Published in ENGINEERING OPTIMIZATION, 51(2): 247-264, 2019
https://doi.org/10.1080/0305215X.2018.1446084

# Evolutionary Reliable Regional Kriging Surrogate for Expensive Optimization

Suprayitno[1] and Jyh-Cheng Yu[2]

[1,2]*Department of Mechanical and Automation Engineering, National Kaohsiung First University of Science and Technology, Kaohsiung 811, Taiwan, ROC.*

[1] *Department of Mechanical Engineering, State University of Malang, Indonesia*

[2] *Corresponding author: jcyu@nkfust.edu.tw, ORCiD: 0000-0003-3716-0289*

## Abstract

This work proposes a sequential optimization algorithm, EORKS, combining a Kriging surrogate from an adaptive sampling and an iterative constrained search in the dynamic reliable regions to reduce the sampling size in expensive optimization. A surrogate established from small samples is liable to limited generality, which leads to a false prediction of optimum. EORKS applies the Kriging variance to establish the reliable region neighbouring the learning samples to constrain the evolutionary searches of the surrogate. The verified quasi-optimum is used as an additional sample to dynamically update the regional model according to the prediction accuracy. A hybrid infilling strategy switches between the iterative quasi-optimums and the maximum expected improvement from Kriging to prevent early convergence of local optimum. EORKS provides superior optimums in several benchmark functions and an engineering design problem using a much smaller samples compared with the literature results, which demonstrates the sampling efficiency and searching robustness.

**Keywords**: Expensive optimization; Surrogate-based optimization; Sequential approximate optimization; Kriging; Evolutionary algorithms

### 1. Introduction

Current advances in evolutionary optimization such as genetic algorithm (GA) (Holland 1973) and particle swarm optimization (PSO) (Eberhart and Kennedy 1995) are effective to solve for the global optimum of highly nonlinear problems (Zang, Zhang, and Hapeshi 2010). Success applications of evolutionary algorithm in engineering optimization have been well reported in literature. However, cost constraints of expensive optimization often impose a limit on the number of samples if time consuming simulations and physical experiments are involved. Direct applications of population-based algorithms to expensive optimization are often impractical because a lot of fitness evaluation are required in the evolution process. Surrogate assisted evolutionary optimizations replace the engineering system with an approximate surrogate trained from finite learning samples. Instead of direct interaction with actual engineering systems, evolutionary optimizers search in the surrogate for an estimate of optimum to reduce the sampling cost (Tenne 2012).

Published in ENGINEERING OPTIMIZATION, 51(2): 247-264, 2019
https://doi.org/10.1080/0305215X.2018.1446084

Typical surrogate modelling includes polynomial response surface method (RSM), Kriging model, radial basis function (RBF), artificial neural network (ANN), moving least-squares (MLS), and support vector regression (SVR) (Forrester and Keane 2009). The choice of modelling scheme will affect prediction accuracy and sampling strategies. Le et al. introduced an adaptive surrogate that selects multiple diverse approximation methods according to the evolvability metric in the optimum search (Le et al. 2013). The evolvability learning of surrogates is constructed for use within a trust-region enabled local search to increase the searching efficiency. Kriging and radial basis function are preferred to conventional quadratic models for robust design optimization because of a superior ability to simulate non-linear responses (Elsayed and Lacor 2014). Kriging is known as Gaussian process modelling consisting of a regression model and a stochastic process (Sacks et al. 1989). Regression models can be zeroth, first, or second order polynomials which globally approximates design space, while stochastic processes compose local deviations. Kriging model provides a measure for the predictor error (Kriging variance) to establish a confidence interval as shown in **Figure 1**. Some  (Liu and Maghsoodloo 2011) adopted Taylor expansion as the base functions to enhance the regression approximation of Kriging. Xia et al. developed Adaptive Dynamic Taylor Kriging (ADTK) to assist Monte Carlo simulation for reliability evaluation (Xia, Ren, and Koh 2017). The Taylor Kriging model is iteratively updated by additional sampling points selected adaptively from the test points until the fitting error of the surrogate is acceptable.

