A kriging-based sequential optimization algorithm with hybrid infill sampling strategy

DOI : 10.36909/jer.ICCSCT.19469

Xiaojing, Wang[1], Yaohui Li[2, 3*], Qilu Liu[3], Jingfang Shen[3], Zebin Zhang[4]

[1]Department of mechanical and electrical engineering, Henan Light Industry Vocational College, Zhengzhou, Henan, 450006, China;
[2]College of Mechanical and Electrical Engineering, Xuchang University, Xuchang, 461000, Henan, China;
[3]College of Science, Huazhong Agricultural University, 430070 Wuhan, China;
[4]Scholl of Mechanical and Power Engineering, Zhengzhou University, Zhengzhou, 450001, Henan, China.

* Corresponding author: lyh@xcu.edu.cn

ABSTRACT

The appropriate infill sampling criteria to generate more promising updated points is crucial for kriging-based global optimization. For this purpose, a kriging-based sequential optimization algorithm with hybrid infill sampling strategy (KSO-HIS) is proposed. In each iteration, three efficient sampling criteria (i.e., predicted objective minimization criterion, improved expected improvement and new curvature maximization criterion) based on kriging are respectively optimized to produce three optimal solutions by TR (Trust Region) method. Then, a new screening strategy is adopted to determine final expensive evaluation points from the three optimal solutions. The proposed method is compared with three other optimization methods. The test results of eight benchmark functions and a simulation case verify that KSO-HIS can deliver better sampling and convergence performance.
Keywords: Global optimization; Infill sampling criterion; Kriging; Surrogate models.

INTRODUCTION

Surrogate models capable of approximation, prediction, stability and sensitivity analysis have been widely used in the design and optimization of computationally intensive problems. The common surrogate models are PRS (Hosder et al.,2001), RBF (Leonard et al.,1992) and kriging (Stein.M.L., et al.,2012). Among them, kriging can also predict an estimated variance at any point, which will offer promising optimization search direction (Martin et al.,2005).
To obtain the global approximate optimal solution satisfying the accuracy requirements with few expensive function estimates, an optimization method that uses the minimum mean sampling criterion of objective function can simultaneously optimize and propagate the

uncertainty of kriging (Janusevskis et al.,2013). At present, a large number of method obtains only one update point in each iteration. When the maximum number of expensive valuations remains unchanged, it will spend a large time cost due to the expensive valuation. Therefore, the optimization method of obtaining multiple update points at one time has attracted more attention.

Combined with the strategies on kriging believer and constant liar, an EI-based multiple infill method is proposed to realize multi-point parallel evaluation (Ginsbourger et al.,2010). Recently, Lee and Kim use a kriging-based adaptive infill sampling method to calculate failure probability to select additional design points (Lee et al.,2017). A kriging-based global optimization with multi-point sampling method (KGOMS) can transform EI into probability distribution function, and extract appropriate new samples from it (Cai et al.,2017). The EGO (Effective Global Optimization) method proposed by Wang et al. can simultaneously use different infill criteria per cycle to generate new sample points (Wang et al.,2018). Multi-infill strategy for kriging optimization (MSKO) (Song et al.,2018) can employ minimization of kriging prediction, EI, probability of improvement function and lower confidence bounding as infill sampling criteria to obtain four new expensive evaluation points.

Although the above methods can produce multiple update points, there is a large correlation between these points. Therefore, the generated points will not provide the largest amount of information for kriging. If these points are all evaluated with expensive function, it may violate the original intention of the kriging-based optimization. How to get more promising points is worthy of further study.

In view of this, a kriging-based sequential optimization method with hybrid infill sampling strategy is proposed. In each iteration process, predicted objective minimization criterion, the improved EI and new curvature maximization criterion are respectively optimized by TR (Trust Region) method to generate three optimal update points. Next, a screening strategy is used to further select final expensive-evaluation points from these candidate points. The proposed method is compared with the KGOMS, MSKO and EGO. The test results of several benchmarks and a gear pump simulation case show that KSO-HIS has better sampling and convergence performance.

KRIGING MODEL

For design points $\mathbf{X}=\left[\mathbf{x}_1,...,\mathbf{x}_m\right]^T$, $\mathbf{X}\in\Re^{m\times d}$ and objective response $\mathbf{Y}=[y_1,...,y_m]^{\mathrm{T}}$, $\mathbf{Y}\in\Re^{m\times 1}$, kriging based on statistical interpolation and composed of trend function $\mathbf{E}\boldsymbol{\omega}$ and random process $Z(\mathbf{x})$ (Stein.M.L., et al.,2012) can be written as

$$y(\mathbf{x})=\mathbf{E}\boldsymbol{\omega}+Z(\mathbf{x}), \tag{1}$$

where the matrix $\mathbf{E}$ is composed of regression function $e_i(\mathbf{x}^{(j)})(i=1,...,p,\ j=1,...,m)$. The vector $\boldsymbol{\omega}$ is formed by the coefficients of all regression functions. When $\mathbf{E}\boldsymbol{\omega}$ is equal to 0, μ, and the sum of $\omega_i e_i(\mathbf{x})$, they correspond to simple kriging, ordinary kriging, and standard kriging, respectively. Z($\mathbf{x}$) is a realization of a stochastic process with zero mean and

covariance of

$$Cov[z(\mathbf{x}_i), z(\mathbf{x}_j)] = \sigma^2 \boldsymbol{R}(\boldsymbol{\theta}, \mathbf{x}^{(i)}, \mathbf{x}^{(j)}), \tag{2}$$

The parameter $\boldsymbol{\theta}$ and σ^2 are respectively correlation coefficient vector and process variance. For *d*-dimensional problem, the spatial correlation function $\boldsymbol{R}(\boldsymbol{\theta}, \mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ that represents the correlation between point $\mathbf{x}^{(i)}$ and point $\mathbf{x}^{(j)}$ is shown by

$$\boldsymbol{R}(\boldsymbol{\theta}, \mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \prod_{k=1}^{d} R_k(\theta_k, x_k^{(i)} - x_k^{(j)}). \tag{3}$$

