuquerque, NM, Aug. 2013.

[47] Han, Z.-H., Goertz, S., and Zimmermann, R., "Improving Variable-Fidelity Surrogate Modeling via Gradient-Enhanced Kriging and a Generalized Hybrid Bridge Function," *Aerospace Science and Technology*, Vol. 25, No. 1, 2013, pp. 177–189.

[48] Liu, W., and Batill, S. M., "Gradient-Enhanced Response Surface Approximations Using Kriging Models," *9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, AIAA paper 2002-5456, Sept. 2002.

[49] Liu, W., "Development of Gradient-Enhanced Kriging Approximations for Multidisciplinary Design Optimization," Ph.D. Thesis, University of Notre Dame, Notre Dame, IN, 2003.

[50] Ulaganathan, S., Couckuyt, I., Ferranti, F., Laermans, E., and Dhaene, T., "Performance Study of Multi-Fidelity Gradient Enhanced Kriging," *Structural and Multidisciplinary Optimization*, Vol. 51, No. 5, 2015, pp. 1017–1033. doi:10.1007/s00158-014-1192-x

[51] Laurenceau, J., Meaux, M., Montagnac, M., and Sagaut, P., "Comparison of Gradient-Based and Gradient-Enhanced Response-Surface-Based Optimizers," *AIAA Journal*, Vol. 48, No. 5, 2012, pp. 981–994.

[52] Yamazaki, W., and Mavriplis, D. J., "Derivative-Enhanced Variable Fidelity Surrogate Modeling for Aerodynamic Functions," *AIAA Journal*, Vol. 51, No. 1, 2013, pp. 126–137.

[53] Jo, Y., Choi, S., and Lee, D., "Variable-Fidelity Design Method Using Gradient-Enhanced Kriging Surrogate Model with Regression," AIAA Paper 2014-2867, 2014.

[54] Dwight, R. P., and Han, Z.-H., "Efficient Uncertainty Quantification Using Gradient Enhanced Kriging," *50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, AIAA Paper 2009-2276, May 2009.

[55] Yamazaki, W., Rumpfkeil, M. P., and Mavriplis, D. J., "Design Optimization Utilizing Gradient/Hessian Enhanced Surrogate Model," *28th AIAA Applied Aerodynamics Conference*, AIAA Paper 2010-4363, June–July 2010.

[56] Reuther, J., Jameson, A., Farmer, J., Martinelli, L., and Saunders, D., "Aerodynamic Shape Optimization of Complex Aircraft Configurations via an Adjoint Formulation," *34th Aerospace Sciences Meeting and Exhibit*, AIAA Paper 1996-0094, Jan. 1996.

[57] Brezillon, J., and Gauger, N. R., "2D and 3D Aerodynamic Shape Optimization Using the Adjoint Approach," *Aerospace Science and Technology*, Vol. 8, No. 8, 2004, pp. 715–727.

[58] Lyu, Z., Kenway, G. K. W., and Martins, J. R. R. A., "Aerodynamic Shape Optimization Studies on the Common Research Model Wing Benchmark," *AIAA Journal*, 2015, Vol. 53, No. 4, pp. 968–985.

[59] Koo, D., and Zingg, D. W., "Progress in Aerodynamic Shape Optimization Based on the Reynolds-Averaged Navier–Stokes Equations," *54th AIAA Aerospace Sciences Meeting*, AIAA Paper 2016-1292, 2016.

[60] Chernukhin, O., and Zingg, D. W., "Multimodality and Global Optimization in Aerodynamic Design," *AIAA Journal*, Vol. 51, No. 6, 2013, pp. 1342–1354.

[61] Antunes, A. P., and Azevedo, J. L. F., "Studies in Aerodynamic Optimization Based on Genetic Algorithms," *AIAA Journal*, Vol. 51, No. 3, May–June 2014, pp. 1002–1012.

[62] Poole, D. J., Allen, C. B., and Rendall, T. C. S., "Comparison of Local and Global Constrained Aerodynamic Shape Optimization," *32nd AIAA Applied Aerodynamics Conference*, AIAA Paper 2014-3223, June 2014.

[63] Han, Z.-H., "Improving Adjoint-Based Aerodynamic Optimization via Gradient-Enhanced Kriging," *50th AIAA Aerospace Sciences Meeting*, AIAA Paper 2012-0670, Jan. 2012.

[64] Giunta, A. A., Wojtkiewicz, S. F., and Eldred, M. S., "Overview of Modern Design of Experiments Methods for Computational Simulations," *41st Aerospace Sciences Meeting and Exhibit*, AIAA Paper 2003-0649, Jan. 2003.

[65] Fang, K. T., Lin, D. K. J., Winker, P., and Zhang, Y., "Uniform Design: Theory and Application," *Technometrics*, Vol. 42, No. 3, 2000, pp. 237–248.