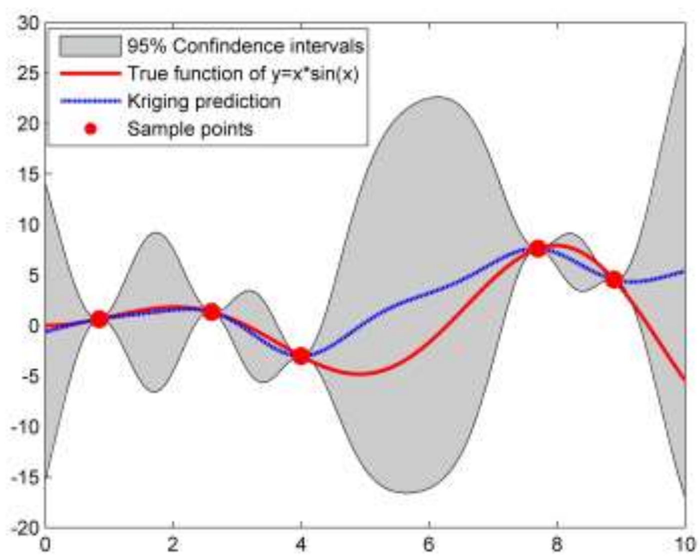



**Figure 1** Example of one-dimensional data interpolation by Kriging with confidence intervals.

The generality of surrogate is directly related to the density of training samples. The efficiency of sample deployment becomes a major concern in expensive optimization. Unlike conventional surrogate-based approaches which prefer well established samples to construct a surrogate with global accuracy, the iterative search known as sequential approximate optimization (SAO) uses adaptive sampling to iteratively search and update the surrogate model (Kitayama, Arakawa, and Yamazaki 2011). The efficient global optimization (EGO) used Kriging surrogate and the maximum expected improvement criteria as the sequential sampling strategy (Jones, Schonlau, and Welch 1998). Eason and Cremaschi introduced pure and mixed adaptive sampling algorithms to iteratively update a neural network model (Eason and Cremaschi 2014). Yu and Juang applied the quasi-optimum obtained from a constrained population search in the reliable region as the sequential sampling to refine an iterative network model (Yu and Juang 2010). Kanakasabai and Dhingra applied a progressive Kriging surrogate for reliability-based design optimization, and proposed the Halton sequence as adaptive sampling strategy (Kanakasabai and Dhingra 2014).

Typical training samples of surrogate are from existed field data and design of experiments (DOE). Most studies prefer DOE since uncontrolled field data might be poorly distributed which will decrease the prediction accuracy of trained models. There are two sampling strategies in surrogate-based optimization approach: space filling sampling (SFS) and sequential infilling sampling (SIS) methods. SFS distributes an initial samples into the design space to gain an overview of the response, while SIS places additional samples at specific locations based on the information derived from the current data. Latin hypercube sampling (LHS) is attractive among SFS strategies for expensive optimization because of an uniform distribution of the projection of samples onto each dimension without repetition (Giunta, Wojtkiewicz, and Eldred 2003). Many literatures (Jones, Schonlau, and Welch 1998, Shao and Krishnamurty 2008, Bhattacharya 2010) suggested 10 times of the number of variables as initial samples. However, the initial samples may not be sufficient and well distributed in expensive optimization, which results in a surrogate model with insufficient generality. Similar constraints appear if ill distributed field data has to be used as initial learning samples to save cost. How to utilize the inadequate surrogate to predict a reliable quasi-optimum and the strategy of additional samples to efficiently refine the model become important issues.

Sequential infilling sampling also called adaptive sampling strategy addresses how to distribute new samples to improve the regional accuracy of iterative surrogate model (Hu et al. 2017, Eason and Cremaschi 2014). Sequential sampling methods can be categorized as exploitation, exploration, or balanced exploitation/exploration approach (Forrester and Keane 2009). Sequential maximin Latin hypercube sampling method (Long et al. 2016) distributed additional samples to the less populated regions by maximizing the minimum distance between the new sample and existing ones to evenly explore the design space and improve the overall prediction generality. The sequential sampling strategies using maximizing expected improvement (EI) (Jones 2001, Parr et al. 2012) balancing exploitation and exploration, which has been proven to be an efficient infilling criterion. However, the sampling strategy tends to distribute additional samples in the area with lower sampling density, and the model optimum is not searched and verified during the iteration until the convergence of sampling. If the search has to be terminated prematurely, the initial improvement of optimum is not secured.