The regression $\mathbf{E}\boldsymbol{\omega} \approx Y$ based on unbiased estimator has a generalized least squares solution (see Eq. (4)) and maximum likelihood estimation (see Eq. (5)) of variance.

$$\hat{\boldsymbol{\omega}} = (\mathbf{E}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{E})^{-1}\mathbf{E}^{\mathrm{T}}\mathbf{R}^{-1}Y \tag{4}$$

$$\hat{\sigma}^2 = (Y - \mathbf{E}\hat{\boldsymbol{\omega}})^{\mathrm{T}}\mathbf{R}^{-1}(Y - \mathbf{E}\hat{\boldsymbol{\omega}})/m. \tag{5}$$

where $\mathbf{R}_{m\times m}$ is composed of $\boldsymbol{R}(\boldsymbol{\theta}, \mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ $(i, j = 1, ..., m)$. The objective $\hat{y}$ and mean square error (i.e., $\hat{s}^2$)(Stein.M.L., 2012) of kriging at any point $\mathbf{x}^*$ is expressed by Eq. (6) and Eq. (7).

$$\hat{y}(\mathbf{x}^*) = \mathbf{E}\hat{\boldsymbol{\omega}} + \mathbf{r}^{\mathrm{T}}(\mathbf{x}^*)\hat{\boldsymbol{\gamma}}. \tag{6}$$

$$\hat{s}^2(\mathbf{x}^*) = \mathrm{MSE}[Y(\mathbf{x}^*)] = \hat{\sigma}^2 \left\{ 1 - [e(\mathbf{x}^*)^{\mathrm{T}} \quad \mathbf{r}(\mathbf{x}^*)^{\mathrm{T}}] \begin{bmatrix} 0 & \mathbf{E}^{\mathrm{T}} \\ \mathbf{E} & \mathbf{R} \end{bmatrix} \begin{bmatrix} e(\mathbf{x}^*) \\ \mathbf{r}(\mathbf{x}^*) \end{bmatrix} \right\} \tag{7}$$

where $\mathbf{r}^{\mathrm{T}}(\mathbf{x}^*) = [R(\boldsymbol{\theta}, \mathbf{x}^*, \mathbf{x}_1^*), ..., R(\boldsymbol{\theta}, \mathbf{x}^*, \mathbf{x}_m^*)]$, $\hat{\boldsymbol{\gamma}} = \mathbf{R}^{-1}(Y - \mathbf{E}\hat{\omega})$.

KSO-HIS METHOD

In this section, the three infill sampling criteria based on kriging model are described in detail. In order to optimize the three sampling criteria, an improved TR strategy assisted by kriging model is then proposed. Finally, the specific implementation of the KSO-HIS method is shown in detail.

3.1 Three sampling criteria in KSO-HIS

3.1.1 Predicted objective minimization criterion (POMC)

As the predicted objective, the kriging model $\hat{y}(\mathbf{x})$ is constructed by the existed sample data, and the infill sampling criterion generating next pint is shown by "$\min \hat{y}(\mathbf{x}) \quad \mathbf{x} \in [\boldsymbol{a}, \boldsymbol{b}]$". The sign $\hat{y}(\mathbf{x})$ denotes an approximate kriging of real function, the vector ***a*** and ***b*** are the upper and lower bounds of problems respectively.

Intuitively speaking, taking "$\min \hat{y}(\mathbf{x}) \quad \mathbf{x} \in [\boldsymbol{a}, \boldsymbol{b}]$" as the sampling standard will make the sampling points distributed in the valley which has the greatest impact on the kriging precision. It is helpful for the distance parameter to enhance global exploration ability.

To this end, the minimum distance maximization method is used to make the sampling points evenly distributed in the design space. If minimal distance $d_{\min}$=min ($\|\mathbf{x}\text{-}\mathbf{x}_1\|$,…, $\|\mathbf{x}\text{-}\mathbf{x}_m\|$) and maximal distance $d_{\max}$=max ($\|\mathbf{x}\text{-}\mathbf{x}_1\|$,…, $\|\mathbf{x}\text{-}\mathbf{x}_m\|$) are calculated by untried point $\mathbf{x}$ and the known sampling points $\mathbf{x}_1$,…$\mathbf{x}_m$, and the exponent u can be obtained from the values u={u_1…, 0.001, 0.01, 0.1, 1, 10, 100, 100, …u_k}, then distance parameter d is defined by

$$d(\boldsymbol{x}) = \begin{cases} \dfrac{d_{\max} - d_{\min}}{\text{norm}(\boldsymbol{b}-\boldsymbol{a})}, \hat{y}(\boldsymbol{x}) > 0 \\ 1 - \dfrac{d_{\max} - d_{\min}}{\text{norm}(\boldsymbol{b}-\boldsymbol{a})}, \text{ other} \end{cases} \tag{8}$$

Therefore, the POMC find $\mathbf{x}$ by

$$\min \ \hat{y}(\mathbf{x}) \cdot d(\mathbf{x}) . \tag{9}$$

The reasons for using POMC as a filling sampling criterion are as follows: (1) the approximate kriging is used as the optimization objective. Its minimization may get a local optimal solution, which is usually a promising point. The expensive evaluation of this point is conducive to improving the kriging accuracy. (2) The parameter d can ensure the distance between the iterative points is not too close, or else, it may make matrix $\mathbf{R}$ singular, and lead to construction failure of kriging.

