

# Kriging surrogate model with coordinate transformation based on likelihood and gradient

Nobuo Namura<sup>1</sup>  · Koji Shimoyama<sup>1</sup> · Shigeru Obayashi<sup>1</sup>

Received: 9 December 2015 / Accepted: 28 March 2017 / Published online: 4 April 2017  
© Springer Science+Business Media New York 2017

**Abstract** The Kriging surrogate model, which is frequently employed to apply evolutionary computation to real-world problems, with a coordinate transformation of the design space is proposed to improve the approximation accuracy of objective functions with correlated design variables. The coordinate transformation is conducted to extract significant trends in the objective function and identify the suitable coordinate system based on either one of two criteria: likelihood function or estimated gradients of the objective function to each design variable. Compared with the ordinary Kriging model, the proposed methods show higher accuracy in the approximation of various test functions. The proposed method based on likelihood shows higher accuracy than that based on gradients when the number of design variables is less than six. The latter method achieves higher accuracy than the ordinary Kriging model even for high-dimensional functions and is applied to an airfoil design problem with spline curves as an example with correlated design variables. This method achieves better performances not only in the approximation accuracy but also in the capability to explore the optimal solution.

**Keywords** Kriging model · Efficient global optimization · Coordinate transformation · Likelihood · Eigenvalue decomposition · Airfoil design

---

✉ Nobuo Namura  
namura@edge.ifs.tohoku.ac.jp

Koji Shimoyama  
shimoyama@edge.ifs.tohoku.ac.jp

Shigeru Obayashi  
obayashi@edge.ifs.tohoku.ac.jp

<sup>1</sup> Institute of Fluid Science, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai, Miyagi 980-8577, Japan

## 1 Introduction

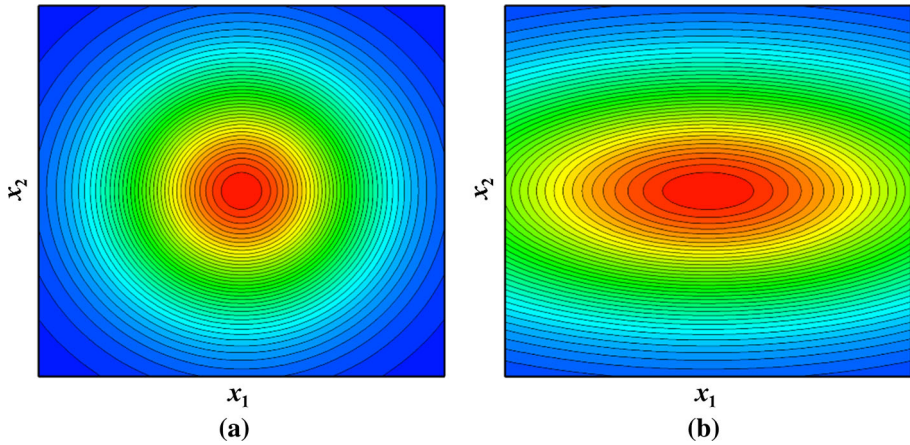
Optimization in real-world problems is usually time consuming and computationally expensive in the evaluation of objective functions [1,2]. Surrogate models are often useful to solve this difficulty. Surrogate models are constructed to promptly estimate the values of the objective functions at any point from sample points where real values of the objective functions are obtained by expensive computations. Therefore, it is important that accurate models can be constructed even with a small number of sample points.

The most common surrogate model is polynomial regression (PR) [3]. In construction of the PR model, users give the polynomial order arbitrarily and then compute the coefficients of each term in the polynomial fitting the sample points using the least-squares method. The accuracy of the PR model significantly depends on the polynomial order, which corresponds to the number of local maxima and/or minima in an objective function. However, it is not always possible to achieve sufficient accuracy by adjusting the order because the real shape of the objective function is usually not known. Generally, quadratic functions are employed to approximate the function locally in the real-world problems [4]. In this case, adequate optimization cannot be performed if the objective function has some local optima and too many sample points are required to obtain the global optimum. Furthermore, a Pareto dominance-based evolutionary multi-objective optimization algorithm explores large design space where diverse Pareto-optimal solutions exist. The surrogate models are also required to approximate large design space as accurately as possible. From this point of view, the local PR model with quadratic functions is not suitable for multi-objective problems.

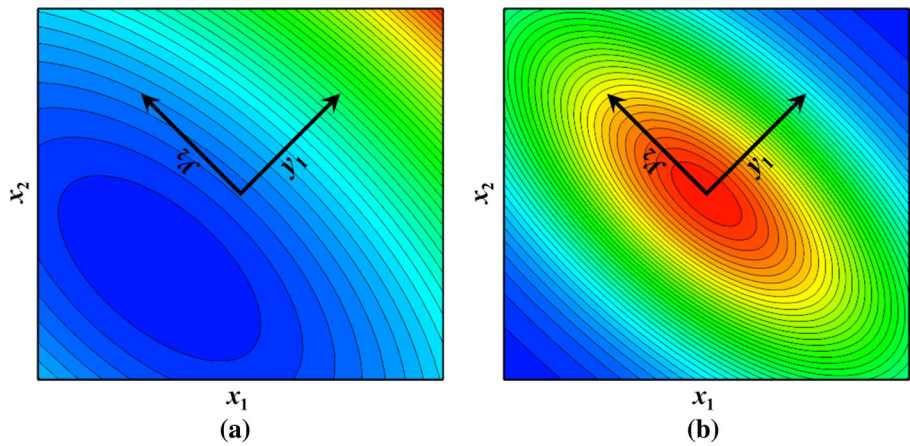
To approximate complex functions, radial basis function (RBF) networks [5] and the Kriging model [6] are often used. Both of them can adapt well to complex functions because they approximate a function as a weighted superposition of basis functions such as Gaussian function. Thus, the model complexity can be controlled by changing the weight coefficients and variance of each basis function. Gaussian basis functions in the Kriging model have independently different variance values along each design variable direction to fit the complexity and scale while those of the RBF have the same values (Fig. 1). This anisotropy enhances the accuracy of the Kriging model. In addition, the Kriging model gives not only estimated function values but also approximation errors, which help users determine the locations of the additional sample points to improve the accuracy of the surrogate model. Jones et al. [7] have proposed the efficient global optimization (EGO) in which additional sample points are selected by maximizing the expected improvement (EI) value derived from the estimated function value and the approximation error. The EI value enables the Kriging model to efficiently explore the optimum as well as to improve the model accuracy.

It is desirable that one of these models approximates the function accurately. However, sometimes more complex models are needed. Hybrid methods that combine two surrogate models may be effective if the function consists of complex macro- and micro-trends [8,9]. The universal Kriging model (PR+Kriging) [10] and the extended RBF (PR+RBF) [11] are typical hybrid methods. In contrast, Xiong et al. [12] proposed the non-stationary covariance based Kriging model whose variance values of each basis function vary depending on the location of the basis functions in the design space. This model showed good performance if the complexity of the objective function changes according to the location.

Some design variables can be correlated with each other in real-world problems, e.g., control points of spline curves and free-form deformation (FFD) [13]. Optimization with



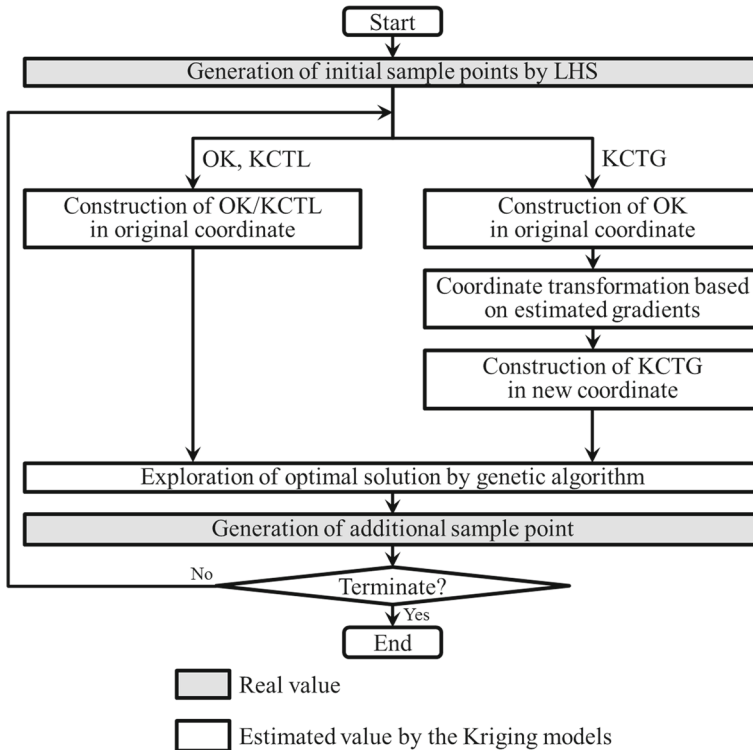
**Fig. 1** Gaussian basis functions. **a** Radial basis function, **b** Kriging model



**Fig. 2** Extraction of suitable coordinates. **a** Objective function, **b** basis function

such design variables can express various configurations while the problem tends to become difficult to solve. However, only the PR model takes account of the correlation as the cross-terms among different variables though the PR model does not approximate the complex function accurately due to the reason described above.