[66] Tang, B., Heywood, M. I., and Shepherd, M., "Input Partitioning to Mixture of Experts," *Proceedings of the 2002 International Joint Conference on Neural Networks*, Vol. 1, IEEE Publ., Piscataway, NJ, 2002, pp. 227–232.

[67] Harris, P., Charlton, M., and Fotheringham, A. S., "Moving Window Kriging with Geographically Weighted Variograms," *Stochastic Environmental Research and Risk Assessment*, Vol. 24, No. 8, 2010, pp. 1193–1209.

[68] Clark, D. L., Jr., Bae, H. R., Gobal, K., and Penmetsa, R., "Engineering Design Exploration Using Locally Optimized Covariance Kriging," *AIAA Journal*, Vol. 54, No. 10, 2016, pp. 3160–3175.

[69] Zimmermann, R., "On the Condition Number Anomaly of Gaussian Correlation Matrices," *Linear Algebra and Its Applications*, Vol. 466, Feb. 2015, pp. 521–526.

[70] Lophaven, S. N., Nielsen, H. B., and Søndergaard, J., "DACE—A Matlab Kriging Toolbox (Version 2.0)," Technical Univ. of

Denmark, Dept. of Informatics and Mathematical Modelling, TR IMM-REP2002-12, Copenhagen, 2002, pp.1–34.

[71] Toal, D. J. J., Bressloff, N. W., and Kean, A. J., "Kriging Hyper Parameter Tuning Strategies," *AIAA Journal*, Vol. 46, No. 5, 2008, pp. 1240–1252.
doi: 10.2514/1.34822

[72] Jo, Y., Yi, S., Choi, S., Lee, D. J., and Choi, D. Z., "Adaptive Variable-Fidelity Analysis and Design Using Dynamic Fidelity Indicators," *AIAA Journal*, Vol. 54, No. 11, 2016, pp. 3564–3579.

[73] Forrester, A. I. J., Bressloff, N. W., and Keane, A. J., "Optimization Using Surrogate Models and Partially Converged Computational Fluid Dynamics Simulations," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, Vol. 462, No. 2071, 2006, pp. 2177–2204.

[74] Giunta, A. A., Wojtkiewicz, S. F., Jr., and Eldred, M. S., "Overview of Modern Design of Experiments Methods for Computational Simulations," *41st Aerospace Sciences Meeting and Exhibit*, AIAA Paper 2003-0649, Jan. 2003.

[75] Fang, K. T., Lin, D., Winker, P., and Zhang, Y., "Uniform Design: Theory and Application," *Technometrics*, Vol. 42, No. 3, 2000, pp. 237–248.

[76] Han, Z.-H., "SurroOpt: A Generic Surrogate-Based Optimization Code for Aerodynamic and Multidisciplinary Design," *Proceedings of ICAS 2016*, International Council of the Aeronautical Sciences Paper 2016-0281, Daejeon, Korea, 2016.

[77] Jones, D. R., Schonlau, M., and Welch, W. J., "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, Vol. 13, No. 4, 1998, pp. 455–492.

[78] Jones, D. R., "A Taxonomy of Global Optimization Methods Based on Response Surfaces," *Journal of Global Optimization*, Vol. 21, No. 4, 2001, pp. 345–383.

[79] Liu, J., Han, Z.-H., and Song, W.-P., "Comparison of Infill Sampling Criteria in Kriging-Based Aerodynamic Optimization," *Proceedings of the 28th Congress of the International Council of the Aeronautical Sciences*, International Council of the Aeronautical Sciences Paper 2012-0269, 2012.

[80] Sasena, M. L., Paplambros, P. Y., and Goodvaerts, P., "Exploration of Metamodeling Sampling Criteria for Constrained Global Optimization," *Engineering Optimization*, Vol. 34, No. 34, 2002, pp. 263–278.

[81] Cox, D. D., and John, S., "SDO: A Statistical Method for Global Optimization," *Proceedings of the IEEE International Conference on Systems, Man & Cybernetics*, IEEE Publ., Piscataway, NJ, 1992, pp. 1241–1246.

[82] Michalewicz, Z., "Genetic Algorithm, Numerical Optimization, and Constraints," *Proceedings of the 6th International Conference on Genetic Algorithms*, Morgan Kaufmann Publishers, San Francisco, CA, 1995, pp. 151–158.

[83] Kulfan, B. M., "Universal Parametric Geometry Representation Method," *Journal of Aircraft*, Vol. 45, No. 1, 2008, pp. 142–158.

[84] Xie, F., Song, W.-P., and Han, Z.-H., "Numerical Study of High-Resolution Scheme Based on Preconditioning Method," *Journal of Aircraft*, Vol. 46, No. 2, 2008, pp. 520–525.

K. E. Willcox
Associate Editor