3.1.2 The improved EI criterion

As a quality factor, EI with the estimated objective and predicted standard error of kriging model in EGO can calculate the best expected level of improvement and balance local and global search. The EI is calculated as follows: Before obtaining the unknown sample point $\mathbf{x}$, its objective $y(\mathbf{x})$ is uncertain, and the uncertainty may be represented by Gaussian normal random distribution $Y(\mathbf{x})$. It assumed that the y_S and $y_{\min} = \min(y^{(1)},...,y^{(m)})$ are the expensive function evaluation of a new the point $\boldsymbol{x}$ and the current known best function value, respectively. The expected improvement function can be expressed by $\text{EI}(\mathbf{x}) = \text{E}[\max(y_{\min} - Y(\mathbf{x}), 0) \mid Y_S = y_S]$. Since it is a conditional Gaussian process with mean $\hat{y}(\mathbf{x})$ and variance $\hat{s}^2(\mathbf{x})$, the expected function can be expressed as

$$\text{EI}(\mathbf{x}) = \hat{s}(\mathbf{x}) \left[u\Phi(u) + \frac{d\Phi(u)}{du} \right], \ u = \frac{y_{\min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})} \tag{10}$$

where $\Phi()$ is a normal cumulative distribution function. New expensive evaluation point will be selected by maximizing EI.However, the complexity of EI may cause long optimization time. In addition, the deception of initial sampling may cause small standard deviation estimation. As a result, only the data points close to the current optimal solution have a large EI. Before a more global search, it will lead to a more detailed search in the area near the initial optimal solution. Finally, in the case that there is an exact optimal solution in initial kriging, this method has no adaptability. Therefore, it is necessary to combine with other sampling criteria to improve global optimization ability of the proposed algorithm.

To this end, The improved EI criterion including cool parameter q was introduced by

$$qEI = E(I^q) = \hat{s}^q \sum_{i=0}^{q} (-1)^i \left(\frac{q!}{i!(q-i)!}\right) u^{q-i} \cdot A_i \,. \tag{11}$$

$I^q = \max\{0, (y_{\min} - Y(\mathbf{x}))^q\}$, the recursive formula $A_i = -(u)^{i-1}\phi(u) + (i-1)A_{i-2}$ with $A_0 = \Phi(u)$ and $A_1 = -\Phi(u)$. The qEI increases with the increase of q. Even if a large qEI produces a small probability, which makes the algorithm more inclined to global search. In addition, small q may be easy to search in the local basin. Therefore, the parameter q has a good guidance for the search direction of the optimization method to some extent. Therefore, M.J. sasena selects different q value to change the qEI in the constant set according to the number of iterations (Sasena. M., 2002). In order to make the value more general, according to the discussion of related parameter $\boldsymbol{\theta}$ of kriging in the reference (Wu et al.,2014), as shown in Eq. (12), Euclidean norm is rounded as q value.

$$q = \text{round}[\text{norm}(\boldsymbol{\theta}_{i+1} - \boldsymbol{\theta}_i)] \,. \tag{12}$$

The reasons why qEI criterion is used are as follows: (1) qEI is suitable for global optimization based on kriging model. Due to taking into account kriging's mean and variance, large qEI in the optimal search process is usually far away from the current optimal solution, which is helpful to find a smaller objective in an unexplored area. Therefore, it can perform a balance between exploration and exploitation. (2) Under the condition of less expensive evaluations or constant kriging model, maximized qEI can quickly converge to a better approximate optimal solution, and the optimization process keeps a good stability, so it is widely used in many engineering designs.

3.1.3 New curvature maximization criterion (NCMC)

Mathematically the curvature (Capozziello et al.,2002) represents the bending degree of at a point of the curve. The big curvature is consistent with the great bending degree of curve. The meaning of curvature is shown in Figure. 1. A point A_0 on the smooth curve M is selected as the basic point. Let the arc length of point A relative to point A_0 be s, the inclination angle of the tangent at point A is β. Seen from Figure. 1, the arc length of another point A' on the curve is $s + \Delta s$, and the inclination of the relative tangent is $\beta + \Delta\beta$. Then, the length of the arc AA' is Δs. If a point is moved from A to A', the change of angle is $\Delta\beta$. When Δs approaches zero (A' is infinitely close to A), the curvature $C(\mathbf{x})$ at point A can be expressed by

$$C = \lim_{\Delta s \to 0} \left| \Delta\beta / \Delta s \right| \tag{13}$$

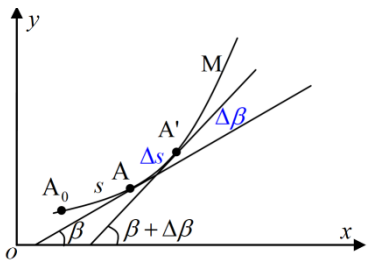

Figure 1. Curvature diagram.

To facilitate the calculation of curvature, ordinary kriging $\hat{y}(\mathbf{x}) = \hat{\mu} + \mathbf{r}^{\mathrm{T}}(\mathbf{x})\mathbf{R}^{-1}(Y - \mathbf{1}\hat{\mu})$ is chosen. The ordinary kriging is continuous, first order and second order differentiable because the Gaussian kernel function used by kriging is exponential type function. Its first-order and second-order partial derivative can be expressed by Eq. (14) and Eq. (15).

$$\frac{\partial \hat{y}}{\partial x_i} = \frac{\partial \mathbf{r}}{\partial x_i}\mathbf{R}^{-1}(Y - \mathbf{1}\hat{\mu}) + \frac{\partial Y}{\partial x_i}(\mathbf{R}^{-1})^{\mathrm{T}}\mathbf{r} \tag{14}$$

$$\frac{\partial^2 \hat{y}}{\partial x_i \partial x_j} = \frac{\partial^2 \mathbf{r}}{\partial x_i \partial x_j}\mathbf{R}^{-1}(Y - \mathbf{1}\hat{\mu}) + \frac{\partial Y}{\partial x_i}(\mathbf{R}^{-1})^{\mathrm{T}}\frac{\partial \mathbf{r}}{\partial x_i} + \frac{\partial^2 Y}{\partial x_i \partial x_j}(\mathbf{R}^{-1})^{\mathrm{T}}\mathbf{r} + \frac{\partial \mathbf{r}}{\partial x_i}\mathbf{R}^{-1}\frac{\partial Y}{\partial x_i} \tag{15}$$

where $i, j = 1, ..., n$. Hessian matrix $\mathbf{H}$ at any point can be obtained from the formula

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 \hat{y}(\mathbf{x})}{\partial x_1^2} & \mathrm{L} & \frac{\partial^2 \hat{y}(\mathbf{x})}{\partial x_1 \partial x_n} \\ \mathrm{M} & \mathrm{O} & \mathrm{M} \\ \frac{\partial^2 \hat{y}(\mathbf{x})}{\partial x_n \partial x_1} & \mathrm{L} & \frac{\partial^2 \hat{y}(\mathbf{x})}{\partial x_n^2} \end{bmatrix}. \tag{16}$$

According to the differential geometry principle, the curvature $C(\mathbf{x})$ can be estimated by the eigenvalue of Hessian matrix. If the eigenvalues of matrix $\mathbf{H}$ are $c_1, c_2, ..., c_n$, then the curvature at point $\mathbf{x}$ can be calculated by Eq. (17).

$$C(\mathbf{x}) = \sqrt{\sum_{i=1}^{n} c_i(\mathbf{x})} \tag{17}$$

Therefore, equation (18) can be expressed by

$$\max\ C(\mathbf{x}). \tag{18}$$

Maximizing curvature in each cycle can find a more promising sampling point, which objective may be a maximum or a minimum. Moreover, the sampling point corresponding to maximum objective value is meaningless for global minimization. Therefore, further screening is necessary.

