Article

An Improved Blind Kriging Surrogate Model for Design Optimization Problems

Hau T. Mai [1,2], Jaewook Lee [1], Joowon Kang [3], H. Nguyen-Xuan [4] and Jaehong Lee [1,*]

1 Deep Learning Architectural Research Center, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Korea
2 Faculty of Mechanical Technology, Industrial University of Ho Chi Minh City, Ho Chi Minh City 72308, Vietnam
3 School of Architecture, Yeungnam University, 280, Daehak-ro, Gyeongsan 38541, Korea
4 CIRTech Institute, HUTECH University, Ho Chi Minh City 72308, Vietnam
* Correspondence: jhlee@sejong.ac.kr; Tel.: +82-2-3408-3287; Fax: +82-2-3408-4331

Abstract: Surrogate modeling techniques are widely employed in solving constrained expensive black-box optimization problems. Therein, Kriging is among the most popular surrogates in which the trend function is considered as a constant mean. However, it also encounters several challenges related to capturing the overall trend with a relatively limited number of function evaluations as well as searching feasible points with complex or discontinuous feasible regions. To address this above issue, this paper presents an improved surrogate blind Kriging (IBK) and a combined infill strategy to find the optimal solution. According to enhancing the prediction accuracy of metamodels of objective and constraints, the high-order effects of regression function in the blind Kriging are identified by promising a variable selection technique. In addition, an infill strategy is developed based on the probability of feasibility, penalization, and constrained expected improvement for updating blind Kriging metamodels of the objective and constraints. At each iteration, two infill sample points are allocated at the positions to achieve improvement in optimality and feasibility. The IBK metamodels are updated by the newly-added infill sample points, which leads the proposed framework search to rapidly converge to the optimal solution. The performance and applicability of the proposed model are tested on several numerical benchmark problems via comparing with other metamodel-based constrained optimization methods. The obtained results indicate that IBK generally has a greater efficiency performance and outperforms the competitors in terms of a limited number of function evaluations. Finally, IBK is successfully applied to structural design optimization. The optimization results show that IBK is able to find the best feasible design with fewer evaluation functions compared with other studies, and this demonstrates the effectiveness and practicality of the proposed model for solving the constrained expensive black-box engineering design optimization problems.

Keywords: blind Kriging; surrogate model; black-box optimization; infill strategy

MSC: 65Kxx

1. Introduction

The optimization process is critical in the engineering design, which requires lower computational cost, robustness, stability, and accuracy. However, most conventional optimization techniques face several challenges in solving the black-box global optimization problem, such as the unavailable expressions, gradient information of objective function or constraints, and the time-consuming effort. One way to circumvent these issues is based on the surrogate model (SM) [1,2]. Here, SM is an invaluable tool that is used to approximate the expensive computational models during the optimization process indirectly or directly. There are commonly used metamodels, such as response surface methodology [3,4], radial basis function (RBF) [5,6], Kriging [7,8], support vector regression

(SVR) [9], neural network (NN) [10–13], inverse distance weighting (IDW) [14], and so on. To achieve the optimal solution, the sampling technique and infill strategy are employed to build and iteratively refine the metamodel in order to ameliorate the solution found during the performance process. Therein, the sampling technique, which is also known as the design of the experiment, creates a set of points over the domain, and the initial metamodel is fitted to the observed points. Various sampling techniques for generating the sample points are developed, including Monte Carlo, random, and Latin hypercubic sampling (LHS) techniques. For more details, interested readers are suggested to refer to ref. [15].

In the past decade, variants of the metamodel-based optimization method have been successfully developed for solving the optimization problems with expensive simulations. For instance, Jones et al. [16] proposed an efficient global optimization (EGO) method where Kriging and expected improvement (EI) are introduced to find the solution. In addition, Huang et al. [17,18] extended the EGO algorithm by using the augmented EI function to determine the next sampling point. Additionally, a new infill strategy with adaptive radius and direction search is developed by Dong et al. [19] to indicate all local optimal values. To reduce the dimension of hyper-parameters, Zhao [20] incorporated the maximal information coefficient into Kriging. Gutmann [21] associated the measure of bumpiness in the RBF model to find the global minimum. Regis and Shoemaker [22–24] proposed a stochastic response surface to identify the promising points. More recently, Shepard [14] introduced the IDW, and then Joseph [25] added a linear regression function to improve accuracy. A conjugate between IDW and RBF is suggested by Bemporad [26].

The constrained black-box optimization problem is popular in many practical engineering designs. Hence, it has attracted the remarkable attention of researchers in recent years. They have been successfully applied to address this issue, such as ConstrLMSRBF [27], COBRA [28], RCGO [6], CARS [29], and so on. Among the surrogate models, Kriging has attracted more attention than other models due to estimating the prediction error and capacity to approximate the highly nonlinear functions. Therein, its regression function can be considered as the part trying to catch the general trend and thus the largest variations of the data. In recent times, Li et al. [30] proposed a new Kriging-based constrained global optimization algorithm in which the global optimal solution is obtained by two pivotal phases. An enhanced approach based on the modification of the probability of an improvement algorithm is presented by Carpio et al. Carpio et al. [31]. Additionally, Qian et al. [32] illustrated a new infill strategy, in which the position of the new sampling point is the intersection between the confidence interval and the constraint boundary. A combination of Kriging and the mixture of experts is proposed by Bartoli et al. [33] to improve the model's accuracy. In addition, Forrester and Keane [34] presented a constrained expected improvement criterion to attain a new sample for updating the surrogate model. Similarly, Shi et al. [35] developed a probability of constrained improvement based on filter technology. However, there are several challenges due to multimodal, nonlinear functions, limited sample size, and a strong overall trend exists. In addition, identifying the right regression function for a set of data with interactions between variables is a difficult task. To address the above issues, researchers have tried to improve the prediction accuracy of Kriging by adjusting either the regression function, the stochastic process, or both. A well-known author, Joseph et al. [36], presented the blind Kriging model, in which the optimal basis functions are estimated by the Bayesian variable selection technique. Additionally, Kersaudy et al. [37] introduced a combination of polynomial chaos expansions and universal Kriging. Zhang et al. [38,39] proposed the regularization method for constructing a trend function and the penalized blind likelihood Kriging. Nevertheless, the above-mentioned models are only established to solve the unconstrained optimization problem. Furthermore, to the best of our knowledge, it has still not been yet utilized for constrained optimization thus far.

In this study, an improved surrogate blind Kriging is first presented to handle computational expensive constrained optimization problems. In our work, the higher-order effects of the trend function are estimated by a Bayesian variable selection technique. Si-

multaneously, a new infill strategy is developed based on the probability of feasibility, penalization, and constrained expected improvement to refine the surrogate models and improve the solution found during the optimization process. According to this scheme, two sample points will be indicated corresponding to the exploration and exploitation stages for each iteration. It helps the IBK metamodel to learn the complex or discontinuous feasible domains. The performance and applicability of the proposed approach are demonstrated through several benchmark optimization problems. The outcomes of the proposed model are compared with other models to evaluate its efficiency and reliability. The obtained results showed that our work is able to find the best feasible design with fewer evaluation functions.

The rest of the paper is organized as follows. Section 2 introduces the IBK model. Next, an infill strategy based on the combination is presented in Section 3. Afterward, several numerical examples are investigated to demonstrate the efficiency of the proposed model in Section 4. Finally, the conclusions are outlined in Section 5.

2. Improved Blind Kriging

2.1. Basics

The primary objective of a metamodel is to search for the approximate function from a set of data points. A brief summary of Kriging and some basic formulas are described in this subsection.

Consider the function $y(x)$ as defined in the domain Ω. Let the n sample points $\mathbf{X} = \left\{ \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(n)} \right\}^T$ with $\mathbf{x} \in \mathbb{R}^d$ and the true response values $y = \left\{ y^{(1)}, y^{(2)}, ..., y^{(n)} \right\}^T$ with $y \in \mathbb{R}$, respectively. The universal Kriging (UK) model postulates a combination of the regression function and a stochastic process, as demonstrated in Equation (1)

$$Y(\mathbf{x}) = f(\mathbf{x}) + Z(\mathbf{x}), \tag{1}$$

where $Y(\mathbf{x})$ represents a black-box function; $f(\mathbf{x}) = \boldsymbol{v}(\mathbf{x})^T \boldsymbol{\beta}_m$ denotes the regression function; $Z(\mathbf{x})$ is characterized by Gaussian process with zero mean and stationary covariance. For the blind Kriging, the trend function utilizes a regression function, as defined as Equation (2)

$$f(\mathbf{x}) = \boldsymbol{v}(\mathbf{x})^T \boldsymbol{\beta}_m, \tag{2}$$

in which $\boldsymbol{v}(\mathbf{x}) = \{1, v_1(\mathbf{x}), ..., v_m(\mathbf{x})\}^T$ is the vector that contains known basic functions; m is the number of basis functions, and $\boldsymbol{\beta}_m = \{\beta_0, \beta_1, ..., \beta_m\}^T$ denotes the unknown coefficients [40]. All of them are identified through the feature selection methods in Section 2.2.