In this study, we propose modified Kriging models suitable for the problems with correlated design variables by focusing on the anisotropy of Gaussian basis functions. It means finding out suitable coordinates in the design variable space, which represent significant trends in the objective function (Fig. 2a) and then defining the variance of basis functions along each coordinate in the transformed system (Fig. 2b). Suitable coordinates are identified using two criteria: one is based on likelihood derived by the Kriging model and the other uses estimated gradients of the objective function. The Kriging models with coordinate transformation based on likelihood (KCTL) and gradients (KCTG) and the ordinary Kriging (OK) model are applied to test functions and airfoil design problems to investigate the feature of the proposed models.



**Fig. 3** Flowchart of EGO with KCTL, KCTG, and OK

## 2 Construction of Kriging models

The flowchart explaining EGO with KCTL, KCTG, and OK is summarized in Fig. 3. Initial sample points are generated uniformly in the design space using the Latin-hypercube sampling (LHS) method [14]. The Kriging models are constructed by interpolating these sample points. Construction of KCTG consists of three parts: construction of the Kriging model (OK) in the original coordinate system, coordinate transformation, and reconstruction of the Kriging model (KCTG) in the transformed coordinate system while OK and KCTL are constructed in a single part. An optimal solution is explored by the EGO framework, which explores the solution where the EI value becomes maximum by an optimizer such as evolutionary algorithm (the non-dominated sorting genetic algorithm II (NSGA-II) [15] is employed in our EGO system toward the application to multi-objective optimization in real-world problems) and adds it as an additional sample point. EGO is completed by iterating the procedure illustrated in Fig. 3 until a termination condition is satisfied. Generally, execution time and cost consumed for surrogate-based optimization are dominated by the function evaluation at sample points while those for surrogate model construction and optimal solution exploration on the model are ignorable. Hence, it can be said that better surrogate models should approximate functions more accurately with less sample points. The following present details of the Kriging models.

### 2.1 Ordinary Kriging model

The Kriging model expresses the unknown function  $f(\mathbf{x})$  as

$$f(\mathbf{x}) = \mu(\mathbf{x}) + \varepsilon(\mathbf{x}), \tag{1}$$

where  $\mathbf{x}$  is an  $m$ -dimensional vector ( $m$  design variables),  $\mu(\mathbf{x})$  is a global model, and  $\varepsilon(\mathbf{x})$  represents a local deviation from the global model, which is defined as the Gaussian process following  $N(0, \sigma^2)$ . The correlation between  $\varepsilon(\mathbf{x}^i)$  and  $\varepsilon(\mathbf{x}^j)$  is strongly related to the distance between the two corresponding points,  $\mathbf{x}^i$  and  $\mathbf{x}^j$ . In the Kriging model, a specially weighted distance is used instead of the Euclidean distance because the latter weighs all design variables equally. The distance function between the points at  $\mathbf{x}^i$  and  $\mathbf{x}^j$  is expressed as

$$d(\mathbf{x}^i, \mathbf{x}^j) = \sum_{k=1}^m \theta_k (x_k^i - x_k^j)^2, \tag{2}$$

where  $\theta_k$  ( $0 \leq \theta_k < \infty$ ) is the weight coefficient and the  $k$ -th element of an  $m$ -dimensional weight vector  $\boldsymbol{\theta}$ . These weights give the Kriging model anisotropy and enhance its accuracy. The correlation between the points  $\mathbf{x}^i$  and  $\mathbf{x}^j$  is defined as

$$\text{Corr}(\varepsilon(\mathbf{x}^i), \varepsilon(\mathbf{x}^j)) = \exp(-d(\mathbf{x}^i, \mathbf{x}^j)). \tag{3}$$

The Kriging predictor is

$$\hat{f}(\mathbf{x}) = \hat{\mu}(\mathbf{x}) + \mathbf{r}^T \mathbf{R}^{-1}(\mathbf{f} - \hat{\boldsymbol{\mu}}), \tag{4}$$

where  $\hat{\mu}(\mathbf{x})$  is the estimated value of  $\mu(\mathbf{x})$ ,  $\mathbf{R}$  denotes the  $n \times n$  matrix whose  $(i, j)$  entry is  $\text{Corr}(\varepsilon(\mathbf{x}^i), \varepsilon(\mathbf{x}^j))$ ,  $\mathbf{r}$  is an  $n$ -dimensional vector whose  $i$ -th element is  $\text{Corr}(\varepsilon(\mathbf{x}), \varepsilon(\mathbf{x}^i))$ , and  $\mathbf{f}$  and  $\hat{\boldsymbol{\mu}}$  denote as follows ( $n$  sample points):

$$\mathbf{f} = (f(\mathbf{x}^1) \cdots f(\mathbf{x}^n))^T, \tag{5}$$

$$\hat{\boldsymbol{\mu}} = (\hat{\mu}(\mathbf{x}^1) \cdots \hat{\mu}(\mathbf{x}^n))^T. \tag{6}$$

Thus, the unknown parameters in the Kriging model are  $\hat{\sigma}^2$  (estimated  $\sigma^2$ ),  $\hat{\mu}(\mathbf{x})$ , and  $\boldsymbol{\theta}$ , which are obtained by maximizing the following log-likelihood function:

$$\begin{aligned} Ln(\hat{\mu}, \hat{\sigma}^2, \boldsymbol{\theta}) = & -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\hat{\sigma}^2) - \frac{1}{2} \ln(|\mathbf{R}|) \\ & - \frac{1}{2\hat{\sigma}^2} (\mathbf{y} - \hat{\boldsymbol{\mu}})^T \mathbf{R}^{-1} (\mathbf{y} - \hat{\boldsymbol{\mu}}). \end{aligned} \tag{7}$$

$\hat{\sigma}^2$  is analytically determined through partial differentiation as

$$\hat{\sigma}^2 = \frac{(\mathbf{f} - \hat{\boldsymbol{\mu}})^T \mathbf{R}^{-1} (\mathbf{f} - \hat{\boldsymbol{\mu}})}{n}. \tag{8}$$

The definition of  $\hat{\mu}(\mathbf{x})$  has some variations. The OK model, which is the most widely used Kriging model, assumes the global model to be a constant value as  $\hat{\mu}(\mathbf{x}) = \hat{\mu}$ . In this case,  $\hat{\mu}$  is also analytically determined as

$$\hat{\mu}(\mathbf{x}) = \hat{\mu} = \frac{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{f}}{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}}, \tag{9}$$

where  $\mathbf{1}$  denotes an  $n$ -dimensional unit vector. Plugging in Eq. (8) for Eq. (7), the log-likelihood function becomes

$$Ln(\hat{\mu}, \hat{\sigma}^2, \boldsymbol{\theta}) = -\frac{n}{2} (\ln(2\pi) + 1) - \frac{n}{2} \ln(\hat{\sigma}^2) - \frac{1}{2} \ln(|\mathbf{R}|). \tag{10}$$

The first term can be ignored in the maximization because it has a constant value. Therefore, the log-likelihood maximization becomes an  $m$ -dimensional unconstrained non-linear optimization problem. In this study, a simple genetic algorithm is adopted to solve this problem.

### 2.2 Expected improvement

The accuracy of the function value predicted by the Kriging model depends largely on the distance from sample points. The closer point  $\mathbf{x}$  is to the sample points, the more accurate the prediction,  $\hat{f}(\mathbf{x})$ , becomes. This is expressed as:

$$s^2(\mathbf{x}) = \hat{\sigma}^2 \left( \mathbf{1} - \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} + \frac{(\mathbf{1} - \mathbf{1}^T \mathbf{R}^{-1} \mathbf{r})^2}{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}} \right), \tag{11}$$

where  $s^2(\mathbf{x})$  is the mean square error at point  $\mathbf{x}$ , which indicates the uncertainty of the estimated value. Thus, estimated values in the Kriging model do not have deterministic values but follows the Gaussian distribution denoted by  $N(\hat{f}(\mathbf{x}), s^2(\mathbf{x}))$ , from which the probability that the solution at point  $\mathbf{x}$  may achieve a new global optimum can be calculated. The EI value, which corresponds to the expected value of the objective function improvement from the current optimal solution among the sample points, is also derived by using this probability. In  $f(\mathbf{x})$  minimization problem, the improvement value  $I(\mathbf{x})$  and the EI value,  $E(I(\mathbf{x}))$  of  $f(\mathbf{x})$  are expressed, respectively, as

$$I(\mathbf{x}) = \max(f_{ref} - f, 0), \tag{12}$$

$$E(I(\mathbf{x})) = \int_{-\infty}^{f_{ref}} (f_{ref} - f)\varphi(f)df, \tag{13}$$

where  $f_{ref}$  is the reference value of  $f$  and corresponds to the minimum value of  $f$  among the sample points in this study.  $\varphi$  is the probability density function denoted by  $N(\hat{f}(\mathbf{x}), s^2(\mathbf{x}))$  and represents uncertainty about  $f$ .

Special modification for EI has been proposed to enhance the constrained optimization [16]. A modified EI value is expressed by multiplying the probability satisfying the constraint to the conventional EI value. If the constraint function which should be approximated by the Kriging model is expressed as  $g(\mathbf{x}) > c$ , the modified EI ( $E_c I$ ) value is calculated as follows:

$$E_c(I(\mathbf{x})) = E(I(\mathbf{x})) \int_c^\infty \varphi(g)dg. \tag{14}$$

### 2.3 Coordinate transformation based on likelihood function

An affine coordinate transformation is used in KCTL. Generally, such transformations are capable of rotation, scaling, translation, and shear mapping. However, scaling has the same effect as tuning the weight coefficient,  $\theta_k$  for each design variable in the Kriging model, and translation has no effect in the approximation of objective functions because design variables are normalized before constructing surrogate models. Only rotation is considered in KCTL because shear mapping violates the orthonormality of the design variables.