3.2 The kriging-based trust region (TR) method

TR method can guarantee the convergence of optimization (Martínez et al.,1995). It can generate a trial step l_k by optimizing problem (20) for the objective function.

$$\min_{d \in \Re^n} \quad \phi(l) = \Delta_k^{\mathrm{T}} l + \frac{1}{2} l^{\mathrm{T}} \mathbf{G}_k l, \quad \text{s.t.} \quad \|l\|_2 \le r_k \tag{20}$$

The equation $\Delta_k = \nabla y(x_k)$ and $\mathbf{G}_k$ are the gradient value at the point x_k, a $n \times n$ Hessian matrix, respectively. And sign r_k is a TR radius, d_k is a norm distance. Next, we define the ratio

$$\rho_k = [y(x_k) - y(x_k + d_k)] / [q_k(0) - q_k(d_k)] \tag{21}$$

The reduction and extension of TR can be realized by formula (22).

$$r_k = \begin{cases} [\eta_1 \rho_k,\ \eta_2 \rho_k], & \text{if } \rho_k < \gamma_1 \\ [\eta_2 \rho_k,\ \rho_k], & \text{if } \gamma_1 \le \rho_k < \gamma_2 \\ [\eta_k,\ \infty), & \text{if } \gamma_2 \le \rho_k \end{cases}. \tag{22}$$

The constant $\eta_1,\ \eta_2$ and $\gamma_1,\ \gamma_2$ satisfy $0 < \eta_1 \le \eta_2 < 1,\ 0 \le \gamma_1 < \gamma_2 < 1$.

The key issue of the TR method is how to calculate the TR and how to decide whether to expand or narrow the trial TR. Due to kriging, the ratio η_k is defined by

$$\eta_k = [y(\mathbf{x}_{k-1}^*) - y(\mathbf{x}_k^*)] / [y(\mathbf{x}_{k-1}^*) - \hat{y}(\mathbf{x}_k^*)] \tag{23}$$

For the current optimal solution x_k^* in Eq. (23), $y(x_k^*)$ is real objective function value, $\hat{y}(x_k^*)$ is the estimated objective produced by kriging. The corresponding extension coefficient ξ_k is

$$\xi_k = \begin{cases} \tau_1, & \rho_k \le \tau_1 \\ \tau_3, & \tau_1 < \rho_k \le \tau_2 \\ \tau_4, & \text{otherwise} \end{cases}, \tag{24}$$

where $0 \le \tau_1 < \tau_2 < 1,\ 1 \le \tau_3 < \tau_4$, their typical values are $\tau_1 = 0.25, \tau_2 = 0.75, \tau_3 = 1,\ \tau_4 = 2$.

Further, the TR is described by

$$\begin{cases} \boldsymbol{u}_k = \boldsymbol{x}_{k-1}^* + \xi_k \cdot (\boldsymbol{u}_{k-1} - \boldsymbol{l}_{k-1}) \\ \boldsymbol{l}_k = \boldsymbol{x}_{k-1}^* - \xi_k \cdot (\boldsymbol{u}_{k-1} - \boldsymbol{l}_{k-1}) \end{cases}. \tag{25}$$

Parameter $\boldsymbol{u}_k$ and $\boldsymbol{l}_k$ in Eq. (25) are respectively upper bound and low bound of TR. When $k = 0$, $\boldsymbol{u}_0$ and $\boldsymbol{l}_0$ are initial bounds, $\boldsymbol{x}_0$ is midpoint of initial bounds.

3.3 Screening strategy

In our approach, three candidate sampling points can be obtained by using trust domain method to optimize the above three sampling criterion. The optimization based on kriging model needs to use as few expensive-evaluation points as possible to obtain a global approximate optimal solution that meets certain approximate accuracy requirements. Therefore, the three candidates need be deeply screened to select sampling points with greater potential. The specific screening process is divided into the three steps.

Step 1: A candidate point with the largest kriging objective estimation need to be removed among the three candidates. There are two other reasons why we need to do this. (i) New curvature maximization criterion may get a global maximum value, which must be removed; (ii) When the three sampling points lack independence or the distance between them is very close, we need to filter out the sampling points with less potential information.

Step 2: Next, the remaining two candidate points is further screened. The minimum distance $d_{\min}$ is firstly calculated between any two points. The $d_{\min}^i = \| \mathbf{x}_i - \mathbf{x}_k \|^2$, $\mathbf{x}_k \in \mathrm{X}$ need also be calculated between any one $\mathbf{x}_i$ $(i \in \{1,2\})$ in remaining two candidate points and all known points. Due to the value $d_{\min}^i$ and the value $d_{\min}$, we can obtain the improvement index

$\delta = | y_m - y_n | / d_{\min}$ and $\delta_i = | \hat{y}_i - y_k | / d^i_{\min}$ (Here $\hat{y}_i$ is the kriging objective estimate at point $\mathbf{x}_i$). when $\delta_i < \delta$, $\mathbf{x}_i$ is removed, or else, accept $\mathbf{x}_i$.

Finally, one or two candidate points will be selected to perform expensive function evaluation and added to the data sample set X. It is worth noting that if both candidate points are removed, the candidate point with the minimum kriging objective estimation will be reserved as the newly added expensive-evaluation points.