Let $\boldsymbol{\Psi}$ be the correlation matrix of the samples, $\psi(\mathbf{x})$ the correlation vector between a point $\mathbf{x}$ and the sample point in the data set, and $\boldsymbol{F}_m$ the model matrix of the sample points. Then the predicted value at any point $\mathbf{x}$ by the blind Kriging model is given as,

$$\hat{y}(\mathbf{x}) = \boldsymbol{v}(\mathbf{x})^T \hat{\boldsymbol{\beta}}_m + \boldsymbol{\psi}(\mathbf{x})^T \boldsymbol{\Psi}^{-1} \left(\mathbf{y} - F_m \hat{\boldsymbol{\beta}}_m \right), \tag{3}$$

where the variance $\hat{\sigma}_m^2$ and mean value $\hat{\boldsymbol{\beta}}_m$ of variance can be calculated as follows

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_m &= \frac{F_m^T \boldsymbol{\Psi}^{-1} \mathbf{y}}{F_m^T \boldsymbol{\Psi}^{-1} F_m} \\
\hat{\sigma}_m^2 &= \frac{\left(\mathbf{y} - F_m \hat{\boldsymbol{\beta}}_m \right)^T \boldsymbol{\Psi}^{-1} \left(\mathbf{y} - F_m \hat{\boldsymbol{\beta}}_m \right)}{n}.
\end{aligned}
\tag{4}
$$

To indicate the optimal hyper-parameters $\boldsymbol{\theta}$, in this paper, the differential evolution (DE) algorithm is employed to maximize the likelihood function. Their searching space is restricted to $(10^{-3}, 10^2)$ [34,41]. It should be noted that $\boldsymbol{\theta}$ is estimated two times. The first

time, they are found at $m = 0$, and after choosing m, the hyper-parameters will be estimated again. The interested reader is referred to [36,39,40] for more details about deriving them.

Mean square error (MSE) at points in the design domain is given by [42]

$$\hat{s}^2(\mathbf{x}) = \hat{\sigma}_m^2 \Big[1 - \boldsymbol{\psi}(\mathbf{x})^T \boldsymbol{\Psi}^{-1} \boldsymbol{\psi}(\mathbf{x}) + \Big(\boldsymbol{F}_m^T \boldsymbol{\Psi}^{-1} \boldsymbol{\psi}(\mathbf{x}) \\ - \boldsymbol{\nu}(\mathbf{x}) \Big)^T \Big(\boldsymbol{F}_m^T \boldsymbol{\Psi}^{-1} \boldsymbol{F}_m \Big)^{-1} \Big(\boldsymbol{F}_m^T \boldsymbol{\Psi}^{-1} \boldsymbol{\psi}(\mathbf{x}) - \boldsymbol{\nu}(\mathbf{x}) \Big). \tag{5}$$

To provide insight, a single variable test function from five initial samples, which is assumed as a black-box objective function, is investigated here, as shown in Figure 1. In which the solid black line represents the true function, the blue dashed line shows the prediction value, the red dots denote the set of sample points, and the solid red line shows the root mean squared error. It is easily seen that the predicted value at sample points is equal to the real value, and its MSE is zero. Note that MSE is used as the measure of uncertainty in each prediction and to estimate the approximate accuracy of the BK model.
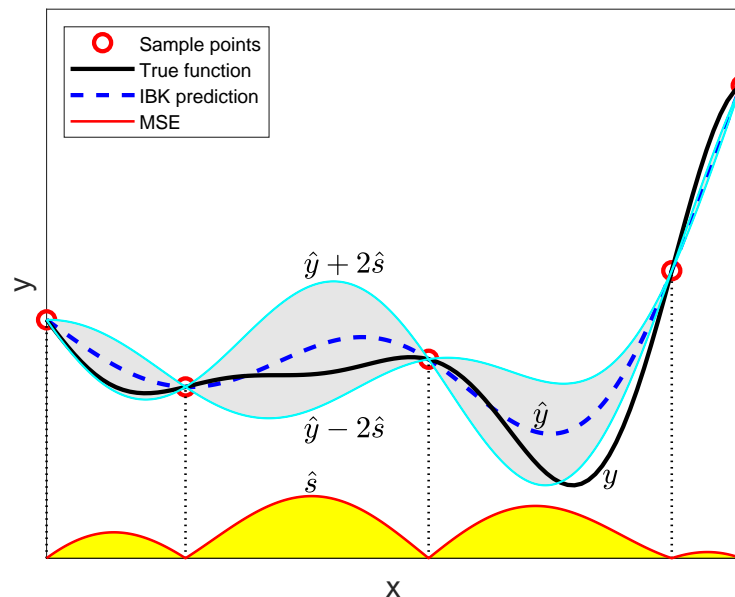

Figure 1. Mean square error of the IBK model.

2.2. Variable Selection

As mentioned by Couckuyt [40], BK can capture the most variance in the data set by data analysis methods. A collection of candidate functions is considered for selection by using cross-validation prediction error (CVPE). Now, we consider a trend function that combines the regression function with the set of candidate functions to fit the data in the following linear model

$$f(\mathbf{x}) = \sum_{i=0}^{m} \beta_i v_i(\mathbf{x}) + \sum_{i=0}^{t} \alpha_i u_i(\mathbf{x}), \tag{6}$$

where t is the number of potential functions; $u(\mathbf{x}) = \{1, u_1(\mathbf{x}), ..., u_t(\mathbf{x})\}$ denotes the set of candidate functions, and $\boldsymbol{\alpha}$ is the vector of corresponding coefficients. Note that $\boldsymbol{\beta}$ has already been determined independently $\boldsymbol{\alpha}$. When the transfer function f can be highly nonlinear, the number of candidate features to approximate are larger than the number of sample points. Therefore, we cannot determine all the coefficients α_i. In this study, a function prior to all parameters in the linear model is utilized to overcome this issue. Additionally, the trend function is improved by considering the high-order effect of the Bayesian variable selection technique, including linear, quadratic, cubic, quartic, and two-factor interaction effects. Consequently, the total of candidate variables and the mean

term is $t = 4\left(d^2 + 2\sum\limits_{i=1}^{d-1} i\right)$. Let us consider five-level, spaced factors with possible levels $l^{(1)}$, $l^{(2)}$, $l^{(3)}$, $l^{(4)}$, and $l^{(5)}$, which are defined as follows

$$
\begin{aligned}
l^{(1)} &= \min(\mathbf{X}) \quad , \\
l^{(2)} &= \frac{mean(\mathbf{X})}{2} \quad , \\
l^{(3)} &= mean(\mathbf{X}) \quad , \\
l^{(4)} &= \frac{3mean(\mathbf{X})}{2}, \\
l^{(5)} &= \max(\mathbf{X}) \quad .
\end{aligned}
\tag{7}
$$

Therefore, the model matrix using orthogonal polynomial coding, which is the column lengths of $\sqrt{5}$, is

$$
\mathbf{U}_j = \begin{bmatrix}
1 & -\sqrt{2} & \sqrt{\frac{10}{7}} & -\sqrt{\frac{1}{2}} & \sqrt{\frac{1}{14}} \\
1 & -\sqrt{\frac{1}{2}} & -\sqrt{\frac{5}{14}} & \sqrt{2} & -\sqrt{\frac{8}{7}} \\
1 & 0 & -\sqrt{\frac{10}{7}} & 0 & \sqrt{\frac{18}{7}} \\
1 & \sqrt{\frac{1}{2}} & -\sqrt{\frac{5}{14}} & -\sqrt{2} & -\sqrt{\frac{8}{7}} \\
1 & \sqrt{2} & \sqrt{\frac{10}{7}} & \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{14}}
\end{bmatrix}.
\tag{8}
$$

The corresponding correlation matrix is

$$
\mathbf{\Psi}_j = \begin{bmatrix}
1 & \psi_j\left(l_j^{(2)}\right) & \psi_j\left(l_j^{(3)}\right) & \psi_j\left(l_j^{(4)}\right) & \psi_j\left(l_j^{(5)}\right) \\
\psi_j\left(l_j^{(2)}\right) & 1 & \psi_j\left(l_j^{(2)}\right) & \psi_j\left(l_j^{(3)}\right) & \psi_j\left(l_j^{(4)}\right) \\
\psi_j\left(l_j^{(3)}\right) & \psi_j\left(l_j^{(2)}\right) & 1 & \psi_j\left(l_j^{(2)}\right) & \psi_j\left(l_j^{(3)}\right) \\
\psi_j\left(l_j^{(4)}\right) & \psi_j\left(l_j^{(3)}\right) & \psi_j\left(l_j^{(2)}\right) & 1 & \psi_j\left(l_j^{(2)}\right) \\
\psi_j\left(l_j^{(5)}\right) & \psi_j\left(l_j^{(4)}\right) & \psi_j\left(l_j^{(3)}\right) & \psi_j\left(l_j^{(2)}\right) & 1
\end{bmatrix},
\tag{9}
$$

and the $\mathbf{R}_j$ (5×5) variance-covariance matrix is calculated as follows

$$
\mathbf{R}_j = \mathbf{U}_j^{-1} \mathbf{\Psi}_j \left(\mathbf{U}_j^{-1}\right)^T.
\tag{10}
$$

Note that we only use the diagonal elements of $\mathbf{R}_j$ [43]. Hence, the ith element of the diagonal matrix $\mathbf{R}((t+1) \times (t+1))$ can be written as

$$
R_{i,i} = \prod_{j=1}^{d} r_{jl}^{l_{ij}} r_{jq}^{q_{ij}} r_{jc}^{c_{ij}} r_{jq_r}^{q_{r_{ij}}},
\tag{11}
$$

in which $l_{ij} = 1$ if α_i includes the factor's linear effect j and 0 otherwise. Similarly, q_{ij}, c_{ij}, $q_{r_{ij}}$ are assigned to the quadratic, cubic, and quartic effects, respectively.

$$
\begin{aligned}
r_{jl} &= \frac{R_j(2,2)}{R_j(1,1)}, \\
r_{jq} &= \frac{R_j(3,3)}{R_j(1,1)}, \\
r_{jc} &= \frac{R_j(4,4)}{R_j(1,1)}, \\
r_{jq_r} &= \frac{R_j(5,5)}{R_j(1,1)}, \\
0 &\leqslant l_{ij} + q_{ij} + c_{ij} + q_{r_{ij}} \leqslant 2.
\end{aligned}
\tag{12}
$$