A rotation in multi-dimensional space is defined on each plane consisting of two arbitrary dimensions (design variables). In  $m$ -dimensional space, a rotation on all  ${}_m C_2 = m(m - 1)/2$

planes needs to be defined. For example, rotations in three-dimensional space are expressed as the product of three rotation matrices as follows.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi_3 & -\sin \phi_3 \\ 0 & \sin \phi_3 & \cos \phi_3 \end{bmatrix} \begin{bmatrix} \cos \phi_2 & 0 & -\sin \phi_2 \\ 0 & 1 & 0 \\ \sin \phi_2 & 0 & \cos \phi_2 \end{bmatrix} \begin{bmatrix} \cos \phi_1 & -\sin \phi_1 & 0 \\ \sin \phi_1 & \cos \phi_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \tag{15}$$

Note that the order of the rotational planes must be defined uniquely before the fact because the final coordinate transformation matrix changes with order even if the same rotation angles are applied to each rotational plane. KCTL is generated in the rotated coordinate system,  $\mathbf{y} = [y_1, \dots, y_m]^T$  instead of the original coordinate system,  $\mathbf{x} = [x_1, \dots, x_m]^T$ .

Rotation angles in each plane are determined to maximize log-likelihood function in Eq. (10) which becomes a  $m(m + 1)/2$ -dimensional optimization problem with rotation angles,  $\phi_\kappa$  ( $0^\circ \leq \phi_\kappa < 90^\circ | \kappa = 1, \dots, m(m - 1)/2$ ) as variables besides  $\theta_\kappa$  in the KCTL construction.  $\phi_\kappa$  from  $0^\circ$  to  $90^\circ$  is sufficient to arbitrarily define the rotated coordinate system because axes consisting of the plane where the rotation is conducted are orthogonal to each other. This optimization problem has a tendency to become extremely high-dimensional one (15-dimension when  $m = 5$ , 55-dimension when  $m = 10$ , and 120-dimension when  $m = 15$ ). Thus, we adopt an artificial bee colony (ABC) algorithm [17], whose robustness to high-dimensional problems has already been confirmed [18], to solve this problem.

### 2.4 Coordinate transformation based on estimated gradients

To identify suitable coordinates and improve the approximation accuracy, gradients of the objective function to each design variable are employed in KCTG as is the case in the active subspace method [19]. First, the  $m \times m$  covariance matrix  $\mathbf{C}$ , whose  $(k, l)$  entry is

$$C_{kl} = \frac{\partial \hat{f}}{\partial x_k} \frac{\partial \hat{f}}{\partial x_l}, \tag{16}$$

is defined. The objective function estimated by the Kriging model,  $\hat{f}$  is used in this study whereas gradients of the real objective function are used in [19]. After the construction of OK in Fig. 3, the estimated gradients are calculated by differentiating Eq. (4) analytically as follows:

$$\frac{\partial \hat{f}(\mathbf{x})}{\partial x_k} = \left( \frac{\partial \mathbf{r}}{\partial x_k} \right)^T \mathbf{R}^{-1} (\mathbf{f} - \hat{\boldsymbol{\mu}}), \tag{17}$$

where  $i$ -th element of  $\partial \mathbf{r} / \partial x_k$  is

$$\frac{\partial \mathbf{r}}{\partial x_k} \Big|_i = -2\theta_k (x_k - x_k^i) \exp \left( - \sum_{l=1}^m \theta_l (x_l^- - x_l^i)^2 \right). \tag{18}$$

Using estimated gradients, neither the finite difference of the real objective function nor the adjoint computation [20] is needed; also the function evaluation costs are reduced immensely. Note that we can deal with the objective function as a black box function. This study calculates  $\mathbf{C}$  at 100,000 points in the design space, which are randomly sampled using the Monte Carlo method, and averages them as  $\bar{\mathbf{C}}$ . Second, an eigenvalue decomposition is performed on  $\bar{\mathbf{C}}$  as

$$\bar{\mathbf{C}} = \mathbf{W}\Lambda\mathbf{W}^T, \tag{19}$$

where  $\mathbf{W} = (\mathbf{w}_1 \cdots \mathbf{w}_m)$  are the eigenvectors, which represent the suitable coordinates and  $\Lambda = \text{diag}(\lambda_1 \cdots \lambda_m)$  is the eigenvalue matrix. Third, the design variable vector in the new coordinate system  $\mathbf{y}$  is calculated from the original vector  $\mathbf{x}$  as

$$\mathbf{y} = \mathbf{W}^T\mathbf{x}. \tag{20}$$

KCTG is constructed in the new coordinate system  $\mathbf{y}$  in the same way as OK except its coordinate system.

### 3 Application to test functions

#### 3.1 Test problem definition

KCTL, KCTG, and OK are applied to the following seven types (five of them are widely employed [7,21–23]) of rotated and un-rotated (original) test functions with various dimensions:

Ellipsoid function:

$$f(\mathbf{x}) = \sum_{i=1}^m (i \cdot y_i)^2; \tag{21}$$

Rosenbrock function:

$$f(\mathbf{x}) = \sum_{i=1}^{m-1} \left( (1 - y_i)^2 + 100 (y_{i+1} - y_i^2)^2 \right); \tag{22}$$

Branin function:

$$f(x_1, x_2) = \left( y_2 - \frac{5.1}{4\pi^2} y_1^2 + \frac{5}{\pi} y_1 - 6 \right)^2 + 10 \left( 1 - \frac{1}{8\pi} \right) \cos y_1 + 10; \tag{23}$$

Translated Hartmann function ( $m = 3$ ):

$$f(x_1, x_2, x_3) = - \sum_{i=1}^4 \alpha_i \exp \left[ - \sum_{j=1}^3 A_{ij} (y_j - (P_{ij} - 0.5))^2 \right], \tag{24}$$

where  $\alpha = (1.0, 1.2, 3.0, 3.2)^T$

$$A = \begin{bmatrix} 3.0 & 10 & 30 \\ 0.1 & 10 & 35 \\ 3.0 & 10 & 30 \\ 0.1 & 10 & 35 \end{bmatrix}, \quad P = \begin{bmatrix} 0.3689 & 0.1170 & 0.2673 \\ 0.4699 & 0.4387 & 0.7470 \\ 0.1091 & 0.8732 & 0.5547 \\ 0.0381 & 0.5743 & 0.8828 \end{bmatrix};$$

Translated Hartmann function ( $m = 6$ ):

$$f(x_1, x_2, x_3, x_4, x_5, x_6) = - \sum_{i=1}^4 \alpha_i \exp \left[ - \sum_{j=1}^6 A_{ij} (y_j - (P_{ij} - 0.5))^2 \right], \tag{25}$$



where  $\alpha = (1.0, 1.2, 3.0, 3.2)^T$

$$A = \begin{bmatrix} 10 & 3 & 17 & 3.5 & 1.7 & 8 \\ 0.05 & 10 & 17 & 0.1 & 8 & 14 \\ 3 & 3.5 & 1.7 & 10 & 17 & 8 \\ 17 & 8 & 0.05 & 10 & 0.1 & 14 \end{bmatrix}$$

$$P = \begin{bmatrix} 0.1312 & 0.1696 & 0.5569 & 0.0124 & 0.8283 & 0.5886 \\ 0.2329 & 0.4135 & 0.8307 & 0.3736 & 0.1004 & 0.9991 \\ 0.2348 & 0.1451 & 0.3522 & 0.2883 & 0.3047 & 0.6650 \\ 0.4047 & 0.8828 & 0.8732 & 0.5743 & 0.1091 & 0.0381 \end{bmatrix}$$

Sine function:

$$f(x_1, x_2) = \sin(2\pi y_1) \sin(2\pi M y_2); \tag{26}$$

where  $M = 1, 2, 4$

Parabolic function:

$$f(\mathbf{x}) = y_1^2; \tag{27}$$

The Hartmann functions are translated from the original domain  $[0, 1]$  to  $[-0.5, 0.5]$  because the functions are rotated around the origin of the coordinate system. All of them are minimization problems and functions of  $m$  design variables  $\mathbf{x}$  ( $m = 2$  for the Branin function, the six-hump camel function, and the sine function and  $m = 3, 6$  for the Hartmann functions) through intermediate variables  $\mathbf{y}$  in the rotated coordinate system, as shown in Eq. (15) when  $m = 3$ . All rotation angles are set to  $\phi_i = 30^\circ$  ( $i = 1, \dots, m$ ) when rotated functions are employed. Experimental conditions are summarized in Table 1, and the shapes of the original and rotated functions with  $m = 2$  are shown in Fig. 4. Before comparing three models according to the EGO framework, they are compared at fixed numbers of sample points in the ranges described in Table 1. Sample points are generated by LHS and the number of samples depends on the non-linearity (dimension and order of polynomial) of test functions.

Subsequently, EGO with three models is conducted in multidimensional ellipsoid and Rosenbrock functions. Initial sample points are generated by LHS and additional sample points are employed one after another at the location where the EI value calculated by Eq. (13) becomes a maximum. The number of initial and additional samples is shown in Table 2. The numbers of population and generation in NSGA-II to maximize the EI value are 500 and 100, respectively. The EI value computed as a double-precision floating point number has the potential to become zero when test functions with low non-linearity are approximated with excessive numbers of sample points. If NSGA-II fails to find a candidate of an additional sample point due to this situation, NSGA-II maximizes mean square error defined by Eq. (11) and employs it as an additional sample points.