3.4 KSO-HIS implementation

The three sampling criteria and optimization sampling strategy used in the KSO-HIS method are introduced in the front. But the smooth operation of the complete algorithm also requires many auxiliary components, such as stopping criteria, initial sampling method, and so on. Therefore, this section mainly gives the specific implementation process. These steps of the KSO-HIS method are shown in Figure. 2.

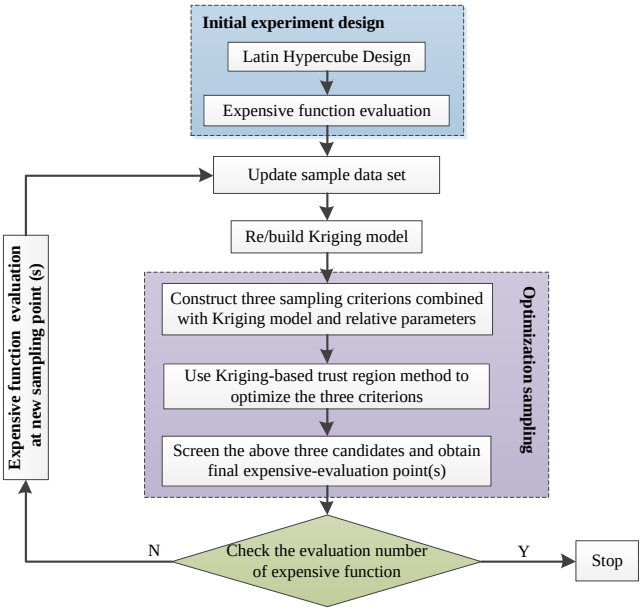

Figure 2. Main steps of the proposed KSO-HIS algorithm.

Table 1 provides the input and output parameters. The main steps of the proposed KSO-HIS algorithm are shown as follows.

Table 1. Input and output parameters of KSO-HIS method.

Input	(I) A black-box function $y(\mathbf{x})$, $\boldsymbol{a} \le \mathbf{x} \le \boldsymbol{b}$, $\mathbf{x} \in \Re^n$. (II) Initial sample set $\mathbf{X}=[\,]$ and $\mathbf{Y}=[\,]$. (III) Maximum expensive-evaluation number $N_{\max}$. (IV) The ordinary kriging model with Gauss kernel function. (V) Constant vector **t**={0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 100}. (VI) Initial TR bounds $\boldsymbol{u}_0 := \boldsymbol{b}$ and $\boldsymbol{l}_0 := \boldsymbol{a}$.
Output	The explored global optimal solution of KSO-HIS.

Step 1. Initial sampling. LHD (Latin Hypercube Design) is used to generate m ($m=2*(n+4)$) initial sample points. For each point $\mathbf{x}_i (i=1,...,m)$, calculate real objective $y(\mathbf{x}_i)$. Then, place $y(\mathbf{x}_i)$ and $\mathbf{x}_i (i=1,...,m)$ in sample set $\mathbf{Y}=[\,]$ and $\mathbf{X}=[\,]$, respectively. Finally, the current optimal objective $y_{\text{best}} := y_{\min}$, the current optimal solution $\mathbf{x}_{\text{best}} := \mathbf{x}_{\min}$ and the current expensive evaluation number $M := m$ are set.

Step 2. Update sample data and build kriging model. The sample data set $\mathbf{Y}$ and $\mathbf{X}$ will further be updated by new expensive-function-valuation point(s), which can be produced optimization sampling and screening strategy. After that, kriging model will be updated by the new sample data set $\mathbf{Y}$ and $\mathbf{X}$.

Step 3. Optimization sampling. The acquisition of new expensive-valuation point(s) will be completed in the following three steps.

Step 3.1. Construct three sampling criteria. Combining the parameter $\hat{y}(\mathbf{x}), d_{\min}(\mathbf{x}), \hat{s}^2(\mathbf{x})$, and $d_{\max}(\mathbf{x})$ of kriging model and the distance factor D, the three sampling criteria are constructed. The specific process can be found in Section 3.1.

Step 3.2. Use kriging-based TR method to generate three candidates. The three sampling criteria are respectively optimized by kriging-based TR method to produce three candidate sampling points in each cycle. The details on TR method using kriging model can be found in Section 3.2.

Step 3.3. Further screening of the above three candidates. The Screening strategy is used to select final expensive-evaluation point(s). Specific process can be found in Section 3.3.

Step 4. Check the expensive-evaluation number. Extract the expensive-evaluation- point number M (i.e., the number of elements in sample set $\mathbf{X}$) from current data sample set $\mathbf{X}$. When parameter M is greater than maximum expensive-evaluation number $N_{\max}$, the KSO-HIS method will be terminated. Otherwise, expensive function evaluations will be performed for the new sampling point(s) obtained from the optimized sampling of Step 3, and they will join in the sample data set $\mathbf{X}$ and $\mathbf{Y}$ respectively.

Step 5. Stop. Terminate KSO-HIS method, output global optimal solution $(\boldsymbol{x}_{best}, y_{best})$.

EXPERIMENTAL STUDIES

To verify the optimization performance of the proposed method, firstly, the different sampling

characteristics of the two-dimensional benchmark functions are graphically illustrated. Then, the iterative optimization processes of all problems are graphed to show its optimization effect. In addition, compared with the other three optimization algorithms (MSKO, KGOMS and EGO), It may show that the KSO-HIS method has better convergence performance. Finally, the structural optimization of a cycloid rotor gear pump problem is transformed into a four-dimensional black box function, which is also optimized by maximizing the volume efficiency to show that the proposed method has good application value.

4.1 Benchmark problems

4.1.1 Test results of KSO-HIS method

Figure 3 provides the graph of two 2-dimensional test functions. The test functions have different characteristics. The Goldstein-price (GP) function with gentle trend has several local basins. The Cross-in-Tray and Holder Table have many local minima, four of which are global minimum global solutions.