The sample data will be normalized to the interval [0, 1]. The encoded samples for the linear, quadratic, cubic, and quartic effects are expressed as follows

$$
\begin{aligned}
x_{j,l} &= \frac{1}{\sqrt{2}} \left(\frac{x_j}{l_j^{(5)}} \right), \\[2mm]
x_{j,q} &= \sqrt{\frac{5}{14}} \left[\left(\frac{x_j}{l_j^{(5)}} \right)^2 - 2 \right], \\[2mm]
x_{j,c} &= \frac{5}{6\sqrt{2}} \left[\left(\frac{x_j}{l_j^{(5)}} \right)^3 - \frac{17}{5} \left(\frac{x_j}{l_j^{(5)}} \right) \right], \\[2mm]
x_{j,q_r} &= \frac{35}{12\sqrt{14}} \left[\left(\frac{x_j}{l_j^{(5)}} \right)^4 - \frac{31}{7} \left(\frac{x_j}{l_j^{(5)}} \right)^2 + \frac{72}{35} \right],
\end{aligned}
\tag{13}
$$

where x_j defines the jth column $\mathbf{X}$. The two-factor interaction terms can be built from these basic effects. After the completion of the construction of $\mathbf{R}$ as shown above, the posterior mean of $\boldsymbol{\alpha}$ is defined as

$$
\hat{\boldsymbol{\alpha}} = \left(\frac{\tau_m}{\sigma_m} \right)^2 \mathbf{R} \mathbf{F}_c{}' \boldsymbol{\Psi}^{-1} (\boldsymbol{y} - \mathbf{F}_m \boldsymbol{\beta}_m),
\tag{14}
$$

$$
\mathrm{var}(\hat{\boldsymbol{\alpha}}) = \tau_m^2 \left(\mathbf{R} - \left(\frac{\tau_m}{\sigma_m} \right)^2 \mathbf{R} \mathbf{F}_c{}' \boldsymbol{\Psi}^{-1} \mathbf{F}_c \mathbf{R} \right),
\tag{15}
$$

where $\mathbf{F}_c$ is the matrix of all candidate variables. As shown in refs. [36,40,43], the absolute of $\hat{\boldsymbol{\alpha}}$ is used instead of the standardized coefficient for variable selection. Hence, variables selected at each step correspond to the largest value $\hat{\boldsymbol{\alpha}}$. It should be noted that $\frac{\tau_m}{\sigma_m}$ is a constant, which is set to 1 for simplicity of the computation. To be specific, for the best value of m, the blind Kriging model has to go through five stages, which are described in detail as follows.

Step 1: The LHS technique is used to create initial sample points $\mathbf{X}$ and the obtained response values, respectively.

Step 2: Construct the ordinary Kriging surrogate model, and the leave-one-out cross-validation prediction error $\left(CVPE_{(m=0)} \right)$ is estimated to measure accuracy.

Step 3: Determine the coefficients $|\hat{\boldsymbol{\alpha}}|$ corresponding to each promising feature, and arrange $|\hat{\alpha}_i|$ from largest to smallest.

Step 4: If improving the accuracy of prediction has not been satisfied,

Step 4.1: Add a candidate function corresponding to each $|\hat{\alpha}_i|$ coefficient to the regression function.

Step 4.2: Construct the intermediate Kriging surrogate model with the new regression function.

Step 4.3: $CVPE_{(m)}$ is estimated to measure the accuracy of the model,

$$
\begin{aligned}
CVPE_{(m)} &= \sqrt{\frac{\sum\limits_{i=1}^{n} cv_i^2}{n}}, \\[2mm]
cv_i &= y_i(\boldsymbol{x}_i) - \hat{y}_i(\boldsymbol{x}_i) \quad .
\end{aligned}
\tag{16}
$$

Step 5: The best set of features to minimize $\left(CVPE_{(m)} \right)$, will be chosen to build the final BK model. Finally, the new hyper-parameters are indicated for the new trend function [40] of the surrogate model.

3. Infill Strategy

The infill strategy is employed to iteratively refine the metamodel and guide the algorithm toward a promising region to improve the solution found. Its trade-off between exploration and exploitation decides the success of the implementation. Although there are several available infill strategies, such as expected improvement (EI), expected violation, probability of improvement function (PI), lower confidence bounding, etc., these procedures may be difficult to find feasible points or feasible disconnected regions [6,28]. In this section, a new combined infill strategy, which consists of the probability of feasibility, penalization, and a constrained expected improvement, is introduced to overcome the above limitations.

The mathematical formulation of the constrained expensive black-box global optimization problem

$$
\begin{aligned}
&\min_{\mathbf{x}\in\mathbb{R}^d} f(\mathbf{x}), \\
&s.t.\ g_j(\mathbf{x}) \le 0,\quad j = 1,\ 2,...,\ m, \\
&\qquad \mathbf{x}_{lb} \le \mathbf{x} \le \mathbf{x}_{ub},
\end{aligned}
\tag{17}
$$

where f and g_j denote the objective function and inequality constraints, respectively. Normally, if there is one equality constraint, it is converted to two inequality constraints using a very small relaxation factor. $\mathbf{x}_{ub}$ and $\mathbf{x}_{lb}$ are the upper and lower bounds of the design variables, respectively. It should be noted that the above problem belongs to the class of derivative-free optimization, so its derivatives are not available. In order to solve Equation (17), the constrained EI is usually employed as the infill criterion and expressed as follows

$$
EI_C(\mathbf{x}) = EI_{C_i}(\mathbf{x}) + EI_{C_r}(\mathbf{x}),
\tag{18}
$$

with

$$
\begin{aligned}
EI_{C_i}(\mathbf{x}) &= (f(\mathbf{x}^*) - \mu_n(\mathbf{x}))\Phi\left(\frac{f(\mathbf{x}^*) - \mu_n(\mathbf{x})}{\sigma_n(\mathbf{x})}\right)\ PF(\mathbf{x}), \\
EI_{C_r}(\mathbf{x}) &= \sigma_n(\mathbf{x})\phi\left(\frac{f(\mathbf{x}^*) - \mu_n(\mathbf{x})}{\sigma_n(x)}\right)\ PF(\mathbf{x}), \\
PF(\mathbf{x}) &= \prod_{j=1}^{m} \Pr_j\big(g_j(\mathbf{x}) \le 0\big), \\
\Pr_j\big(g_j(\mathbf{x}) \le 0\big) &= \Phi\left(\frac{-\mu_{g_j}(\mathbf{x})}{\sigma_{g_j}(\mathbf{x})}\right),
\end{aligned}
$$

where $\mu_{g_j}(\mathbf{x})$ and $\sigma_{g_j}(\mathbf{x})$ are the posterior mean and covariance function of the jth constraint, respectively.

As indicated by Haftka et al. [44], $EI_{C_i}(\mathbf{x})$ indicates the exploitation sample points where the posterior mean of the objective function is small and has low uncertainty, and the constraints are likely to be near the boundaries of feasible operation. In contrast, $EI_{C_r}(\mathbf{x})$ aims to identify exploration sample points, which tend to be biased towards regions that are more likely to satisfy the constraints and high uncertainty of the objective function. However, the standard EI_C often fails when there is no initial feasible point in the data set [44]. To circumvent this, $PF(\mathbf{x})$ is one of the most used methods to favor the feasible region. Note that the value of $PF(\mathbf{x})$ is easy to be zero when one of the constraints violates the design specification [44]. Consequently, the high-priority data points are missed or overlooked.

To tackle these limitations, a combined computation strategy is first introduced in this study to achieve an initial feasible point as well as find the optimal solution. This scheme helps to reduce the number of constraint violations and explore other feasible regions, as follows:

$$
PFC(\mathbf{x})\ = \sigma_n(\mathbf{x})PF(\mathbf{x}),
\tag{19}
$$

$$PIC(\mathbf{x}) = -\sum_{j=1}^{m} \max\left(0, \mu_{g_j}(\mathbf{x})\right), \tag{20}$$

$$ISC_{14}(\mathbf{x}) = \prod_{j=1}^{m} Pr_j\left(g_j(\mathbf{x}) \leq 0\right).\, D_j, \tag{21}$$

where $PFC(\mathbf{x})$ is combined by the covariance function and the probability of feasibility function; $PIC(\mathbf{x})$ is the penalization function corresponding to the constraint set; $ISC_{14}(\mathbf{x})$ is the infill sampling criterion for disconnected feasible regions; D is the distance to the nearest feasible point, and given by:

$$D = \min_{\mathbf{x}_{feas}}\left(\frac{\left\|\mathbf{x}_{feas} - \mathbf{x}\right\|}{range}\right), \tag{22}$$

in which *range* is the lag distance at which it reaches the sill from the variogram model.

According to the proposed strategy, if the feasible point is not found in the data, $PFC(\mathbf{x})$ and $PIC(\mathbf{x})$ will be used as an alternative for obtaining two infill sample points. In this case, providing that there exists a feasible region, two feasible sample points will be identified. More specifically, the first point is located in the sparsely sampled area by maximizing $PFC(\mathbf{x})$, so $\sigma_n(\mathbf{x})$ characterizes the sample density of the objective function in the design space. The other point indicates the corresponding minimum value of the total constraints $PIC(\mathbf{x})$. When the models fit poorly and do not have a feasible region, $PIC(\mathbf{x})$ is utilized to improve the locations that violated constraints. On the contrary, once an initial feasible point is found, Equations (18) and (21) are employed to determine infill sample points. $ISC_{14}(\mathbf{x})$ aims to effectively explore other feasible regions. The algorithm framework is provided in Algorithm 1.