To compare the accuracy of the three models, the following root mean square error (RMSE) between the surrogate model and the real function is calculated at uniformly-distributed  $N \approx 10000$  validation points.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (f(\mathbf{x}^i) - \hat{f}(\mathbf{x}^i))^2}. \tag{28}$$

One hundred independent trials starting with different initial sample points are performed and the averaged RMSE is evaluated for comparison. The optimal solutions obtained by each model are also compared by averaged values in 100 trials when EGO is conducted.

**Table 1** Experimental conditions

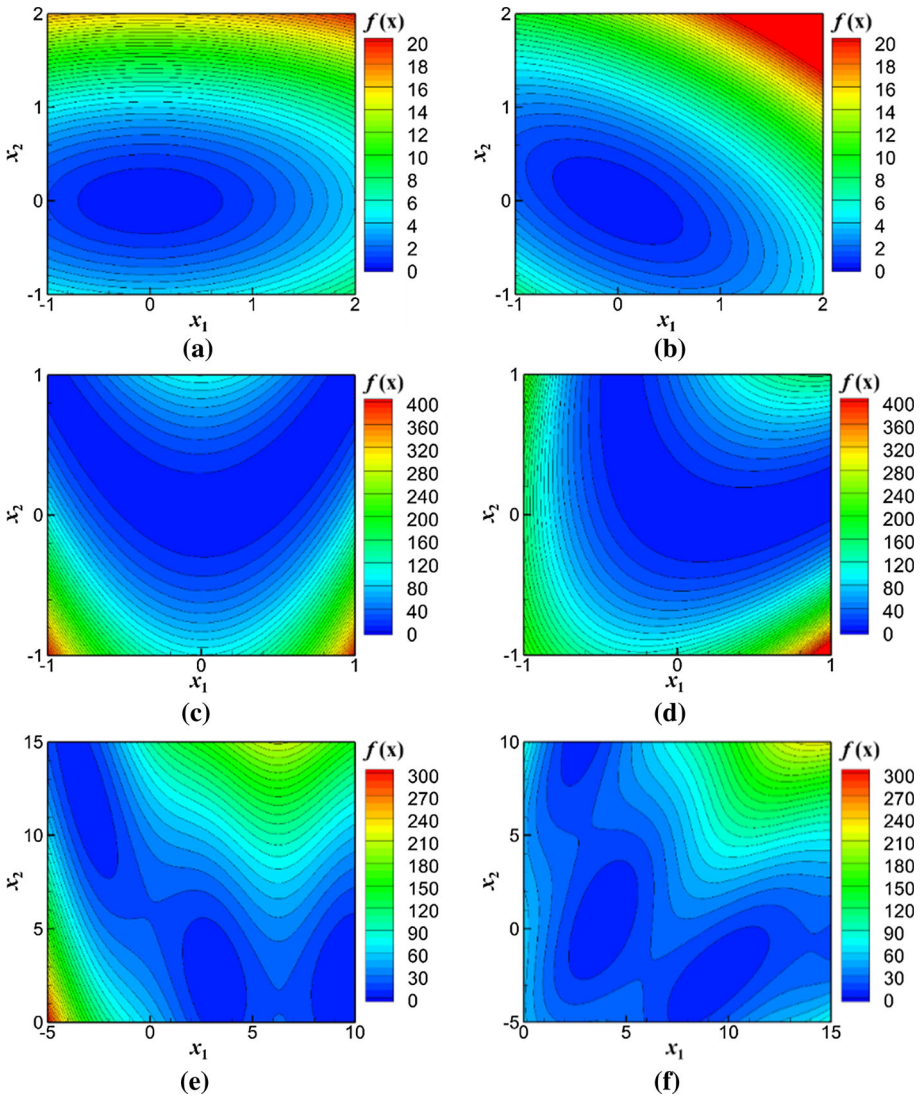
Test functions	$m$	Rotation	Abbreviation	Domain	Minimum samples	Maximum samples
Ellipsoid	2	No	E2	$[-1, 2]^2$	10	30
	2	Yes	E2R	$[-1, 2]^2$	10	30
	3	Yes	E3R	$[-1, 2]^3$	10	50
	4	Yes	E4R	$[-1, 2]^4$	20	60
	5	Yes	E5R	$[-1, 2]^5$	20	100
	10	Yes	E10R	$[-1, 2]^{10}$	30	150
Rosenbrock	2	No	R2	$[-1, 1]^2$	10	30
	2	Yes	R2R	$[-1, 1]^2$	10	30
	3	Yes	R3R	$[-1, 1]^3$	30	90
	4	Yes	R4R	$[-1, 1]^4$	50	150
	5	Yes	R5R	$[-1, 1]^5$	80	200
Branin	2	No	B2	$[-5, 10] \times [0, 15]$	10	50
	2	Yes	B2R	$[0, 15] \times [-5, 10]$	10	50
Translated Hartmann	3	No	H3	$[-0.5, 0.5]^3$	20	60
	3	Yes	H3R	$[-0.5, 0.5]^3$	20	60
	6	No	H6	$[-0.5, 0.5]^6$	20	100
	6	Yes	H6R	$[-0.5, 0.5]^6$	20	100
Sine ( $M = 1$ )	2	Yes	S2R1	$[-0.5, 0.5]^2$	10	50
Sine ( $M = 2$ )	2	Yes	S2R2	$[-0.5, 0.5]^2$	15	75
Sine ( $M = 4$ )	2	Yes	S2R4	$[-0.5, 0.5]^2$	20	100
Parabolic	10	Yes	P10R	$[-1, 2]^{10}$	20	100

## 3.2 Results and discussion for test functions

### 3.2.1 Comparison between original and rotated test functions

E2, R2, B2, H3, H6, and their rotated versions are approximated by KCTL, KCTG, and OK to compare their model accuracy for rotated and un-rotated functions. Figure 5 show the averaged histories of RMSE. Table 3 summarizes statistics (mean, standard deviation, best, and worst) of RMSE in 100 trials when numbers of sample points are intermediate values between minimum and maximum presented in Table 1.

To begin, we focus on E2 and E2R which are the simplest functions in the test functions. KCTL achieves the lowest RMSE at any number of sample points in both functions because it can immediately identify the suitable coordinate system corresponding to the exact rotation angle. KCTG can also identify the suitable coordinate although KCTG and OK show comparable RMSEs in E2 and E2R. Both KCTL and KCTG can identify the suitable coordinate but only KCTL shows extremely high accuracy at 10 sample points in both functions even though OK is constructed with the exact coordinate system in E2. This is because only KCTL can tune the rotation angle to increase the likelihood function. The Kriging model approximates a function as a superposition of Gaussian functions whose centers correspond



**Fig. 4** Shapes of the two-dimensional test functions. **a** E2, **b** E2R, **c** R2, **d** R2R, **e** B2 and **f** B2R

to the locations of sample points. Thus, slightly changing the rotation angle from the exact angle sometimes improves the model accuracy if a few sample points can be used.

In R2 and R2R, RMSE of KCTL decreases rapidly at around 15 sample points whereas KCTL cannot identify the suitable coordinate system with a few sample points. Note that KCTL should be employed with a sufficient number of initial sample points to guarantee its model accuracy. The rotation angles in KCTL correspond to the exact angles ( $0^\circ$  for R2 and  $30^\circ$  for R2R) when a sufficient number of sample points are employed. Additionally, transformed coordinate system of KCTG has converged toward  $0^\circ$  in R2 because R2 roughly symmetric about  $x_1 = 0$ , which leads the off-diagonal elements in  $\bar{C}$  to zero. In approximation of R2R, it converges toward  $40^\circ$  which is slightly different from the exact rotation angle

**Table 2** Experimental conditions for EGO

Test functions	$m$	Rotation	Abbreviation	Domain	Initial samples	Additional samples
Ellipsoid	2	Yes	E2R	$[-1, 2]^2$	10	20
	3	Yes	E3R	$[-1, 2]^3$	15	30
	4	Yes	E4R	$[-1, 2]^4$	20	40
	5	Yes	E5R	$[-1, 2]^5$	25	50
	10	Yes	E10R	$[-1, 2]^{10}$	30	120
Rosenbrock	2	Yes	R2R	$[-1, 1]^2$	10	20
	3	Yes	R3R	$[-1, 1]^3$	30	60
	4	Yes	R4R	$[-1, 1]^4$	50	100
	5	Yes	R5R	$[-1, 1]^5$	70	130