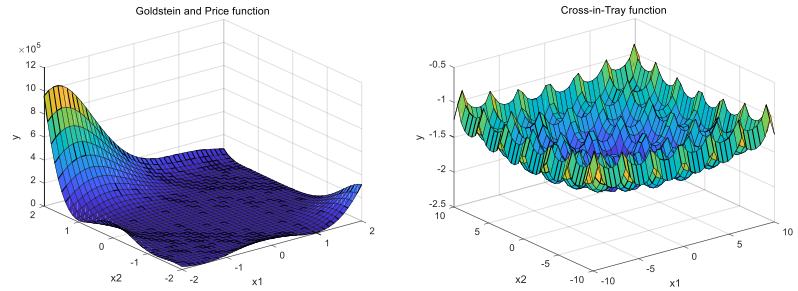

Figure 3. Graphs of eight 2-dimensional benchmark functions.

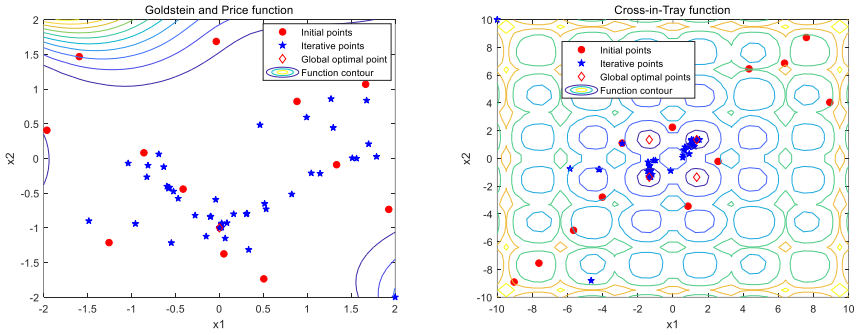

Figure 4. Sampling point distribution of GP function and Cross-in-Tray function.

In Figure. 4, the sampling-points distribution diagrams of the four two-dimensional problems is provided respectively (this also includes the sampling points obtained from the initial experimental design). Within a given number of expensive evaluations, the KSO-HIS method can make sure that each of test function can converge to the region near the global optimal solution quickly. For the GP and Cross-in-Tray, KSO-HIS almost finds the actual global optimal solutions. What's exciting is that Cross-in-Tray function has found the actual global minimum. Therefore, the proposed method can find the satisfactory optimal solutions for all types of two-dimensional test problems.

To explore the optimization ability of KSO-HIS, the characteristics including dimension (D), search space (SP), initial sample number (ISN), maximum number of expensive evaluations

(MNEE) and real global optimal solution (RGOS) of the multi-dimensional test functions are shown in Table 2. There is a big hole at the center of the Ackley5 function, and the area around it is almost flat. Hartman6 functions has 6 local minima. Trid9 with bowl shape has only a global solution. The F16 is differentiable and multimodal.

Table 2. The multi-dimensional benchmarks and their feature.

Test function	D	SP	ISN	MNEE	RGOS
Ackley5	5	$[-32.768\ 32.768]^5$	18	120	0
Hartman6	6	$[0\ 1]^6$	20	140	-3.32237
Trid9	9	$[-81,81]^9$	26	200	-156
F16	16	$[-1\ 1]^{16}$	40	340	25.875

These iteration processes (also including initial design) of objective function values for two/multi-dimensional problems are shown from Figure 5 to Figure 6.

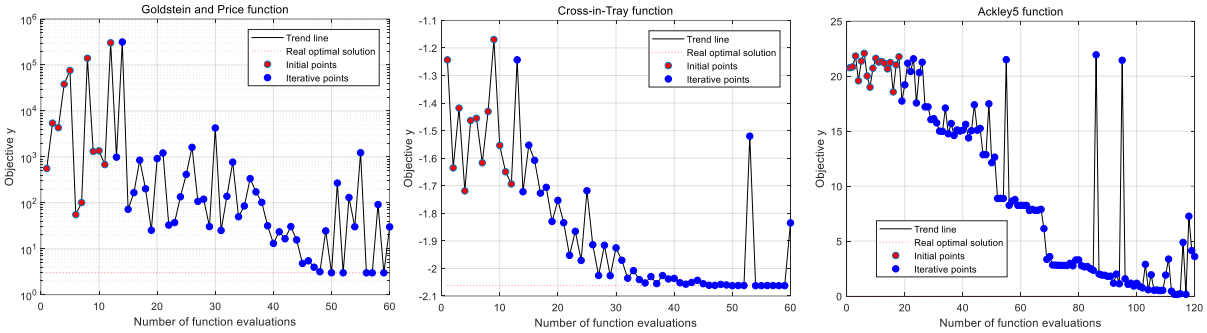

Figure 5. Iterative results of GP function, Cross-in-Tray function and Ackley5 function.

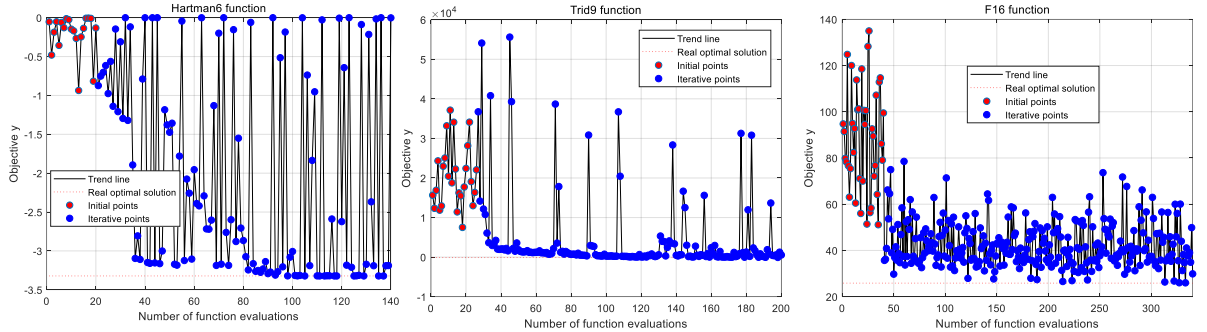

Figure 6. Iterative results of Hartman6 function, Trid9 function and F16 function.

For two-dimensional problems, the relative errors of two functions are all less than 0.001. Especially for cross in tray functions, when their expensive function evaluations are 37 times, 20 times and 19 times respectively, the relative errors are less than 0.001. The above analysis shows KSO-HIS can meet the accuracy requirements of two-dimensional problems.