Algorithm 1 The combined infill strategy for the expensive constrained optimization problem

---

Input: d : number of design variables
     m : number of constraints
     N_{max}: maximum expensive evaluation number
     $[\mathbf{x}_{lb};\ \mathbf{x}_{ub}]$: design space
Output: $[\mathbf{x}_{best}, \mathbf{f}_{best}]$: the optimum solution
  1: Generate n_0 sample points using LHS from the design space, evaluate the fitness values, and collect a set of initial observations $\mathcal{D}_{n_0} = \{\mathbf{x}_i, y_i, \mathbf{g}_i\}\ \forall i = 1, ..., n_0$
  2: Determine $\mathcal{D}^{fes}$ from $\mathcal{D}_{n_0}$
  3: Set $n = n_0$
  4: **while** $n \leq N_{\max}$ **do**
  5:    Build surrogate models for the objective and each of the constraint functions
  6:    **if** $\mathcal{D}_n^{fes} = \emptyset$ **then**
  7:       Find $\mathbf{x}_{n+1}$ by maximizing Equation (19)
  8:       Find $\mathbf{x}_{n+2}$ by maximizing Equation (20)
  9:    **else**
10:       Find $\mathbf{x}_{n+1}$ by maximizing Equation (18)
11:       Find $\mathbf{x}_{n+2}$ by maximizing Equation (21)
12:    **end if**
13:    Evaluate function values at $\mathbf{x}_{n+1,2}$, append $\mathcal{D}_{n+2} = \mathcal{D}_n \cup \{(\mathbf{x}_{n+1,2}, y_{n+1,2}, \mathbf{G}_{n+1,2})\}$, and update $\mathcal{D}_{n+1,2}^{fes}$
14:    $n = n + 2$
15:    Update $[\mathbf{x}_{best}, \mathbf{f}_{best}]$
16: **end while**

---

4. Numerical Examples

4.1. Study in High-Order Effects of the Trend Function

To evaluate the high-order efficiency of the trend function, a planar truss structure for maximum passive vibration isolation [45], as shown in Figure 2, is introduced. The authors of Keane and Bright [46] achieved results for this problem through analysis and experiment. The structures consist of 42 members with the same material properties. The right-hand end nodes are fixed. The unit force excitation is subjected to node 11 over a frequency range 100–200Hz. The x- and y-coordinates of nodes seventh) and eighteenth are considered as design variables, and the other nodes are fixed as per the regular structure. The objective function is stress at the left-hand end node. The interested reader is referred to [45,46] for more details. LHS is used to obtain data for building metamodels that differ in the number of sample points, and 100 validation runs were performed to validate the results. In this case, the average Euclidean error (AEE) is used to measure the prediction error using

$$AEE(\hat{y}, y) = \frac{1}{k} \sum_{i=1}^{k} \sqrt{(\hat{y} - y)^2}. \tag{23}$$
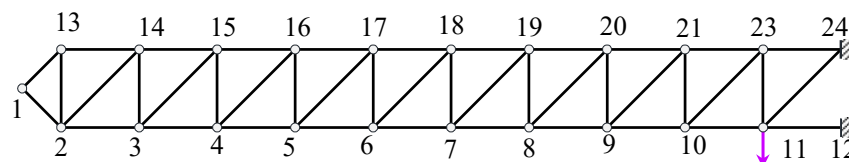

Figure 2. Planar truss structure. Adapted from [47].

As shown in Figure 3, it is easily seen that the BK model improves compared to the Kriging model. Clearly, the CVPE is smaller than the Kriging model. It is easy to explain that the linear, quadratic effect causes a significant rise in accuracy. On the other hand, the difference in AEE scores between the models was an insignificant amount with the number of sample points less than 30. This is explained by the fact that if the sample size is small, then the high-order effect of the regression function will not be much. However, the high-order BK model achieved a vast improvement over another model when the data size increased. Then, the regression function with the high-order effects is represented as Equations (24) and (25) for 100 and 200 sample points, respectively. Clearly, IBK can capture the overall trend to the largest variations of the data.

$$1 + x_1 x_3 + x_1 + x_4 + x_1 x_4 + x_3 x_4 + x_2 + x_1 x_2^3 + x_1^4 x_2^4 + x_2 x_4 + x_1^3 x_2 + x_1^4 x_3 + x_1^2 x_3 + x_1^2 x_4. \tag{24}$$

$$1 + x_1 x_4 + x_1 + x_1 x_3 + x_1^4 x_3 + x_1 x_2 + x_3 x_4 + x_1^4 x_2^4 + x_2 + x_1^4 x_3^2 + x_2^4 x_4 + x_1 x_3^2 + x_2^4 x_4^2. \tag{25}$$

4.2. Synthetic Test Problems

In this section, nine well-known benchmark problems, in which the number of variables, constraints, and the best know solutions are given Table 1, are investigated to evaluate the efficiency of the proposed model. More information about these problems can be seen in Appendix A and in ref. [28]. Although these optimization problems are not expensive to evaluate, they are treated as all computationally expensive functions to conduct meaningful comparisons of the performance of the alternative methods. The obtained results will be compared with the result of RCGO, FLT-AKM, COBRA-local, COBRA-global, and ConstrLMSRBF. Therein, the two algorithms RCGO and FLT-AKM are implemented by the authors following Wu et al. [6] and Shi et al. [35]. In the RCGO method, the distance coefficient in the first phase is $\Pi = [0.001, 0.005, 0.01, 0.05, 0.1]$, and $\Gamma = [0, 1, 2, 3, 4]$ is the value of the exponent in the second phase. In order to obtain a fair comparison between the different methods, all the tests are performed in a Matlab^TM environment using an Intel® Core(TM) i5-8500 CPU 3.0 GHz desktop machine (Microsoft, New York, NY, USA). Each metamodel is run ten times on each test problem to reduce the effect of random error.

Otherwise, the searching process of the model is terminated when either the maximum number of evaluation functions is reached or the best result does not improve after 10 iterations. LHS created ten different initial data sets. The differential evolution algorithm (DE) is employed to find hyper-parameters and the next points in the infill strategy.
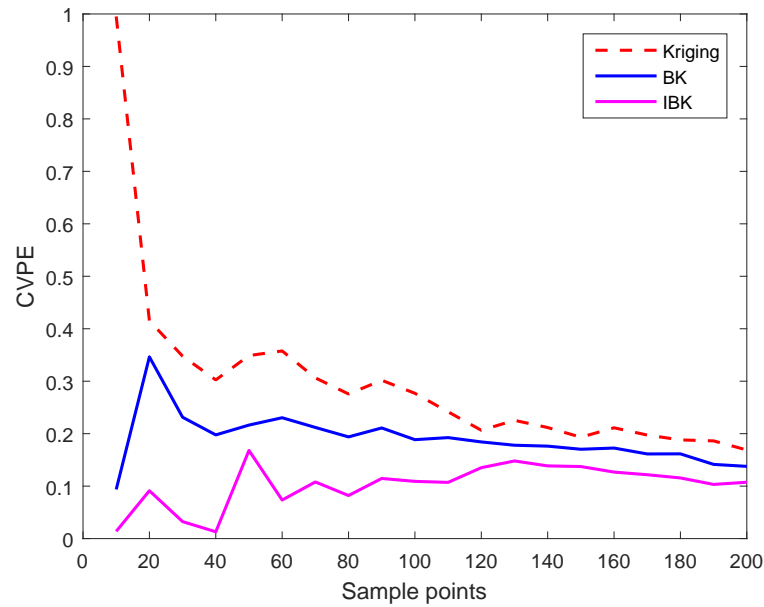
(**a**)
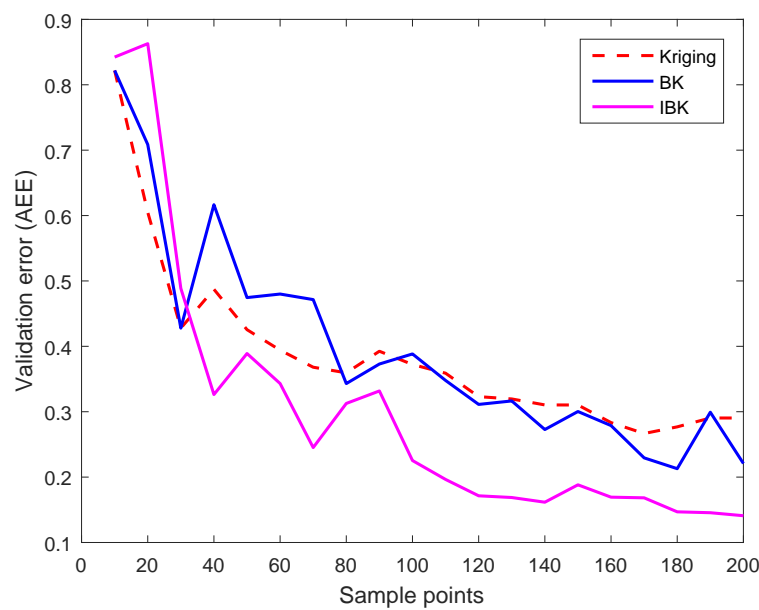
(**b**)

Figure 3. The relationship between flesh elements and thickness values. (**a**) Cross-validation prediction error of metamodels with different sample size. (**b**) Validation error of metamodels with different sample size.

Table 1. Constrained optimization benchmark problems.