Prediction reliability is as important as optimality in expensive optimization. A surrogate from a small number of training samples might result in an inadequate model with lower generality, which leads to a false prediction of optimum or a trap to local optimum. The space surrounding learning samples of an approximate model often provides better prediction accuracy. Therefore, instead of searching the regions far away from sample sites which are liable to larger errors, it is more reliable to constrain the search in

Published in ENGINEERING OPTIMIZATION, 51(2): 247-264, 2019
https://doi.org/10.1080/0305215X.2018.1446084

the regions surrounding the training samples. A trust region is defined by a trust radius surrounding the updated optimum, which is initialized at a user defined value from an arbitrary point and iteratively updated to shrink or to expand the region depending on the prediction accuracy (Alexandrov et al. 1998). The trust region approach shows excellent convergence to a local optimum (Forrester and Keane 2009, Abdel-Malek, Ebid, and Mohamed 2017).

Younis and Dong proposed the space exploration and unimodal region elimination algorithm that divides the design space into key unimodal regions from the initial experimental data to identify the promising regions (Younis and Dong 2010). Additional samples using LHS are introduced to each of the unimodal regions to establish regional Kriging models to predict the local optima which are compared to determine the global optimum. However, the quality of the initial samples will be important to identify the promising regions, and the total number of samples increases significantly with the number of regions due to additional LHS samples.

Yu et al. adapted the trust region concept to population-based optimization and proposed the reliable regions of an artificial neural network model to be the union of hyper-spheres surrounding training samples (Yu, Liang, and Hung 2014). The searched quasi-optima serve as additional samples to refine the network model, and the prediction accuracy of the quasi-optimum will be applied to adjust the reliable radii of the regional neural network using a fuzzy inference. The searching and retraining processes iterate until the convergence of optimum. The definition of the reliable regions using hyper-spheres surrounding training samples is simple but lack of the consideration of the correlation among samples. The estimation error of the prediction provided by Kriging modelling will provide a more realistic reference to define the reliable region (Suprayitno 2016). Also, the vicinity search surrounding the samples won't guarantee a global optimum if no initial sample is in the global basin of attraction. To locate the true global optimum, exploration samples are essential in infilling sampling for sequential approximate optimization.

This work presents a novel evolutionary framework with an adaptive sampling strategy for expensive optimization. The sequential surrogates for engineering applications are liable to low generality sequential approximate optimization due to constraint of sample size and system nonlinearity. The estimation error provided by Kriging will be applied to define the reliable region of surrogate that is dynamically updated according to the prediction accuracy. Evolutionary optimizer is introduced to search for the quasi-optimum in the reliable region to improve the search efficiency. A hybrid infilling strategy is also introduced to deploy additional samples to ensure sampling efficiency and global optimality during the evolution of surrogate model.

## 2. Evolutionary Optimization Using Reliable Regional Kriging Surrogate.

This study presents a novel evolutionary algorithm based on the iterative constrained search in the progressive regional surrogate defined by Kriging model and prediction error. **Figure 2** shows the flowchart of the proposed scheme, the evolutionary optimization using reliable regional Kriging surrogate (EORKS). A Kriging based surrogate model is first established from small initial samples using DOE. To improve the search reliability the Kriging standard error is applied to define the reliable region to constrain an evolutional optimizer for a quasi-optimum. The quasi-optimum will become the additional infilling sample unless the updated quasi-optimum converges to the current designs, and the infilling strategy will relocate an alternative sample with the maximum expected improvement (EI). The additional sample will be augmented to the learning samples to refine the surrogate. The prediction accuracy of the surrogate for the sequential sample will also be applied to intelligently evolve the regional model using a fuzzy inference. The process iterates until the convergence to the global optimum. The detailed process of EORKS will be illustrated as follows.