For multi-dimensional problems, Ackley5 and Hartman6 functions can explore some

unknown regions continuously in the iterative process, and they have achieved better convergence accuracy by using only a few iterations. Although the dimensions of Trid9 are high, it can find a global approximate optimal solution with an error accuracy of about 0.01. The 16- dimensional F16 is optimized to find the approximate optimal solution which is close to the real minimum value under the given expensive estimation. This shows that KSO-HIS can also be applied to some high-dimensional problems.

To sum up, KSO-HIS shows good convergence and stability in most of the function testing process.

4.1.2 Comparison with other optimization methods

To reflect the difference of the optimal results caused by the algorithm uncertainty in the comparison process, each optimization problem is run for 15 times, and the average value of the optimal solution obtained is taken as the evaluation standard of the final optimal solution. These comparison results of MSKO, KGOMS and EGO are shown in Table 3. The comparison parameters include D (Dimension), UBAOS (the Upper Bound of Approximate Optimal Solutions), LBAOS (the Lower Bound of the Approximate Optimal Solution), RE (Relative Error) between the average optimal solution and the real minimal values, and the average times of expensive valuation (ATEV) when the RE is less than 0.1. When the true optimal value of some optimization problem is 0, RE cannot be calculated. In this case, the maximum UBAOS generated by the four algorithms is chosen as the denominator of relative error, while the numerator is still represented by the actual absolute error.

For the two-dimensional test problem, the proposed algorithm can find the real global optimal solutions for the optimization Cross-in-Tray function. The global approximate optimal solution with relative error accuracy less than 0.1 is found for another functions. With the increase of dimension, the KSO-HIS algorithm also produces better optimization results for Ackley5, Hartman6 problems. Except Trid9, the REs of other high-dimensional problems are also very close to 0.1. Especially for F16 functions, the convergence accuracy is satisfactory under a certain amount of expensive evaluation conditions. The complexity of the test function, the specification and some limitations in the modeling process of kriging are the key factors affecting the convergence precision. In general, KSO-HIS has smaller LBAOS, RE and ATEV than other three optimization algorithms in most cases. The optimization results of MSKO and KGOMD are close, and the test results of EGO are relatively poor. In addition, the relatively small UBAOS also shows that KSO-HIS has good stability. Therefore, the proposed method is appropriate for the optimization of these problems.

Table 3. Comparison results of MSKO, KGOMS and EGO.

BF	D	RGOS	Method	UBAOS	LBAOS	RE	ATEV
GP	2	3	KSO-HIS	3.0458	3.0001	7.6000e-03	42.8
			MSKO	3.0982	3.0198	1.9700e-02	46.3

			KGOMS	3.2036	3.0150	3.6400e-02	44.6
			EGO	4.6983	3.0437	0.2903	>60
			KSO-HIS	-2.0625	**-2.06261**	4.8482e-06	39.2
Cross-in-Tray	2	-2.06261	MSKO	-2.0602	-2.0626	5.8664e-04	41.5
			KGOMS	-2.0602	-2.06261	5.8421e-04	40.6
			EGO	-2.0622	-2.0626	1.0181e-04	38.8
			KSO-HIS	3.9258	0.0131	0.0033	119.8
Ackley5	5	0	MSKO	3.0681	0.0163	0.0042	>120
			KGOMS	2.6093	0.0152	0.0039	>120
			EGO	2.5279	0.7662	0.1952	>120
			KSO-HIS	-3.3223	-3.32236	1.2040e-05	98.2
Hartman 6	6	-3.32237	MSKO	-3.3062	-3.32235	0.0024	96.8
			KGOMS	-3.3130	-3.32236	0.0014	100.5
			EGO	-3.1866	-3.3220	0.0205	107.8
			KSO-HIS	-49.9737	-104.4772	0.5050	>200
Trid9	9	-156	MSKO	-39.4866	-106.2894	0.5328	>200
			KGOMS	-50.2874	-108.0996	0.4923	>200
			EGO	-34.5571	-76.5528	0.6439	>200
			KSO-HIS	30.1967	26.5756	0.0970	302.6
F16	16	25.875	MSKO	32.4008	29.3981	0.1924	326.8
			KGOMS	33.7554	28.4277	0.2016	331.5
			EGO	42.9713	37.5962	0.5569	>340

4.2 Simulation case

The KSO-HIS algorithm is applied to the structural optimization of a cycloid rotor gear pump, and compared with MSKO, KGOMS and EGO algorithm. The pump is widely used in engineering field because of its compact structure, stable operation, low cavitation and high-volume efficiency. Figure. 7 is its three-dimensional structure diagram established by Pro/E. The main structure includes front end cover, rear end cover and inner and outer rotors. During the working process, the inner and outer rotors are meshed and rotated around their respective axes according to different speeds, which will make the cavity size between the inner and outer rotors have a constant change. Further, it will lead to the internal pressure change. In this case, the oil in/outlet cavities and the changing cavities are effectively connected to achieve the oil suction and discharge.

The volume efficiency corresponds to the large actual flow, which will make the cycloid pump have a good oil suction and drainage. Due to the parameter setting of cycloid pump (for example, the theoretical angle of setting the oil inlet and outlet cavity is 26 degrees and 13.2 degrees respectively) and the practical experience (Li et al.,2014), the volume efficiency is set as the optimization target, then the problem can be described as

$$\max \ V/V_0, \ \ \alpha_{1,2} \in [22,26], \ \beta_{1,2} \in [13.2,16.2] \tag{26}$$

Parameter V_0 and V in the formula (26) respectively represent the theoretical flow and actual flow of the oil outlet for the cycloid pump. The sign $\alpha_{1,2}$ and $\beta_{1,2}$ are the corresponding closed line angles when the inner and outer rotors are in the actual containment area.

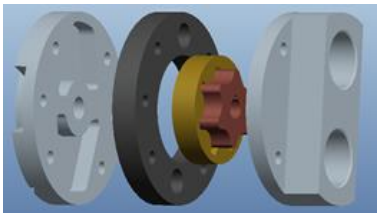

Figure 7. The structure of the cycloid rotor pump.