Problems	Best Known Value	Dimensionality	No. Constraints
G24	-5.508	2	2
G8	-0.0958	2	2
Two_bars	0.0309	3	2
Four_bars	1400	4	1
Ibeam	0.0131	4	2
G5MOD	5126.5	4	5
G4	$-30,665.539$	5	6
Hesse	-310	6	6
SR	2994.42	7	11

The obtained results are summarized in Table 2, which provides the average number of evaluation functions (NEF), objective values of the best solution, median solutions, and mean solutions found by the six metamodels. Firstly, it is easily seen that RCGO has failed to indicate the feasible solution for G24, SR in all trials, and G5MOD in three trials. These are mainly due to the small feasible region, all the initial points are infeasible, or the incoherent distance coefficients. Clearly, IBK outperformed other algorithms for G24, G8, Two-bars, G4, Ibeam, and Hesse with the best solution. In addition, the IBK requires fewer evaluations than other metamodels. Figure 4 provides a graphical illustration of the overall process of identifying the infill samples of the G24 problem. Firstly, the feasible domain is discontinuous and includes one global and two local minimums corresponding to peaks of the feasible domain. All initial sample points are quite analogous to the metamodels for a fair comparison. Clearly, the infill samples obtained by IBK are distributed near the peaks of the feasible region as well as the boundary of the constraint set. On the contrary, the obtained result from FLT-AKM divided two groups of the infill samples, and the first group concentrates on the infeasible region far from the constrained boundary. This can easily be explained by the fact that the order of constraints is polynomials of degree four. Hence, IBK can easily capture the high-order effects compared with Kriging of the FLT-AKM. Furthermore, our infill strategy allows us to reduce the number of constraint violations and explore other feasible domains. More specifically, our model requires only 32 evaluation functions, while FLT-AKM requires 45 analyses for the convergence performance. This shows that the features of IBK and the combination infill strategy outperform other well-known existing algorithms.

For the Four-bars test, the obtained results of all these models are similar, and there are not many differences between the models. It is interesting that, among the problems, it requires few evaluation functions. Hence, IBK will not show much effect with small sample points or a simple, feasible domain. It can be observed that the optimal results found by the COBRA provide the best solution to the G5MOD and SR problems. However, IBK performs better than FLT-AKM and RCGO in terms of median, mean and best values of feasible optimum. Note that COBRA and RCGO require the initial data that contain the feasible points. However, it is difficult to find a feasible point for highly nonlinear problems, which then leads to failure. Clearly, RCGO could not find a feasible point in all trials for the SR problem. Although the result achieved from SR by IBK is close to FLT-AKM, IBK shows stability when the median and mean are close to the best value.
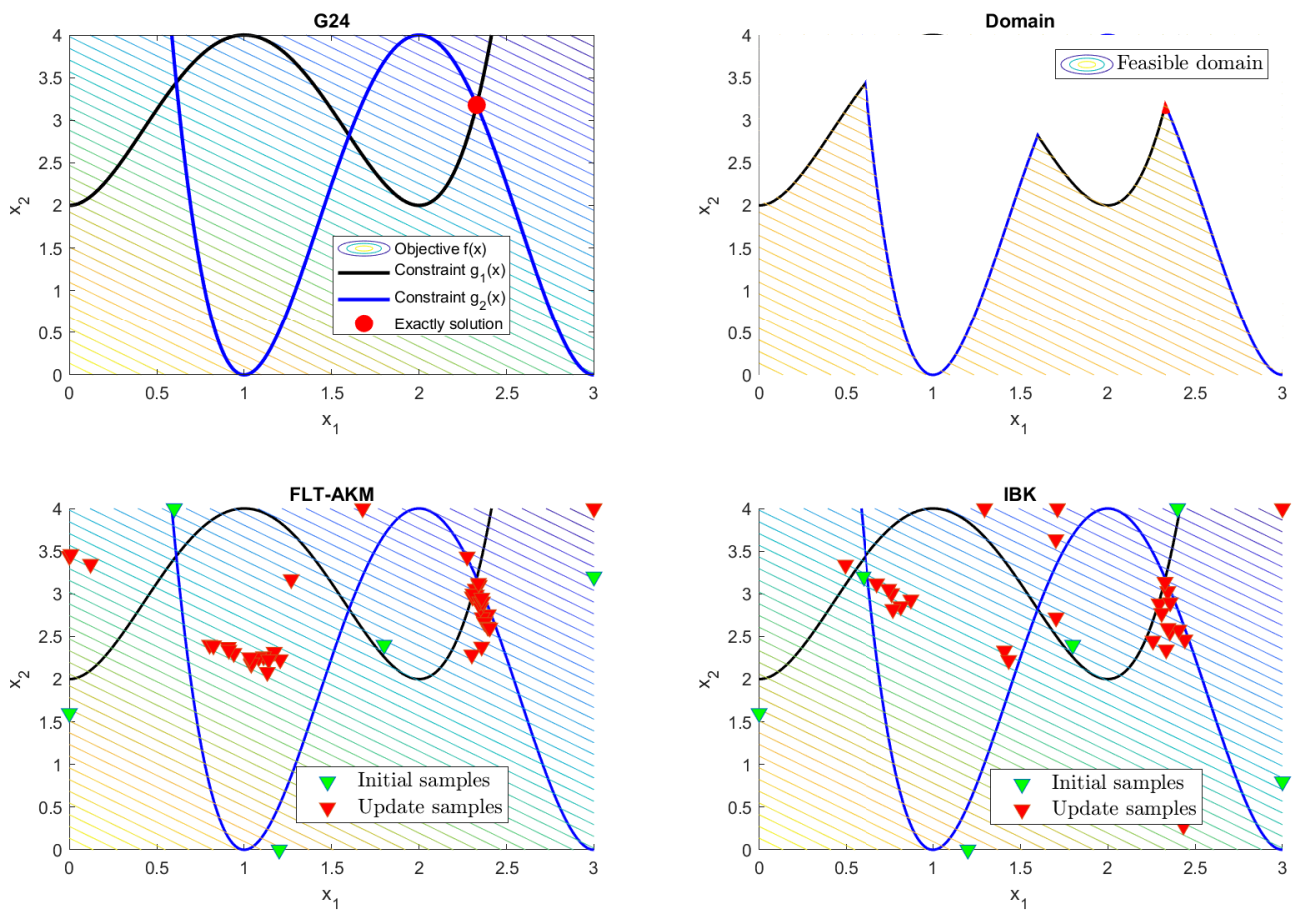
Figure 4. Optimization process-based infill strategy using IBK and FLT-AKM.

Table 2. Comparison of the obtained results for the test problems.

| | | Test Problems | | | | | | | | |
		G24	G8	Four Bars	Two Bars	Ibeam	G5MOD	G4	Hesse	SR
IBK	Best	−5.4669	−0.095	1400	0.031	0.013	5126.5	−30,665.5	−310	3042.2
	Median	−5.4075	−0.091	1400	0.031	0.014	5135.2	−30,665.5	−310	3051.3
	Mean	−5.3832	−0.089	1400	0.031	0.014	5167.1	−30,665.5	−310	3053.3
	NEF	33.2	46.8	17.4	36.2	60	74.0	24.2	38.6	68.4
FLT-AKM	Best	−5.46	−0.095	1400	0.031	0.013	5739.9	−30,611.7	−306.553	3024.4
	Median	−5.4173	−0.084	1400	0.031	0.015	6157.9	−30,416.8	−297.34	3061
	Mean	−5.415	−0.085	1400	0.031	0.018	6187.8	−30,411.2	−297.093	3058.5
	NEF	34	54.8	17.5	37.8	74.8	59.2	35.2	58.6	59.6
RCGO	Best		−0.096	1400	0.031	0.013	5323.5	−30,628.8	−306.428	
	Median		−0.096	1400	0.031	0.014	6099.8	−30,389.5	−294.812	
	Mean	N/A(10)	−0.073	1400	0.033	0.016	5992.9	−30,395.3	−296.239	N/A(10)
	NEF		51.7	17.2	40.4	67.8	65.4(3)	35	73.2	
COBRA-Local	Best		−0.1				5126.5	−30,665.5	−309.94	2994.4
	Median		−0.1				5126.51	−30,665.2	−297.87	2994.7
[28]	Mean	-	−0.09	-	-	-	5126.51	−30,665.1	−296.25	2994.7
	NEF		50				50	50	50	50
COBRA-Global	Best		−0.1				5126.5	−30,665.4	−309.97	2994.7
	Median		−0.1				5126.51	−30,664.9	−297.87	2994.7
[28]	Mean	-	−0.09	-	-	-	5126.62	−30,664.9	−296.25	2994.7
	NEF		50				50	50	50	50

4.3. Structural Design Optimization

Finally, the proposed model is applied to perform the shape optimization of truss structures. This benchmark problem has been previously analyzed by Shao [48] and Suprayitno [49] using the clustering-based surrogate model and Kriging, respectively. A 21-bars planar truss shown in Figure 5 is investigated for design optimization. Coordinates of all free nodes are treated as continuous design variables. The design aim is to maximize the truss performance, which is defined as the maximum load ratio (F_{max}) that the truss can carry to the self-weight of the truss (W). The structure is subjected to stress and stability ratio constraints. The stress limitations of all members are the same for tension and compression and set to 200 MPa. The cross-sectional areas of all bars are 25 cm^2. The number of initial sample points is set at 14 [48]. Here, the critical buckling stress $S_{cr,i} = \pi^2 E/(L_{e,i}/\rho)^2$. Whereas E and ρ are Young's modulus of the truss material and the radius of gyration of the truss bar, respectively. It can be formulated as follows

$$
\begin{aligned}
&Maximize \quad \frac{F_{max}}{W} = f(x_1, x_2, x_3, y_3, x_4, y_4, y_6), \\
&subjected to \ \{\sigma_t, \sigma_c\} \leqslant \sigma_a, \\
&\qquad\qquad \frac{\sigma_{c,i}}{S_{cr,i}} \leqslant 1, \\
&\qquad\qquad 0.8 \leqslant x_1, x_3, y_3, y_4, y_6 \leqslant 3 \text{ m}, \\
&\qquad\qquad 3.3 \leqslant x_2, x_4 \leqslant 5.5 \text{ m}.
\end{aligned}
\tag{26}
$$

The number of maximum evaluation functions is set to 100 for all models. The obtained optimal results are listed in Table 3 and Figure 6. It can easily be seen that IBK found the best feasible solution, where the objective value is 25.8541, while the maximum tensile stress and maximum compressive stress are 94.21 and -115.701 Mpa, respectively. In addition, IBK significantly decreased the number of simulations. These have demonstrated that IBK can solve expensive black-box structure design optimization problems effectively.