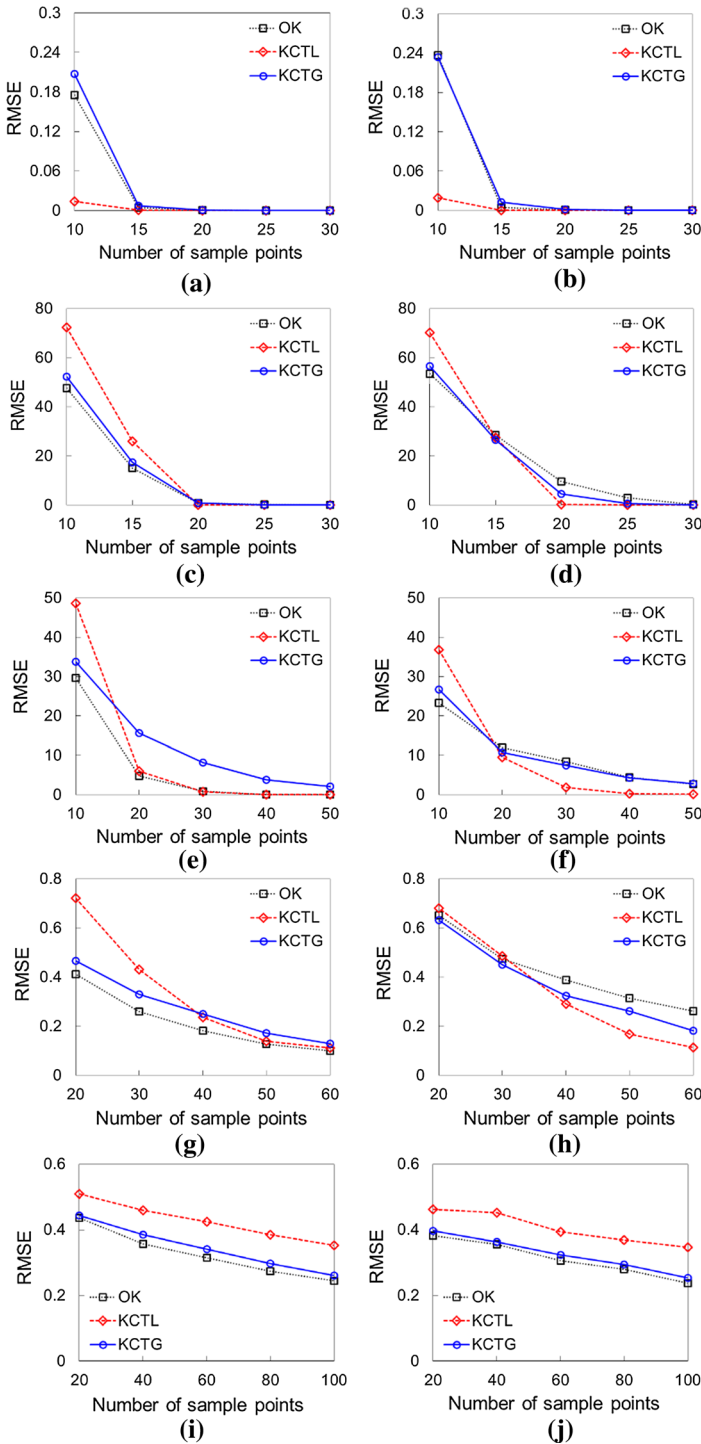
due to the asymmetric property of R2R and its large gradient around  $(x_1, x_2) = (1, -1)$ . Nevertheless, KCTG has better coordinate system than OK and shows lower RMSE in Fig. 5d.

KCTL and KCTG can approximate B2R with the same accuracy as their approximation of B2: the histories of RMSE and the objective function value for B2 and B2R are virtually identical when more than 20 sample points are employed. It is applied to E2–E2R, R2–R2R, and H3–H3R as well. The coordinate transformation enables the Kriging models to approximate functions independent of the arbitrarily defined coordinate system (design variables). Thus, KCTL and KCTG are applicable to the optimization problems with design variables which do not correlate directly with the objective functions and have strong correlation among themselves (spline curves and FFD). High RMSEs of KCTG in B2 and B2R are derived from the misled coordinate system. B2 has three minimal points and two of them (left two in Fig. 4e) constitute a large valley along which KCTG makes the coordinate system rotate although the suitable coordinate system found by KCTL is the original one ( $x_1$  and  $x_2$ ). Additionally, KCTG and OK have the same RMSE in approximation of B2R because B2R also has this valley roughly along  $x_2$  axis.

The suitable coordinate systems of H3 and H3R correspond to their exact rotation angles because they consist of a superposition of four Gaussian functions. Therefore, RMSEs of KCTL and KCTG are lower than those of OK if sufficient sample points are employed.

H6 and H6R should have the same trend as H3 and H6R. However KCTL has extremely high RMSE in these functions because of their high dimensionality which makes it difficult to find the suitable coordinate system; maximizing likelihood function becomes a 21-dimensional optimization problem in KCTL when  $m = 6$ . This optimization itself is solved appropriately, but the likelihood function fails to lead KCTL to a suitable coordinate system in these cases. Instead of KCTL, KCTG should be employed to approximate the functions with more than five correlated design variables. KCTG shows the same trend as OK for not only H6 but also H6R, although KCTG is expected to show higher accuracy than OK in H6R. H3 and H3R have large gradients along the first design variable in all four Gaussian functions constructing them, as the first column has lowest values in each row of matrix  $A$  in Eq. (24). In contrast, the steepest directions of the four Gaussian functions constructing H6 and H6R are not the same [matrix  $A$  in Eq. (25)], which makes KCTG fail in finding a suitable coordinate system.

The statistics of RMSE are strongly correlated with mean values shown in Fig. 5. KCTL mostly achieves the lowest statistics in the functions where KCTL has the lowest mean RMSE

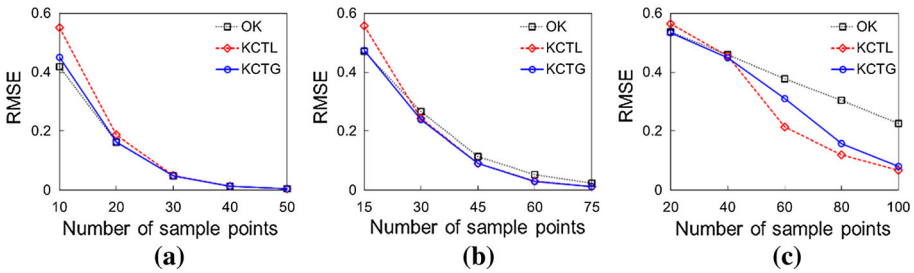


**Fig. 5** Averaged histories of RMSE between original and rotated test functions. **a** E2, **b** E2R, **c** R2, **d** R2R, **e** B2, **f** B2R, **g** H3, **h** H3R, **i** H6 and **j** H6R

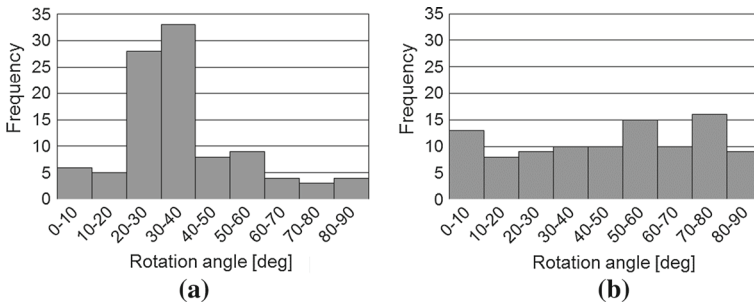
**Table 3** Statistics of RMSE for original and rotated test functions

Test function	Samples	Statistics	OK	KCTL	KCTG
E2	20	Mean	<i>9.840E-04</i>	<b>2.095E-04</b>	7.196E-04
		Std. dev.	<i>1.773E-03</i>	<b>3.049E-04</b>	8.811E-04
		Best	1.759E-05	<b>7.350E-06</b>	<i>1.876E-05</i>
		Worst	<i>7.903E-03</i>	<b>1.833E-03</b>	4.685E-03
E2R	20	Mean	9.346E-04	<b>1.665E-04</b>	<i>1.904E-03</i>
		Std. dev.	1.730E-03	<b>2.611E-04</b>	<i>5.858E-03</i>
		Best	<i>1.620E-05</i>	<b>1.131E-05</b>	1.210E-05
		Worst	1.074E-02	<b>1.890E-03</b>	<i>5.296E-02</i>
R2	20	Mean	7.708E-01	<b>5.743E-02</b>	<i>8.748E-01</i>
		Std. dev.	8.917E-01	<b>7.672E-02</b>	<i>8.954E-01</i>
		Best	1.629E-02	<b>3.970E-03</b>	<i>6.869E-02</i>
		Worst	4.409E+00	<b>5.301E-01</b>	<i>4.503E+00</i>
R2R	20	Mean	<i>9.707E+00</i>	<b>1.968E-01</b>	4.510E+00
		Std. dev.	3.690E+00	<b>5.503E-01</b>	<i>3.711E+00</i>
		Best	<i>2.958E+00</i>	<b>7.379E-03</b>	1.163E-01
		Worst	<i>1.958E+01</i>	<b>3.786E+00</b>	1.734E+01
B2	30	Mean	9.069E-01	<b>8.045E-01</b>	<i>8.239E+00</i>
		Std. dev.	6.132E-01	<b>5.913E-01</b>	<i>4.054E+00</i>
		Best	<b>1.511E-01</b>	1.552E-01	<i>2.951E+00</i>
		Worst	3.944E+00	<b>3.826E+00</b>	<i>2.165E+01</i>
B2R	30	Mean	<i>8.392E+00</i>	<b>1.907E+00</b>	7.441E+00
		Std. dev.	2.882E+00	<b>2.399E+00</b>	<i>3.631E+00</i>
		Best	<i>3.643E+00</i>	<b>2.214E-01</b>	6.462E-01
		Worst	<b>1.731E+01</b>	1.925E+01	<i>2.022E+01</i>
H3	40	Mean	<b>1.814E-01</b>	2.355E-01	<i>2.503E-01</i>
		Std. dev.	<b>5.889E-02</b>	<i>9.703E-02</i>	7.398E-02
		Best	<b>9.753E-02</b>	1.190E-01	<i>1.289E-01</i>
		Worst	<b>4.226E-01</b>	5.024E-01	<i>5.050E-01</i>
H3R	40	Mean	<i>3.890E-01</i>	<b>2.907E-01</b>	3.249E-01
		Std. dev.	<b>7.134E-02</b>	<i>1.232E-01</i>	8.890E-02
		Best	<i>2.456E-01</i>	<b>1.403E-01</b>	1.654E-01
		Worst	6.533E-01	<i>6.729E-01</i>	<b>6.011E-01</b>
H6	60	Mean	<b>3.144E-01</b>	<i>4.234E-01</i>	3.410E-01
		Std. dev.	<i>5.325E-02</i>	5.324E-02	<b>4.805E-02</b>
		Best	<b>2.249E-01</b>	<i>3.238E-01</i>	2.444E-01
		Worst	<b>4.391E-01</b>	<i>6.284E-01</i>	4.568E-01
H6R	60	Mean	<b>3.052E-01</b>	<i>3.940E-01</i>	3.239E-01
		Std. dev.	<b>4.137E-02</b>	<i>5.028E-02</i>	4.652E-02
		Best	<b>2.341E-01</b>	<i>2.971E-01</i>	2.404E-01
		Worst	<b>4.538E-01</b>	<i>5.168E-01</i>	4.550E-01