Figure 8. Mesh on internal flow field.

The STL format of the cycloid pump model established by Pro/E is imported into Pumplinx and meshed as shown in Figure 8. Under the standard atmospheric pressure, the oil density is 800kg/m3, the dynamic viscosity is 5×10^{-6} N·s/m^2, and the saturation pressure is 400 Pascal. The speed of cycloid pump is 3000 rpm, the theoretical value $V0$ the flow is equal to 3.22 L/min according to the calculation formula (Li et al.,2014). For this optimization problem, the mathematical expression of function can't be given, so designers can only improve the volumetric efficiency through expensive simulation optimization. Therefore, it is appropriate to assume the optimization problem as a black-box function.

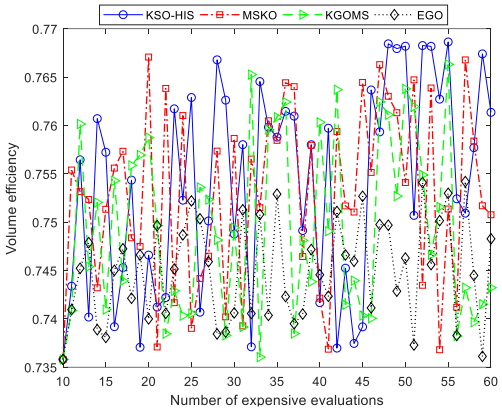

Figure 9. Iterative results on volumetric efficiency.

In view of this, 10 identical initial experimental design points (The maximum volumetric efficiency of the initial sampling is 0.7359) are used for KSO-HIS, MSKO, KGOMS and EGO algorithms, and the maximum expensive simulation number is set as 60. Then the whole optimization process of each algorithm will take about 43 hours. Figure 9 shows the relationship between expensive-valuations number and volumetric efficiency. It can be seen from the figure that the volume efficiency of KSO-HIS is 0.2%, 0.3% and 1.92% higher than that of MSKO, KGOMS and EGO respectively. It shows that the proposed method has better optimization performance in the actual simulation.

CONCLUSION

Many engineering fields have used kriging-based global optimization algorithms to solve expensive black box problems. What kind of strategy is used to explore new sample points with greater potential information is very important. In order to improve this situation, a KSO-HIS is proposed. In this method, the prediction target minimization criterion, the improved EI and the new curvature maximization criterion, which are generated by the constructed kriging model, are optimized using the TR method to sample the three candidate points. Finally, the final expensive evaluation point(s) will be selected from these candidate points by using a new screening strategy. Compared with other methods, this method shows better convergence performance. The future research direction can start from the following two aspects: (1) When the obtained candidate points do not have much information, some correction or adjustment methods may be selected to make the new sampling point deviate from the original position and face more prospective areas; (2) Improve the approximate accuracy by adaptively selecting the kernel functions or trend functions of the kriging model, so as to find more valuable points in sequential optimization.

ACKNOWLEDGEMENTS

This work is supported by National Natural Science Foundation of China (No.12272354), Fundamental Research Funds for the Central Universities (No. 2662022LXYJ003), Science &Technology Innovation Talents in Universities of Henan Province (No. 21HASTIT027), Henan Natural Science Foundation (No. 202300410346).

REFERENCES

Hosder, S., Watson, L.T., Grossman, B., Mason, W.H., Kim, H., Haftka, R.T. & Cox, S. E. 2001. Polynomial response surface approximations for the multidisciplinary design optimization of a high speed civil transport. Optimization and Engineering.2(4) p. 431-452.

Leonard, J.A., Kramer, M.A. & Ungar, L.H. 1992. Using radial basis functions to approximate a function and its error bounds. IEEE transactions on neural networks. 3(4) p. 624-627.

Stein, M.L. 2012. Interpolation of spatial data: some theory for kriging: Springer Science & Business Media.

Martin, J.D. and Simpson, T.W. 2005. Use of kriging models to approximate deterministic computer models. AIAA journal. 43(4) p. 853-863.

Penadés-Plà, V., García-Segura, T. & Yepes, V. 2020. Robust Design Optimization for Low-Cost Concrete Box-Girder Bridge. Mathematics, 2020. 8(3) p. 398.

Janusevskis, J. & Riche, R.L. 2013. Simultaneous kriging-based estimation and optimization of mean response. Journal of Global Optimization. 55(2) p. 313-336.

Ginsbourger, D., Riche, R.L. & Carraro, L. 2010. kriging is well-suited to parallelize

optimization, in Computational intelligence in expensive optimization problems. Springer. p. 131-162.

Lee, S. & Kim, J.H. 2017. An adaptive importance sampling method with a kriging metamodel to calculate failure probability. Journal of Mechanical Science Technology. 31(12) p. 5769-5778.

Cai, X., Qiu, H., Gao, L., Yang, P. & Shao, X. 2017. A multi-point sampling method based on kriging for global optimization. Structural Multidisciplinary Optimization. 56(1) p. 71-88.

Wang, Y., Han, Z.H., Zhang, Y. & Song, W.P. 2018. Efficient Global Optimization using Multiple Infill Sampling Criteria and Surrogate Models. in 2018 AIAA Aerospace Sciences Meeting

Song, C., Yang, X. & Song. W. 2018. Multi-infill strategy for kriging models used in variable fidelity optimization. Chinese Journal of Aeronautics. 31(3) p. 448-456.

Sasena, M. 2002. Flexibility and Efficiency Enhancements For Constrainted Global Design Optimization with kriging Approximations.

Li, Y., Wu, Y. & Huang Z. 2014. An Incremental kriging Method for Sequential Optimal Experimental Design. 97(4) p. 323-357.

Capozziello, S. 2002. Curvature quintessence. International Journal of Modern Physics D. 11(04) p. 483-491.

Martínez, J.M. & Santos, S.A. 1995. A trust-region strategy for minimization on arbitrary domains. Mathematical Programming. 68(1-3) p. 267-301.

Li, Y., Wu, Y. & Huang, Z. 2014. An Incremental kriging Method for Sequential Optimal Experimental Design. CMES-Computer Modeling in Engineering & Sciences. 97(4) p. 323--357.