Table 3. Comparison of the obtained results for the 21-bars planar truss.

Design Variables	IBK	FLT-AKM	RCGO	CMLS [48]	EORKS [49]
x_1	3	3	2.98	2.37	3
x_2	4.31	4.251	4.19	4.174	4.209
x_3	1.907	1.91	1.91	1.877	1.902
y_3	2.267	2.275	2.278	1.721	2.285
x_4	4.218	4.218	4.237	4.024	4.222
y_4	1.706	1.708	1.705	1.438	1.709
y_6	2.723	2.702	2.624	2.186	2.644
Best	25.85	25.46	24.72	22.79	25.66
Median	25.27	25.21	23.46	-	-
Mean	25.36	25.33	23.75	-	-
NFE	55.6	65.7	72.3	95	87
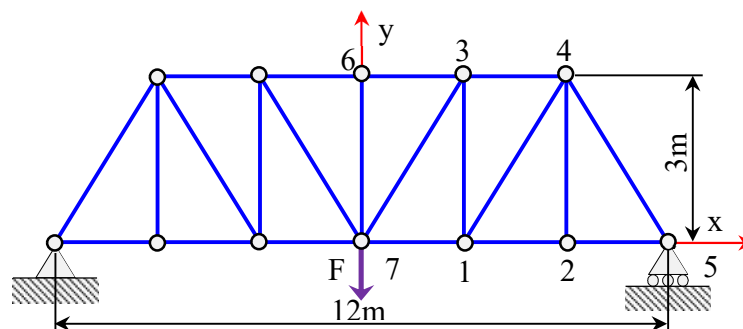
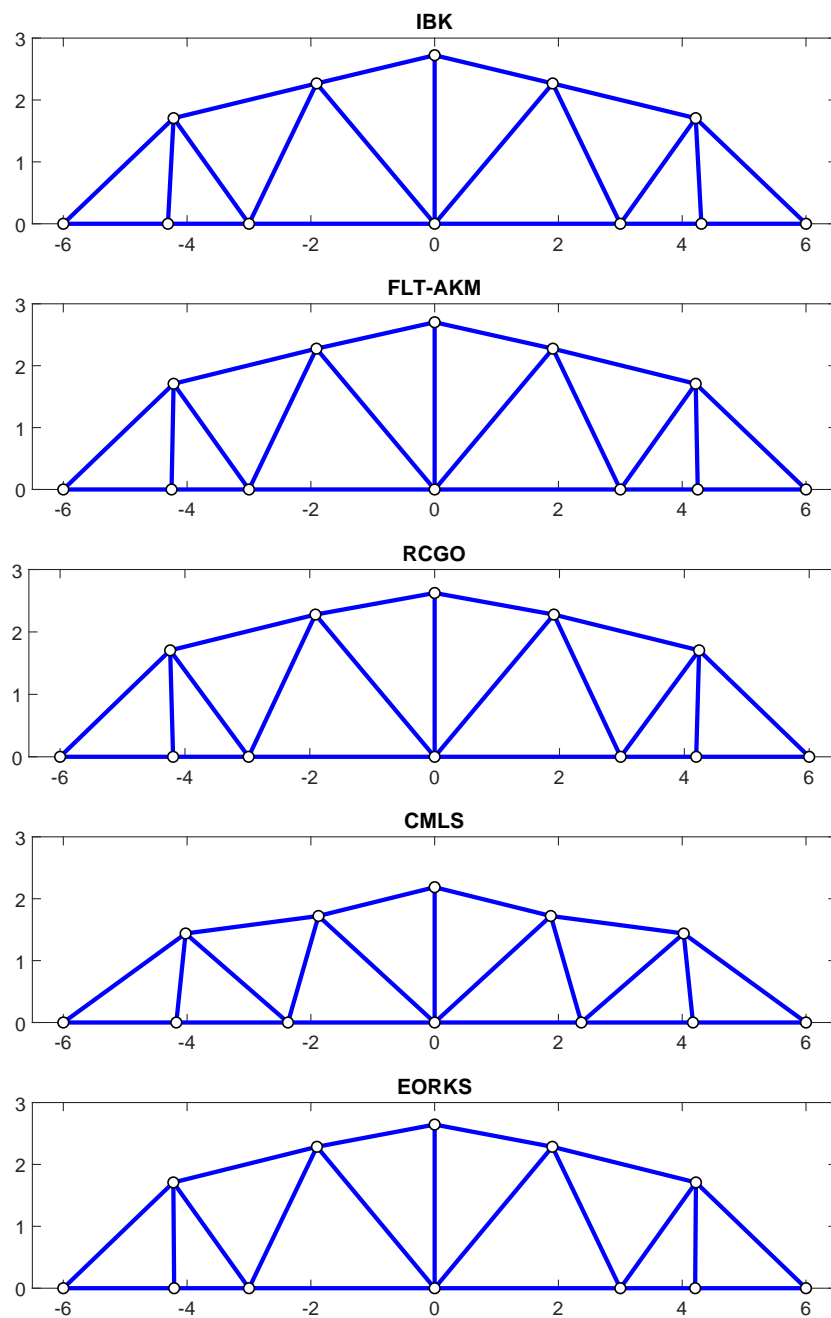**Figure 5.** A 21-bars planar truss structure.

Figure 6. Comparison of optimal shapes of the 21-bars planar truss obtained by IBK.

5. Conclusions

In this study, an improved surrogate blind Kriging combined with the new infill strategy is introduced to solve the constrained expensive black-box optimization problems. According to that core idea, the high-order effects of the trend function are identified from the training data using the Bayesian variable selection method. Simultaneously, the novel infill strategy based on the probability of feasibility, penalization, and constrained expected improvement is constructed to reduce the number of constraint violations, explore other feasible domains, and handle the discontinuous design domain. The efficiency of the proposed approach is demonstrated through several numerical examples for synthetic test problems and shapes the optimization of the truss structure. The obtained results have indicated that the optimum solution obtained by this work is in good agreement with ten of the eleven tests. The proposed model saves a number of evaluation functions in almost

all problems in comparison with other metamodels. In addition, it is also shown that the IBK predictor is simpler to interpret and is more robust than an ordinary Kriging predictor. This proved to be an efficient model for solving real-world engineering problems. Hence, it promises to be robust and effective at resolving complex problems in the expensive optimization of a complex design domain.

Appendix A

G5MOD [28]:

$$\min f(x) = 3x_1 + 10 - 6x_1^3 + 2x^2 + (2 \times 10^{-6}/3)x_2^3$$

$$s.t.\ g_1(x) = x_3 - x_4 - 0.55 \leqslant 0$$

$$g_2(x) = x_4 - x_3 - 0.55 \leqslant 0$$

$$g_3(x) = 1000sin(-x_3 - 0.25) + 894.8 - x_1$$
$$+ 1000sin(-x_4 - 0.25) \leqslant 0$$

$$g_4(x) = 1000sin(x_3 - 0.25) + 894.8 - x_2$$
$$+ 1000sin(x_3 - x_4 - 0.25) \leqslant 0$$

$$g_5(x) = 1000sin(x_4 - 0.25) + 1294.8$$
$$+ 1000sin(x_4 - x_3 - 0.25) \leqslant 0$$

$$0 \leqslant x_1, x_2 \leqslant 1200;\ -0.55 \leqslant x_3,\ x_4 \leqslant 0.55$$

Hesse [28]:

$$\min f(x) = -25(x_1 - 2)^2 - (x_2 - 2)^2 - (x_3 - 1)^2$$
$$- (x_4 - 4)^2 - (x_5 - 1)^2 - (x_6 - 4)^2$$

$$s.t.\ g_1(x) = (2 - x_1 - x_2)/2 \leqslant 0$$

$$g_2(x) = (x_1 + x_2 - 6)/6 \leqslant 0$$

$$g_3(x) = (2 - x_1 + x_2)/2 \leqslant 0$$

$$g_4(x) = (x_1 - 3x_2 - 2)/2 \leqslant 0$$

$$g_5(x) = \left(4 - x_4 - (x_3 - 3)^2\right)/4 \leqslant 0$$

$$g_6(x) = \left(4 - x_6 - (x_5 - 3)^2\right)/4 \leqslant 0$$

$$0 \leqslant x_1 \leqslant 5;\ 0 \leqslant x_2 \leqslant 4;\ 1 \leqslant x_{3,5} \leqslant 5;$$
$$0 \leqslant x_4 \leqslant 6;\ 0 \leqslant x_6 \leqslant 10$$