Bold face: lowest value, italic face: highest value



**Fig. 6** Averaged histories of RMSE among the sine functions. **a** S2R1, **b** S2R2 and **c** S2R4



**Fig. 7** Rotation angle histograms for S2R1 with 50 sample points. **a** KCTL and **b** KCTG

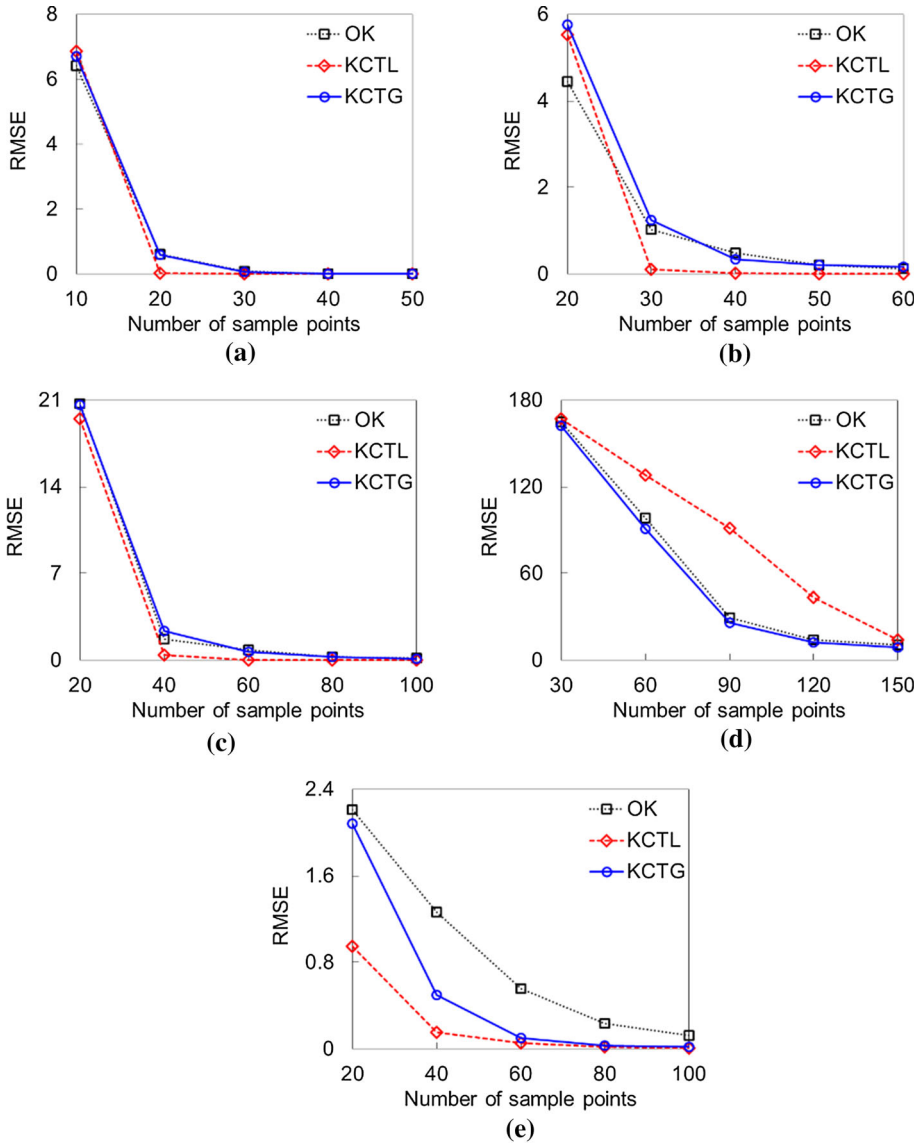
among three models (E2, E2R, R2, R2R, B2, and B2R). By contrast, KCTL has the highest values in most statistics in H6 and H6R where approximation accuracy of KCTL is poor.

### 3.2.2 Comparison among the sine functions with different wavelengths

Three methods are applied to the sine functions to investigate effects of wavelength difference between two directions because shorter wavelength with higher gradient means stronger contribution to an objective function. The averaged histories of RMSE for each function are shown in Fig. 6. Compared to OK, both KCTL and KCTG cannot improve approximation accuracy in S2R1 where wavelengths in two directions are the same. Obviously, KCTG is not suitable for this function because gradients along any directions are the same. On the other hand, KCTL frequently find out the suitable coordinates although its approximation accuracy is not improved. Figure 7 shows rotation angle histograms for S2R1 with 50 sample points. The histogram of KCTL has a peak around 30° while that of KCTG is uniformly distributed. Hence, KCTL can identify the suitable coordinates even though gradients have no trend along any directions. In S2R2 and S2R4, both KCTL and KCTG find out the suitable coordinates and accomplish better approximation accuracy than OK. Their accuracy improvement are emphasized as the wavelength difference becomes large.

### 3.2.3 Comparison among multi-dimensional functions

A multi-dimensional rotated ellipsoid and parabolic functions are employed to investigate the capability of KCTL and KCTG for relatively high-dimensional optimization problems. Figure 8 shows the averaged histories of RMSE. KCTL shows the best accuracy except for E10R. Note that KCTL successfully approximates objective functions with the number of



**Fig. 8** Averaged histories of RMSE for multi-dimensional functions. **a** E3R, **b** E4R, **c** E5R, **d** E10R and **e** P10R

effective correlated design variables no greater than five. Both E10R and P10R includes 10 design variables though KCTL behaves in different manners. E10R is a function of all translated design variables (10 effective variables) though P10R is a function of only  $y_1$  in the translated coordinate system.

KCTG is applicable to high-dimensional functions with many effective variables such as E10R. Comparing E5R and E10R, the difference in RMSE between KCTG and OK is clearer in E10R. KCTG is more suitable for real-world optimization problems with correlated design variables than KCTL. This is because spline curves and FFD usually include 10–300 control



points as design variables [24,25], and more than five variables may be effective even in the translated coordinate system.

### 3.2.4 Comparison in efficient global optimization

Multi-dimensional rotated ellipsoid and Rosenbrock functions are minimized by EGO with KCTL, KCTG, and OK to compare their exploration capability for rotated functions. Figure 9 shows the averaged histories of the objective function value of optimal solution in the sample points. KCTL shows the best performance in each optimization except for E10R. Thus, KCTL not only approximates objective functions more accurately but also optimizes them more efficiently when the number of correlated design variables is no greater than five.

KCTG also finds optimal solutions with less sample points than OK in most functions including E10R although the optimal solution obtained by OK sometimes exceeds that obtained by KCTG when all additional sample points are employed. KCTG finds the quasi-optimal solution around the exact optimal solution in the early stages of the EGO process. It prohibits KCTG from exploring the exact optimal solution to exploit another domain of the design space because EGO is based on two principles: exploration and exploitation.

## 3.3 Application to an airfoil design problem

The results in Sect. 3 show that KCTG approximates the function accurately regardless of the coordinate system and the dimension and has an advantage in the optimization with correlated design variables. However, the test functions in Sect. 3 include simple ones and roughly symmetric ones. We must evaluate the practicality of KCTG through a real-world shape design optimization problem, which includes correlated design variables for shape representation such as the control points of spline curves. In Sect. 4, an airfoil design optimization is considered with KCTG and OK to investigate the effects of the coordinate transformation in a real-world problem.