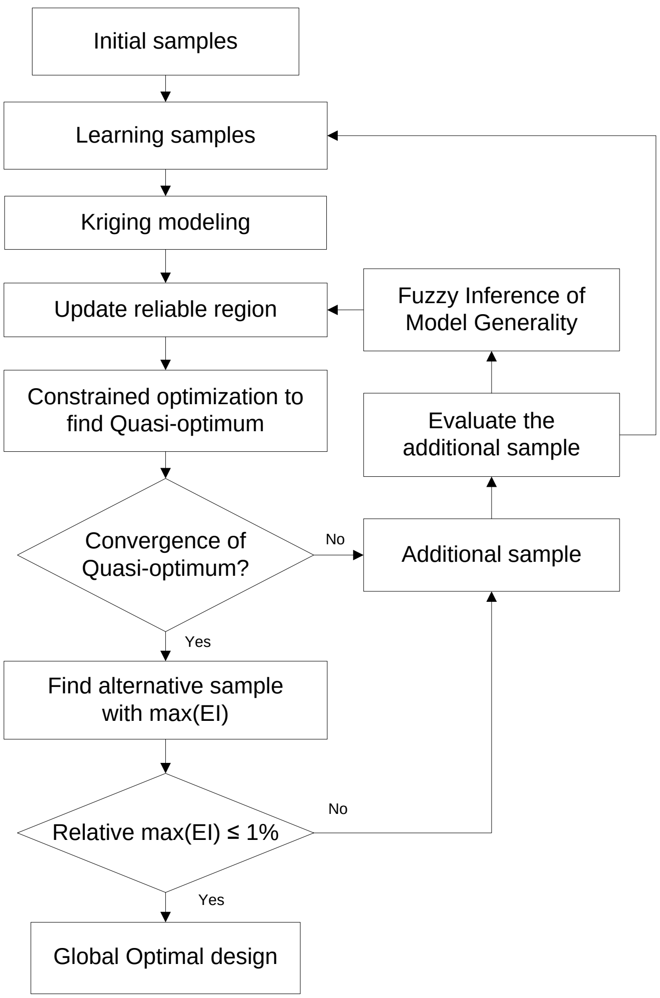



**Figure 2** Schematic flow chart of proposed optimization algorithm EORKS.

### *2.1. Reliable region of an inadequate Kriging surrogate*

The number of samples is often limited for expensive optimization. LHS strategy is suggested if DOE is available since LHS provides a small and symmetric distribution of samples. The number of LHS initial samples is assumed two times of the design variables in this work. However, how to address the low generality of an inadequate surrogate due to a small set of learning samples is a practical issue, especially if a reliable quasi-optimum is expected form the inadequate surrogate in a finite iterations. Similar challenges occur and become more critical if the use of unplanned field data is applied as training samples; even the field data might be sparse and ill distributed.

Kriging, known as Gaussian process modelling, assumes a response function composed of a regression model and a stochastic process, as in (1).

$$Y(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{\beta} + Z(\boldsymbol{x}) \qquad (1)$$

where $\boldsymbol{f}(\boldsymbol{x}) = [f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_p(\boldsymbol{x})]^T$ is a vector of regression functions; $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots \beta_p]^T$ is a vector of unknown coefficients, and $Z(\boldsymbol{x})$ is a stochastic process with zero mean and nonzero covariance $\sigma^2 R(\boldsymbol{x}, \boldsymbol{x}')$. The regression model $\boldsymbol{f}(x)$ often adopts a constant or low-order polynomials. From a preliminary study, first order polynomial is sufficient for most numerical benchmark test functions.