G4 [28]:

$$\min f(x) = 5.3578547x_3^2 + 0.8356891x_1x_5$$
$$+ 37.293239x_1 - 40792.141$$

$s.t. \, g_1(x) = -u \leqslant 0;$

$\quad g_2(x) = u - 92 \leqslant 0$

$\quad g_3(x) = -v + 90 \leqslant 0;$

$\quad g_4(x) = v - 110 \leqslant 0$

$\quad g_5(x) = -w + 20 \leqslant 0;$

$\quad g_6(x) = w - 25 \leqslant 0$

$\quad u = 85.334407 + 0.0056858x_2x_5$

$\qquad + 0.0006262x_1x_4 - 0.0022053x_3x_5$

$\quad v = 80.51249 + 0.0071317x_2x_5$

$\qquad + 0.0029955x_1x_2 + 0.0021813x_2^3$

$\quad w = 9.300961 + 0.0047026x_3x_5$

$\qquad + 0.0012547x_1x_3 + 0.0019085x_3x_4$

$\quad 78 \leqslant x_1 \leqslant 102; \;\; 33 \leqslant x_2 \leqslant 45;$

$\quad 27 \leqslant x_i \leqslant 45 \;\; for \; i = 3, 4, 5$

G24 [28]:

$\min f(x) = -x_1 - x_2$

$s.t. \, g_1(x) = -2x_1^4 + 8x_1^3 - 8x_1^2 + x_2 - 2 \leqslant 0$

$\quad g_2(x) = -4x_1^4 + 32x_1^3 - 88x_1^2 + 96x_1$

$\qquad\quad + x_2 - 36 \leqslant 0$

$\quad 0 \leqslant x_1 \leqslant 3; \;\;\; 0 \leqslant x_2 \leqslant 4$

Ibeam [28]:

$$\min f(x) = \frac{5000}{\frac{1}{12}x_3(x_1 - 2x_4)^3 + \frac{1}{6}x_2x_4^3 + 2x_2x_4\left(\frac{x_1 - x_4}{2}\right)^2}$$

$s.t. \, g_1(x) = 2x_2x_4 + x_3(x_1 - 2x_4) \leqslant 300$

$$\quad g_2(x) = \frac{180000x_1}{x_3(x_1 - 2x_4)^3 + 2x_2x_4\left(4x_4^2 + 3x_1(x_1 - 2x_4)\right)}$$

$$\qquad + \frac{15000x_2}{(x_1 - 2x_4)x_3^3 + 2x_2^3x_4} \leqslant 6$$

$\quad 10 \leqslant x_1 \leqslant 80; \;\;\; 10 \leqslant x_2 \leqslant 50; \;\;\; 0.9 \leqslant x_{3,4} \leqslant 5$

SR [28]:

$\min f(x) = 0.7854x_1x_2^2A - 1.508x^1B$

$\qquad\qquad + 7.477C + 0.7854D$

$s.t. \, g_1(x) = (27 - x_1x_2^2x_3)/27 \leqslant 0$

$\quad g_2(x) = (397.5 - x_1x_2^2x_3^2)/397.5 \leqslant 0$

$\quad g_3(x) = (1.93 - (x_2x_6^4x_3)/x_4^2)/1.93 \leqslant 0$

$\quad g_4(x) = (1.93 - (x_2x_7^4x_3)/x_5^3)/1.93 \leqslant 0$

$\quad g_5(x) = ((A_1/B_1) - 1100)/1100 \leqslant 0$

$\quad g_6(x) = ((A_2/B_2) - 850)/850 \leqslant 0$

$\quad g_7(x) = (x_2x_3 - 40)/40 \leqslant 0$

$\quad g_8(x) = (5 - (x_1/x_2))/5 \leqslant 0$

$\quad g_9(x) = ((x_1/x_2) - 12)/12 \leqslant 0$

$\quad g_{10}(x) = (1.9 + 1.5x_6 - x_4)/1.9 \leqslant 0$

$\quad g_{11}(x) = (1.9 + 1.1x_7 - x_5)/1.9 \leqslant 0$

$$A = 3.3333x_3^2 + 14.9334x_3 - 43.0934;$$
$$B = x_6^2 + x_7^2; \ C = x_6^3 + x_7^3; \ D = x_4 x_6^2 + x_5 x_7^2$$
$$A_1 = [(745x_4/(x_2x_3))^2 + (16.91 \times 10^6)]^{0.5}$$
$$A_2 = [(745x_5/(x_2x_3))^2 + (157.5 \times 10^6)]^{0.5}$$
$$B_1 = 0.1x_3; \ B_2 = 0.1x_3$$
$$2.6 \leqslant x_1 \leqslant 3.6; \quad 0.7 \leqslant x_2 \leqslant 0.8$$
$$17 \leqslant x_3 \leqslant 28; \quad 7.3 \leqslant x_4, x_5 \leqslant 8.3$$
$$2.9 \leqslant x_6 \leqslant 3.9; \quad 5.0 \leqslant x_7 \leqslant 5.5$$

G8 [28]:

$$\min f(x) = \frac{-\sin^3(2\pi x_1)\sin(2\pi x_2)}{x_1^3(x_1 + x_2)}$$

$$s.t. \ g_1(x) = x_1^2 - x_2 + 1 \leqslant 0$$
$$g_2(x) = 1 - x_1 + (x_2 - 4)^2 \leqslant 0$$
$$0 \leqslant x_1, x_2 \leqslant 10$$

Four-bar truss [50]:

The objective here is to minimize the total structural volume subject to the stress constraints on the members as in Figure A1.

$$\min \left(2a_1 + \sqrt{2}a_2 + \sqrt{2}a_3 + a_4\right)$$

$$s.t. \ \frac{FL}{E}\left(\frac{2}{a_1} + \frac{2\sqrt{2}}{a_2} - \frac{2\sqrt{2}}{a_3} + \frac{2}{a_4} \leqslant 0.04\right)$$

$$1 \leqslant a_{1,4} \leqslant 3; \quad \sqrt{2} \leqslant a_{2,3} \leqslant 3$$
$$L = 200 \ \text{cm}; \ F = 10 \ \text{kN}; \ E = 200{,}000 \ \text{kN/cm}^2$$
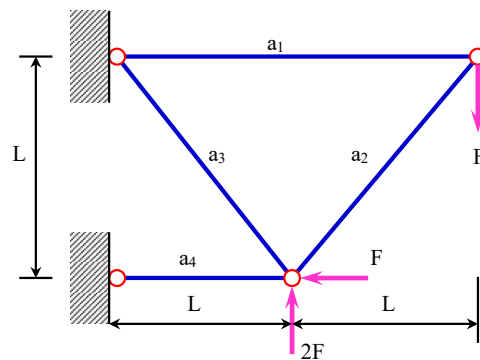

Figure A1. Four-bar truss.

Two-bar truss [50]:

The objective here is to minimize the total structural volume under stress constraints on member cross-sectional areas as in Figure A2.

$$\min \left(a_1\sqrt{b^2 + L^2} + a_2\sqrt{L^2 + (2L - b)^2}\right)$$

$$s.t. \ \frac{W_2 b(2L - b) + W_1 Lb}{2L^2 a_1} \leqslant S_{\max}$$

$$\frac{(W_2 b - W_1 L)\sqrt{L^2 + (2L - b)^2}}{2L^2 a_2} \leqslant S_{\max}$$

$$5.16 \leqslant a_{1,2} \leqslant 19.4 \ \text{cm}^2; \ 9.144 \leqslant b \leqslant 27.432 \ \text{m}$$
$$W_1 = 445 \ \text{kN}; \ W_2 = 4450 \ \text{kN}$$
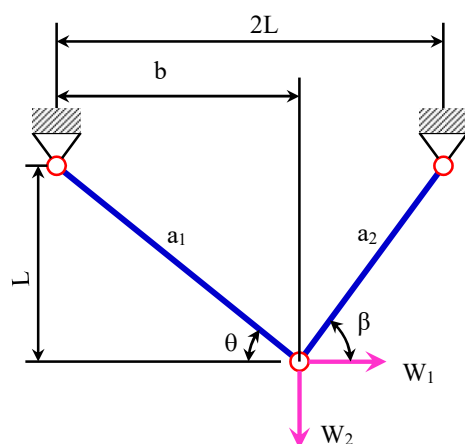$$L = 18.288 \ \text{m}; \ S_{\max} = 3.79 \times 10^6 \ \text{kN/m}^2$$

Figure A2. Two-bar truss.