### 3.4 Design problem definition

The objective function and the constraints in the airfoil design problem are defined as follows:

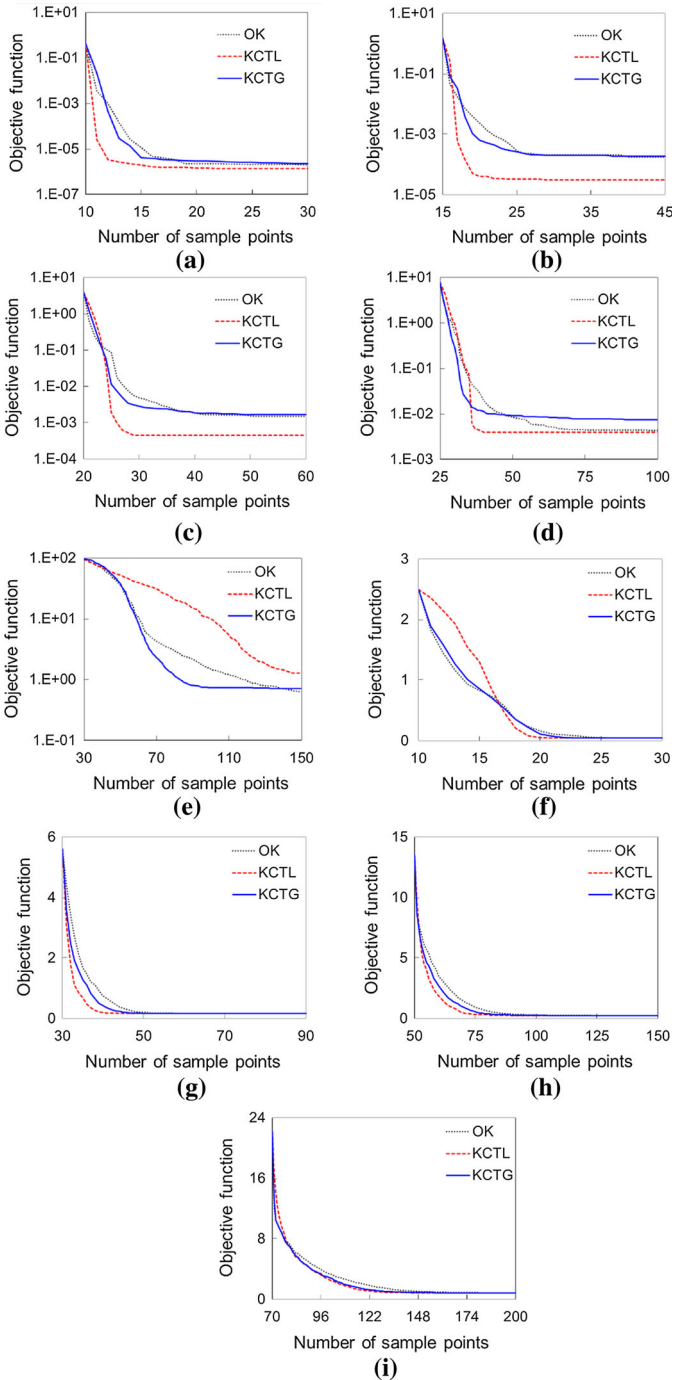
$$\text{Maximize } L/D, \quad (29)$$

$$\text{subject to } C_m \geq -0.1535, \quad (30)$$

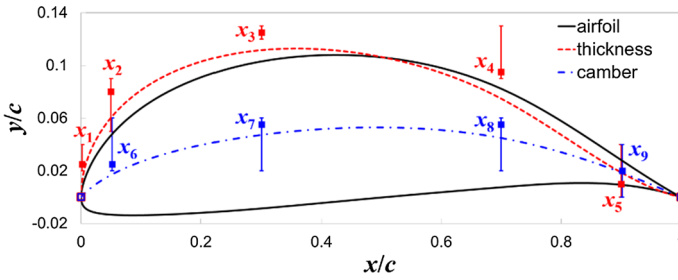
$$t_{\max}/c \geq 0.11, \quad (31)$$

at the angle of attack  $\alpha = 4^\circ$  and the Reynolds number  $\text{Re} = 5 \times 10^5$ .  $L/D$  and  $C_m$  denote the lift-drag ratio and the pitching moment coefficient, respectively.  $t_{\max}$  is the maximum thickness of the airfoil and  $c$  is the chord length.  $L/D$  and  $C_m$  at the sample points are evaluated by a subsonic flow solver “XFOIL” [26] which calculates incompressible viscous flow in this study. Hence, KCTG and OK are used to estimate these two values whereas  $t_{\max}/c$  is calculated directly by representing the airfoil. Generally, computational times for XFOIL are less than one second for one flow condition. This study employs XFOIL to achieve many independent trials of airfoil design optimization and evaluate statistics of the results properly. The constraint values in Eqs. (30) and (31) correspond to those of DAE31 airfoil whose  $L/D$  is 138.6.

The nine design variables correspond to the locations of nine control points for two non-uniform rational basis spline curves defining the airfoil thickness distribution and camber



**Fig. 9** Averaged histories of the objective function value. **a** E2R, **b** E3R, **c** E4R, **d** E5R, **e** E10R, **f** R2R, **g** R3R, **h** R4R and **i** R5R



**Fig. 10** Design variables and their ranges

line in Fig. 10. The five red dots and four blue dots are the control points for thickness and camber, respectively; the red and blue bars show the ranges of each design variable. Only  $x_3$  has a relatively small range to help the maximum thickness meet the constraint in Eq. (31). Each range is shown as follows:

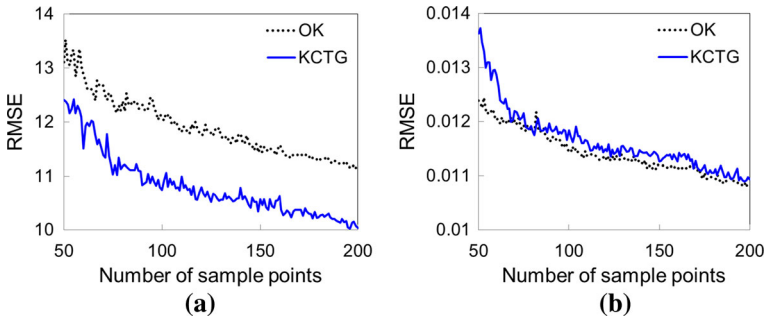
$$\begin{aligned}
 &0.02 \leq x_1 \leq 0.04, \quad 0.05 \leq x_2 \leq 0.09, \quad 0.12 \leq x_3 \leq 0.13, \\
 &0.09 \leq x_4 \leq 0.13, \quad 0.00 \leq x_5 \leq 0.04, \quad 0.02 \leq x_6 \leq 0.06, \\
 &0.02 \leq x_7 \leq 0.06, \quad 0.02 \leq x_8 \leq 0.06, \quad 0.00 \leq x_9 \leq 0.04.
 \end{aligned}
 \tag{32}$$

Fifty initial sample points are generated by LHS and 150 additional sample points are employed one after another at the location where the  $E_c I$  value calculated by Eq. (14) becomes a maximum. KCTG and OK are compared by optimal solutions obtained by EGO with each model and RMSE in Eq. (28) where validation points are generated by LHS and  $N = 10, 000$ . Seventy independent trials are performed and their averaged  $L/D$ ,  $C_m$ , and RMSE are evaluated. The numbers of population and generation in NSGA-II are 500 and 100, respectively, as for the case in Sect. 3.

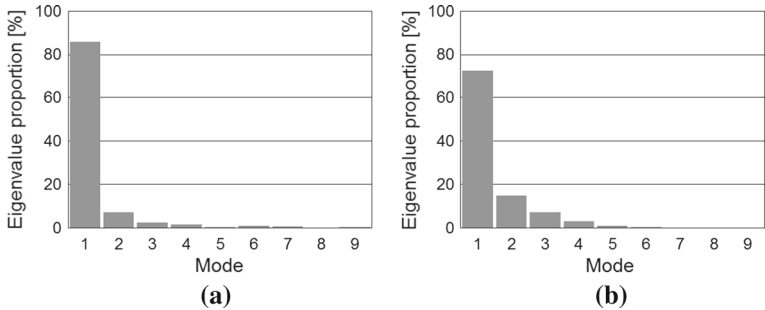
Depending on the airfoil shape, the flow computation with XFOIL sometimes does not converge. An alternative initial sample point and validation point are randomly selected from the entire design space if convergence is not achieved. The same treatment is applied if an initial sample point and a validation point do not meet at least one of the constraints.

### 3.5 Results and discussion for airfoil design problem

Figure 11 shows the averaged histories of RMSE for  $L/D$  and  $C_m$ . KCTG reduces RMSE of  $L/D$  compared with that for OK at any number of sample points in EGO process. Therefore, it is shown that KCTG can approximate the function for a real-world problem using spline curves as design variables more accurately than OK. Regarding  $C_m$ , OK has lower RMSE than KCTG although both models converge along similar trends with increasing sample points. From aerodynamic theory,  $C_m$  is regarded as a function that depends on the camber line and is not affected by the thickness, i.e., the effective number of design variables for  $C_m$  is almost 4. OK does not consider the correlation between camber and thickness, which enables OK to easily ignore the variables related with the thickness by decreasing the weight coefficients in Eq. (2). KCTG can also ignore these variables although the coordinate transformation may disturb it in an early stage of EGO.