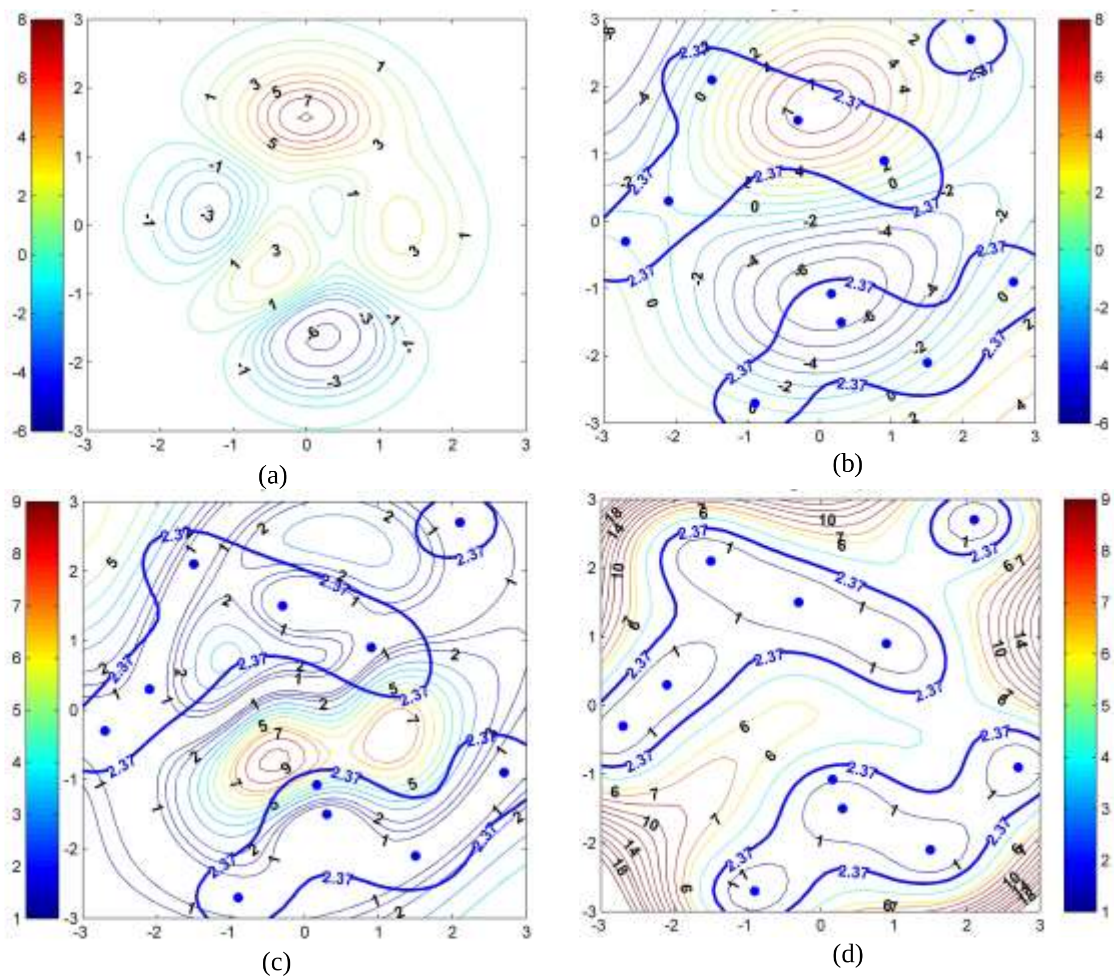



**Figure 3** Illustration of the prediction error using Kriging model (a) True response surface of peaks function, (b) Prediction response surface from Kriging model using 10 LHS learning samples, (c) True error contour plot of Kriging model with an overlaid exemplar reliable region, and (d) Kriging standard error map.

For the example of a two dimensional Peaks function as in (2), the theoretical contour plot is highly non-linear which includes 3 peaks and 2 valleys and flattens out gradually as shown in **Figure 3**(a). **Figure 3**(b) is the response surface of a Kriging surrogate established from 10 LHS samples. First order polynomial is selected as the basic functions. The surrogate model simulates a rough profile of the true function, but significant errors present especially for the prediction farther away from the samples. Increasing the number of samples seems a simple solution, but will raise a cost concern. How could we make the most out of this inadequate surrogate and infer a quasi-optimum from the current model are important in applications.

$$z = 3(1-x)^2 e^{-x^2-(y+1)^2} - 10\left(\frac{x}{5} - x^3 - y^5\right) e^{-x^2-y^2} - \frac{1}{3} e^{-(x+1)^2-y^2} . \quad (2)$$

The unique feature of Kriging modelling is the provision of measuring possible error of the prediction. The Kriging error variance of the predictor (Jones 2001) is shown as in (3),

$$s^2(\boldsymbol{x}) = \hat{\sigma}^2 \left[1.0 - \boldsymbol{r}^T \boldsymbol{R}^{-1} \boldsymbol{r} + \frac{\left(1.0 - \boldsymbol{1}^T \boldsymbol{R}^{-1} \boldsymbol{r}\right)^2}{\boldsymbol{1}^T \boldsymbol{R}^{-1} \boldsymbol{1}}\right] . \quad (3)$$

where **1** is a vector filled with ones. ***R*** and ***r*** are the correlation matrix and the correlation vector, respectively. Gaussian correlation function is widely used with the form as in (4),

$$R(\boldsymbol{x}, \boldsymbol{x}') = e^{-\sum_{k=1}^{n} \theta_k |x_k - x'_k|^2} . \quad (4)$$

where $\theta_k$ are the correlation parameters from the maximum likelihood estimation , $n$ is the number of design variables, and $|x_k - x'_k|$ is the Euclidean distance between any two sites $\boldsymbol{x}$ and $\boldsymbol{x}'$. Kriging error variance goes to zero at the sampled points, and becomes larger as the location is farther away from the sampled points.