References

1. Amine Bouhlel, M.; Bartoli, N.; Regis, R.G.; Otsmane, A.; Morlier, J. Efficient global optimization for high-dimensional constrained problems by using the Kriging models combined with the partial least squares method. *Eng. Optim.* **2018**, *50*, 2038–2053. [CrossRef]
2. Gu, J.; Zhang, H.; Zhong, X. Hybrid meta-model-based global optimum pursuing method for expensive problems. *Struct. Multidiscip. Optim.* **2019**, *61*, 543–554. [CrossRef]
3. Myers, R.H.; Montgomery, D.C.; Anderson-Cook, C.M. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*; John Wiley & Sons: Hoboken, NJ, USA, 2016.
4. Box, G.E.; Draper, N.R. *Empirical Model-Building and Response Surfaces*; Wiley: New York, NY, USA, 1987; Volume 424.
5. Powell, M. *The Theory of Radial Basis Function Approximation in 1990, Advances in Numerical Analysis II: Wavelets, Subdivision, and Radial Functions (WA Light, ed.)*; Oxford University Press: Oxford, UK, 1992; Volume 105, p. 210.
6. Wu, Y.; Yin, Q.; Jie, H.; Wang, B.; Zhao, J. A RBF-based constrained global optimization algorithm for problems with computationally expensive objective and constraints. *Struct. Multidiscip. Optim.* **2018**, *58*, 1633–1655. [CrossRef]
7. Sacks, J.; Welch, W.J.; Mitchell, T.J.; Wynn, H.P. Design and analysis of computer experiments. *Stat. Sci.* **1989**, *4*, 409–423. [CrossRef]
8. Sakata, S.; Ashida, F.; Zako, M. Structural optimization using Kriging approximation. *Comput. Methods Appl. Mech. Eng.* **2003**, *192*, 923–939. [CrossRef]
9. Clarke, S.M.; Griebsch, J.H.; Simpson, T.W. Analysis of support vector regression for approximation of complex engineering analyses. *J. Mech. Des.* **2005**, *127*, 1077–1087. [CrossRef]
10. Berke, L.; Hajela, P. Applications of artificial neural nets in structural mechanics. In *Shape and Layout Optimization of Structural Systems and Optimality Criteria Methods*; Springer: Berlin/Heidelberg, Germany, 1992; pp. 331–348.
11. Mai, H.T.; Kang, J.; Lee, J. A machine learning-based surrogate model for optimization of truss structures with geometrically nonlinear behavior. *Finite Elem. Anal. Des.* **2021**, *196*, 103572. [CrossRef]
12. Zhang, W.; Li, X.; Ma, H.; Luo, Z.; Li, X. Federated learning for machinery fault diagnosis with dynamic validation and self-supervision. *Knowl.-Based Syst.* **2021**, *213*, 106679. [CrossRef]
13. Zhang, W.; Li, X.; Ma, H.; Luo, Z.; Li, X. Universal domain adaptation in fault diagnostics with hybrid weighted deep adversarial learning. *IEEE Trans. Ind. Inform.* **2021**, *17*, 7957–7967. [CrossRef]
14. Shepard, D. A two-dimensional interpolation function for irregularly-spaced data. In Proceedings of the 1968 23rd ACM National Conference, New York, NY, USA, 27–29 August 1968; pp. 517–524.
15. Shields, M.D.; Zhang, J. The generalization of Latin hypercube sampling. *Reliab. Eng. Syst. Saf.* **2016**, *148*, 96–108. [CrossRef]
16. Jones, D.R.; Schonlau, M.; Welch, W.J. Efficient global optimization of expensive black-box functions. *J. Glob. Optim.* **1998**, *13*, 455–492. [CrossRef]
17. Huang, D.; Allen, T.T.; Notz, W.I.; Zeng, N. Global optimization of stochastic black-box systems via sequential kriging meta-models. *J. Glob. Optim.* **2006**, *34*, 441–466. [CrossRef]
18. Huang, D.; Allen, T.T.; Notz, W.I.; Miller, R.A. Sequential kriging optimization using multiple-fidelity evaluations. *Struct. Multidiscip. Optim.* **2006**, *32*, 369–382. [CrossRef]
19. Dong, H.; Song, B.; Wang, P.; Huang, S. A kind of balance between exploitation and exploration on kriging for global optimization of expensive functions. *J. Mech. Sci. Technol.* **2015**, *29*, 2121–2133. [CrossRef]
20. Zhao, L.; Wang, P.; Song, B.; Wang, X.; Dong, H. An efficient kriging modeling method for high-dimensional design problems based on maximal information coefficient. *Struct. Multidiscip. Optim.* **2019**, *61*, 39–57. [CrossRef]
21. Gutmann, H.M. A radial basis function method for global optimization. *J. Glob. Optim.* **2001**, *19*, 201–227. [CrossRef]

22. Regis, R.G.; Shoemaker, C.A. Improved strategies for radial basis function methods for global optimization. *J. Glob. Optim.* **2007**, *37*, 113–135. [CrossRef]

23. Regis, R.G.; Shoemaker, C.A. A stochastic radial basis function method for the global optimization of expensive functions. *INFORMS J. Comput.* **2007**, *19*, 497–509. [CrossRef]

24. Regis, R.G.; Shoemaker, C.A. Parallel stochastic global optimization using radial basis functions. *INFORMS J. Comput.* **2009**, *21*, 411–426. [CrossRef]

25. Joseph, V.R.; Kang, L. Regression-based inverse distance weighting with applications to computer experiments. *Technometrics* **2011**, *53*, 254–265. [CrossRef]

26. Bemporad, A. Global optimization via inverse distance weighting. *arXiv* **2019**, arXiv:1906.06498.

27. Regis, R.G. Stochastic radial basis function algorithms for large-scale optimization involving expensive black-box objective and constraint functions. *Comput. Oper. Res.* **2011**, *38*, 837–853. [CrossRef]

28. Regis, R.G. Constrained optimization by radial basis function interpolation for high-dimensional expensive black-box problems with infeasible initial points. *Eng. Optim.* **2014**, *46*, 218–243. [CrossRef]

29. Nuñez, L.; Regis, R.G.; Varela, K. Accelerated random search for constrained global optimization assisted by radial basis function surrogates. *J. Comput. Appl. Math.* **2018**, *340*, 276–295. [CrossRef]

30. Li, Y.; Wu, Y.; Zhao, J.; Chen, L. A Kriging-based constrained global optimization algorithm for expensive black-box functions with infeasible initial points. *J. Glob. Optim.* **2017**, *67*, 343–366. [CrossRef]

31. Carpio, R.R.; Giordano, R.C.; Secchi, A.R. Enhanced surrogate assisted framework for constrained global optimization of expensive black-box functions. *Comput. Chem. Eng.* **2018**, *118*, 91–102. [CrossRef]

32. Qian, J.; Yi, J.; Cheng, Y.; Liu, J.; Zhou, Q. A sequential constraints updating approach for Kriging surrogate model-assisted engineering optimization design problem. *Eng. Comput.* **2019**, *36*, 993–1009. [CrossRef]

33. Bartoli, N.; Lefebvre, T.; Dubreuil, S.; Olivanti, R.; Priem, R.; Bons, N.; Martins, J.R.; Morlier, J. Adaptive modeling strategy for constrained global optimization with application to aerodynamic wing design. *Aerosp. Sci. Technol.* **2019**, *90*, 85–102. [CrossRef]

34. Forrester, A.; Sobester, A.; Keane, A. *Engineering Design via Surrogate Modelling: A Practical Guide*; John Wiley & Sons: Hoboken, NJ, USA, 2008.

35. Shi, R.; Liu, L.; Long, T.; Wu, Y.; Tang, Y. Filter-based adaptive Kriging method for black-box optimization problems with expensive objective and constraints. *Comput. Methods Appl. Mech. Eng.* **2019**, *347*, 782–805. [CrossRef]

36. Joseph, V.R.; Hung, Y.; Sudjianto, A. Blind kriging: A new method for developing metamodels. *J. Mech. Des.* **2008**, *130*, 031102. [CrossRef]

37. Kersaudy, P.; Sudret, B.; Varsier, N.; Picon, O.; Wiart, J. A new surrogate modeling technique combining Kriging and polynomial chaos expansions–Application to uncertainty analysis in computational dosimetry. *J. Comput. Phys.* **2015**, *286*, 103–117. [CrossRef]

38. Zhang, Y.; Yao, W.; Ye, S.; Chen, X. A regularization method for constructing trend function in Kriging model. *Struct. Multidiscip. Optim.* **2019**, *59*, 1221–1239. [CrossRef]

39. Zhang, Y.; Yao, W.; Chen, X.; Ye, S. A penalized blind likelihood Kriging method for surrogate modeling. *Struct. Multidiscip. Optim.* **2019**, *61*, 457–474. [CrossRef]

40. Couckuyt, I.; Forrester, A.; Gorissen, D.; De Turck, F.; Dhaene, T. Blind Kriging: Implementation and performance analysis. *Adv. Eng. Softw.* **2012**, *49*, 1–13. [CrossRef]

41. Forrester, A.I.; Keane, A.J. Recent advances in surrogate-based optimization. *Prog. Aerosp. Sci.* **2009**, *45*, 50–79. [CrossRef]

42. Palar, P.S.; Shimoyama, K. On efficient global optimization via universal Kriging surrogate models. *Struct. Multidiscip. Optim.* **2018**, *57*, 2377–2397. [CrossRef]

43. Joseph, V.R.; Delaney, J.D. Functionally induced priors for the analysis of experiments. *Technometrics* **2007**, *49*, 1–11. [CrossRef]

44. Haftka, R.T.; Villanueva, D.; Chaudhuri, A. Parallel surrogate-assisted global optimization with expensive functions—A survey. *Struct. Multidiscip. Optim.* **2016**, *54*, 3–13. [CrossRef]

45. Forrester, A.; Jones, D. Global optimization of deceptive functions with sparse sampling. In Proceedings of the 12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Victoria, BC, Canada, 10–12 September 2008; p. 5996.

46. Keane, A.; Bright, A. Passive vibration control via unusual geometries: Experiments on model aerospace structures. *J. Sound Vib.* **1996**, *190*, 713–719. [CrossRef]

47. Couckuyt, I.; De Turck, F.; Dhaene, T.; Gorissen, D. Automatic surrogate model type selection during the optimization of expensive black-box problems. In Proceedings of the 2011 Winter Simulation Conference (WSC), Phoenix, AZ, USA, 11–14 December 2011; IEEE: New York, NY, USA, 2011; pp. 4269–4279.

48. Suprayitno; Yu, J.C. Evolutionary reliable regional Kriging surrogate for expensive optimization. *Eng. Optim.* **2019**, *51*, 247–264. [CrossRef]

49. Shao, T.; Krishnamurty, S. A clustering-based surrogate model updating approach to simulation-based engineering design. *J. Mech. Des.* **2008**, *130*, 041101. [CrossRef]

50. Andrei, N.; Andrei, N. *Nonlinear Optimization Applications Using the GAMS Technology*; Springer: Berlin/Heidelberg, Germany, 2013.