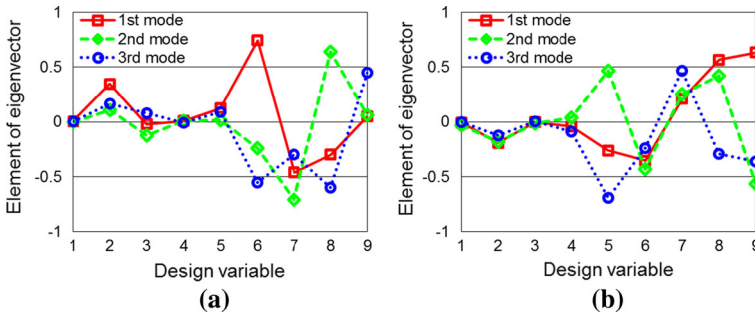
**Fig. 11** Averaged histories of RMSE for airfoil design problem. **a**  $L/D$  and **b**  $C_m$



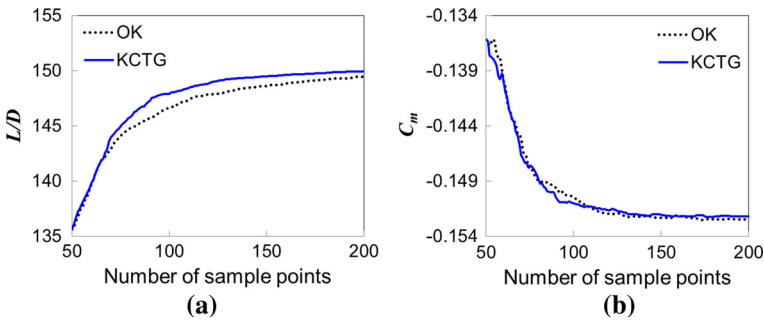
**Fig. 12** Eigenvalue proportion for each mode in KCTG. **a**  $L/D$  and **b**  $C_m$

Approximation accuracy is strongly related to the transformed coordinate system. Figure 12 shows eigenvalue proportions for each mode (eigenvector), which correspond to the squared gradients of the objective and constraint functions estimated by KCTG along each transformed coordinate. The estimated gradient of  $L/D$  is mostly occupied by the first and second modes whose eigenvalue proportions are about 95% in total whereas 95% is accomplished with the first three modes in  $C_m$ .  $L/D$  has a suitable coordinate system representing the significant trend of the function and KCTG successfully identifies this coordinate system. In contrast, KCTG fails to find a suitable coordinate system in the approximation of  $C_m$ , or  $C_m$  originally does not have such a suitable coordinate system. As a result, eigenvalue proportions are more widely dispersed in  $C_m$  than  $L/D$ , and KCTG and OK have comparable RMSEs. Additionally, Fig. 13 shows eigenvectors of the first three modes, which indicate design variables related to the airfoil’s camber line ( $x_6-x_9$ ) are dominant in both  $L/D$  and  $C_m$ .  $C_m$  is also sensitive to the airfoil’s thickness at the trailing edge ( $x_5$ ) because the elements of the eigenvectors for  $x_5$  have large values in the second and third modes. This makes KCTG even more difficult to approximate  $C_m$ .

Averaged histories of  $L/D$  and  $C_m$  at the optimal solution among feasible sample points are shown in Fig. 14. KCTG obtains better solutions than OK on average when the number of sample points is over 67.  $C_m$  values are comparable between two models. Therefore, it is suggested that KCTG has an advantage over OK not only in approximation accuracy but also in the ability to explore the optimal solution if design variables are correlated with each other.



**Fig. 13** Eigenvectors for each mode in KCTG. **a**  $L/D$  and **b**  $C_m$



**Fig. 14** Averaged histories of the objective function and constraint values at the optimal solution for airfoil design problem. **a**  $L/D$  and **b**  $C_m$

### 4 Conclusions

The Kriging models with coordinate transformation based on likelihood and gradients were proposed and validated in various test functions and an airfoil design problem with correlated design variables. A suitable coordinate system was identified by maximizing likelihood function or by finding the eigenvalues of the covariance matrix of the estimated objective function gradients along each design variables. Averaged RMSEs between surrogate models and the real objective function and the optimal solutions obtained by efficient global optimization were used to evaluate the practicality of the proposed methods.

In the application to test functions, the proposed methods approximated the entire function shape more accurately than the conventional method if design variables were correlated with each other and a sufficient number of sample points were employed. The proposed methods also showed greater comparative accuracy to the conventional method even if the correlation between design variables was not strong. The capability to explore the optimal solution was also better than the conventional method if the proposed method achieved higher approximation accuracy. Additionally, maximizing the likelihood function in the Kriging model was preferable in finding a suitable coordinate system for functions with correlated design variables although this method did not function well when the number of design variables was greater than five. The coordinate transformation based on estimated gradients was available for functions with more than five design variables. This method was more suited to real-world optimization problems with correlated design variables than the former because real-world problems usually include 10–1000 design variables.

Control points of the non-uniform rational basis spline curves defining an airfoil's thickness distribution and camber line were employed as the correlated design variables in the airfoil design optimization. The latter proposed method approximated the objective function (lift–drag ratio) more accurately and found better solutions than the conventional method although the constraint function (pitching moment coefficient) was difficult to approximate by the proposed method. Therefore, the proposed method revealed usefulness in real-world optimization problem with correlated design variables.

In this study, optimization problems with up to ten design variables were adopted. The number is relatively smaller than the usual number of design variables in real-world optimization problems which use spline curves and free form deformation for shape definition. Moreover, these real-world problems tend to have more than two objective functions. Therefore, in the future, it is desirable that the proposed method is validated in the multi-objective optimization problems with more design variables.

**Acknowledgements** This work was supported by JSPS KAKENHI 14J07397 through JSPS Research Fellowships for Young Scientists.

## References

1. Queipo, N.V., Haftka, R.T., Shyy, W., Goel, T., Vaidyanathan, R., Tucker, P.K.: Surrogate-based analysis and optimization. *Progr. Aerosp. Sci.* **41**, 1–28 (2005)
2. Namura, N., Obayashi, S., Jeong, S.: Efficient global optimization of vortex generators on a super critical infinite-wing using Kriging-based surrogate models. In: 52nd AIAA Aerospace Sciences Meeting, AIAA-2014-0904. National Harbor (2014)
3. Forrester, A.I.J., Keane, A.J.: Recent advances in surrogate-based optimization. *Progr. Aerosp. Sci.* **45**, 50–79 (2009)
4. Regis, R.G., Shoemaker, A.S.: Local function approximation in evolutionary algorithms for the optimization of costly functions. *IEEE Trans. Evol. Comput.* **8**, 490–505 (2004)
5. Broomhead, D.S., Lowe, D.: Multivariate functional interpolation and adaptive networks. *Complex Syst.* **2**, 321–355 (1988)
6. Matheron, G.: Principles of geostatistics. *Econ. Geol.* **58**, 1246–1266 (1963)
7. Jones, D.R., Schonlau, M., Welch, W.J.: Efficient global optimization of expensive black-box function. *J. Global Optim.* **13**, 455–492 (1998)
8. Joseph, V.R., Hung, Y., Sudjianto, A.: Blind kriging: a new method for developing metamodels. *ASME J. Mech. Des.* **130**, 031102-1–031102-8 (2008)
9. Namura, N., Shimoyama, K., Jeong, S., Obayashi, S.: Kriging/RBF-hybrid response surface methodology for highly nonlinear functions. *J. Comput. Sci. Technol.* **6**, 81–96 (2012)
10. Matheron, G.: The Intrinsic random functions and their applications. *Adv. Appl. Probab.* **5**, 439–468 (1973)
11. Mullur, A.A., Messac, A.: Extended radial basis functions: more flexible and effective metamodeling. In: 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization. Conference, AIAA-2004-4573. Albany (2004)
12. Xiong, Y., Chen, W., Apley, D., Ding, X.: A non-stationary covariance-based Kriging method for metamodeling in engineering design. *Int. J. Numer. Methods Eng.* **71**, 733–756 (2007)
13. Samareh, J.A.: Aerodynamic shape optimization based on free-form deformation. In: 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, AIAA-2004-4630, Albany (2004)
14. McKay, M.D., Beckman, R.J., Conover, W.J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239–245 (1979)
15. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evolut. Comput.* **6**, 182–197 (2002)
16. Jeong, S., Yamamoto, K., Obayashi, S.: Kriging-based probabilistic method for constrained multi-objective optimization problem. In: AIAA 1st Intelligent Systems Technical Conference, AIAA-2004-6437, Chicago (2004)
17. Karaboga, D.: An idea based on honey bee swarm for numerical optimization. In: Technical Report-TR06, Erciyes University, Engineering Faculty, Computer Engineering Department. (2005)

18. Karaboga, D., Akay, B.: A comparative study of artificial Bee colony algorithm. *Appl. Math. Comput.* **214**, 108–132 (2009)
19. Constantine, P.G., Dow, E., Wang, Q.: Active subspace methods in theory and practice: applications to Kriging surfaces. *SIAM J. Sci. Comput.* **36**, A1500–A1524 (2014)
20. Lukaczyk, T., Palacios, F., Alonso, J.J.: Active Subspaces for Shape Optimization. In: 10th AIAA Multidisciplinary Design Optimization Conference, AIAA-2014-1171. National Harbor (2014)
21. Rosenbrock, H.H.: An automatic method for finding the greatest or least value of a function. *Comput. J.* **3**, 175–184 (1960)
22. Liu, H., Xu, S., Ma, Y., Wang, X.: Global optimization of expensive black box functions using potential Lipschitz constants and response surfaces. *J. Global Optim.* **63**, 229–251 (2015)
23. Zhang, J., Chowdhury, S., Messac, A.: An adaptive hybrid surrogate mode. *Struct. Multidiscip. Optim.* **46**, 223–238 (2012)
24. Lyu, Z., Martins, J.R.R.A.: Aerodynamic design optimization studies of a blended-wing-body aircraft. *J. Aircr.* **51**, 1604–1617 (2014)
25. Palacios, F., Economon, T.D., Wendorff, A.D., Alonso, J.J.: Large-scale aircraft design using SU2. In: 53rd AIAA Aerospace Sciences Meeting, AIAA-2015-1946. Kissimmee (2015)
26. Drela, M.: XFOIL: An analysis and design system for low Reynolds number airfoils. In: *Low Reynolds Number Aerodynamics*, Lecture Notes in Engineering, 54, pp. 1–12. Springer-Verlag, New York (1989)