**Figure 3**(c) shows the true error between the Peaks function in **Figure 3**(a) and the Kriging approximation in **Figure 3**(b). Searching the entire space of an inadequate surrogate might lead to a false optimum due to limited generality. Despite insufficient global generality, the standard error provided by the Kriging estimation as in **Figure 3**(d) shows a similar tendency to the true error in **Figure 3**(c) in the proximity of learning samples. An exemplar reliable region based on the predicted standard error of 2.37 is overlaid on the true error contour plot as shown in **Figure 3**(c). A constrained search in the neighbouring regions surrounding the samples will provide a more reliable quasi-optimum than a global search.

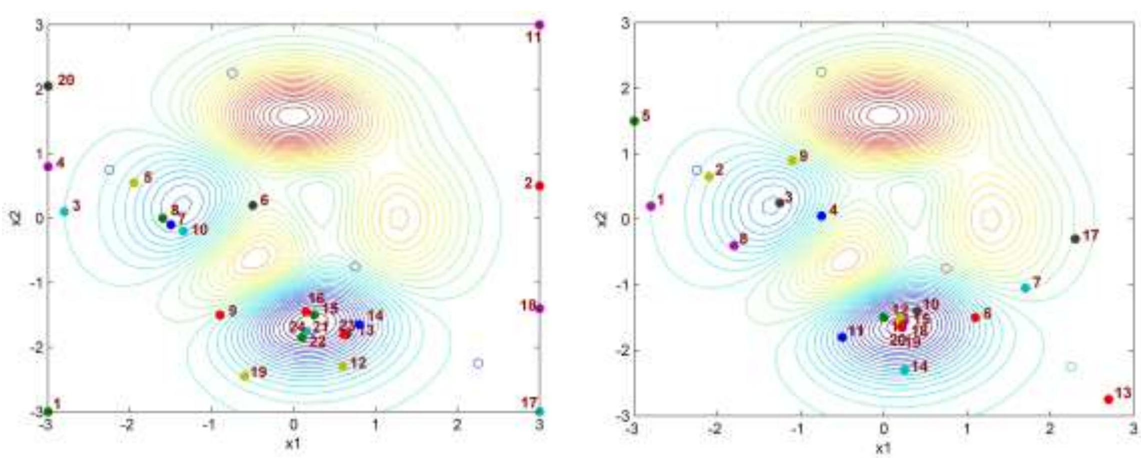
This study proposes the reliable regions of a Kriging surrogate as the space with a Kriging standard error smaller than a weighted function of the maximum Kriging standard error predicted as shown in (5).

$$\begin{gathered} Reliable\ Regional\ Kriging\ Surrogate\ (RKS) = \\ \{\boldsymbol{x} \mid Krigimg\ stdev.(\boldsymbol{x}) \leq [C \times max(Kriging\ stdev.)]\} \end{gathered} \quad (5)$$

The coverage coefficient $C$ varies from 0 to 1. The maximum Kriging standard error can be obtained from the search of (3) in the design space. A small $C$ will lead to a conservative search surrounding the samples, while a large $C$ will explores a larger portion of the surrogate but might lead to an erroneous prediction if the surrogate is inadequate. The selection of the coverage coefficient $C$ depends on the system complexity and model accuracy. A fuzzy inference system for the evolution of $C$ will be introduced based on the prediction accuracy of the additional sample.

### *2.2. Quasi-optimum search using evolutionary optimizer*

Evolutionary optimisers, such as GA and PSO, can be applied to capture the optimum of the Kriging surrogate. The searched optimum is called a quasi-optimum since the search is confined to the reliable regions. As a population-based algorithm, a diverse solution is essential in GA to avoid premature convergence and slow convergence (Gupta and Ghafir 2012). The Kriging standard errors of the offspring are calculated using (3) and compared with (5) to make sure a constrained evolutionary search in the reliable region. The verified result of the searched optimum is added to the learning samples to refine the surrogate in an iterative fashion until convergence. The comparison of the surrogate-based optimization using the constrained GA search in the reliable regions and the conventional search in the complete surrogate illustrates the influence of inadequate surrogate on the optimum search. The initial samples are denoted as hollow dots in **Figure 4**(a). The iterative optima provided from the conventional approach scattered to unlikely regions such as iterations 1, 2, and 11 due to limited generality of the surrogate. On the other hand, the distribution of optima using the reliable regions inclined toward trustworthy and promising area which leads to a faster convergence as shown in **Figure 4**(b).



(a) (b)

**Figure 4** Distributions of iterative predicted optima over the true contour plot of Peaks function using surrogate-based GA search (a) in the whole regions and (b) in the updated reliable regions.

A similar process can be applied to search for the quasi-optimum of the reliable regional Kriging surrogate using any preferred evolutionary optimizers. PSO is inspired by social interaction in human beings and animals such as the behaviour of bird flocking and fish schooling to guide the particles to search for the global optimum. The particles fly over the searching space with dynamic velocities typically composed of three components: inertia velocity, memory, and social knowledge (Rezaee Jordehi and Jasni 2013). The position of particles are updated with the weighted sum of the three components using the coefficients $w$, $c_1$, and $c_2$ respectively.

To demonstrate that the proposed evolutionary reliable regional Kriging surrogate can be readily adapted to the selected optimizers, the comparison of the optimization of 2-D sphere function and Branin function using GA and PSO are presented in **Table 1**. The corresponding parameters for both optimizers are listed in **Table 2.** The contour surface of the 2-D sphere function is very smooth with a single optimum, while the Branin function is multi-modal. The optimization results of 10 runs show that the applications of EORKS with PSO and GA have comparable performance in both the best solution and total number of required samples. Since the optimizer is applied to search the regional surrogate, the computational cost is small compared with the actual samples required to establish the surrogate which often involve expensive experiments and simulations. In general, PSO is more efficient to locate the quasi-optimum of the regional surrogate compared with GA but liable to trap to local optimum if the initial selection of particles is ill distributed. Therefore, GA is adopted as the optimizer in the following study.

**Table 1** Comparison of the applications of EORKS using GA and PSO as the optimizer for 2-D sphere function and Branin function

| | EORKS - PSO | | EORKS - GA | |
|---|---|---|---|---|
| | Optimum | Total no. of samples | Optimum | Total no. of samples |
| *2-D sphere function* | | | | |
| Average[a] | 2.54E-07 | 15.90 | 1.29E-06 | 15.80 |
| Standard Deviation[a] | 2.06E-07 | 0.74 | 1.65E-06 | 0.63 |
| *Branin function* | | | | |
| Average[a] | 0.3979 | 26.60 | 0.3979 | 25.40 |
| Standard Deviation[a] | 0.0000 | 3.86 | 0.0000 | 2.76 |

[a] Optimization results in 10 runs

**Table 2** Parameters setting for GA and PSO optimizers

| Real number GA | Adaptive PSO * |
|---|---|
| Initialization:<br>Number of population: 10 × number of problem variables<br>Number of new offspring: 0.8 × number of population<br>Maximum generation: 1000<br>Parents selection method: roulette wheel selection<br>Crossover: single point crossover<br>Mutation: one point mutation with 0.3 mutation rate<br>New generation: Elitism method – preserve the best two instances from the parent population | Swarm size: 10 × number of problem variables<br>Maximum iteration: 1000<br>Initial inertia weight: $w$ = 0.9<br>Initial acceleration: $c_1$=$c_2$=2.0 |

* Adaptive strategy from (Zhan et al. 2009)

Published in ENGINEERING OPTIMIZATION, 51(2): 247-264, 2019
https://doi.org/10.1080/0305215X.2018.1446084

### *2.3. Evolution of reliable region*

The coverage coefficient *C* will determine the reliable region and the searching space of surrogate for optimizer, which depends on the system complexity and surrogate accuracy. This study presents an automated update mechanism starting from an arbitrary initial selection. The searched quasi-optimum will serve as a test sample, and the prediction error of the test sample will become a feedback mechanism to adjust the coverage coefficient *C*. The model reliability is measured by the relative prediction error of the test sample as in (6).

$$Relative\ prediction\ error = abs\left(\frac{y_{verified} - y_{prediction}}{y_{verified}}\right). \tag{6}$$

If the relative prediction error is small, the generality of the model is good and the reliable region will be expanded to explore a larger space, otherwise the reliable region should be reduced for a conservative search of the quasi-optimum in the next iteration. A heuristics based fuzzy inference scheme is proposed as follows to dynamically update the reliable region based on the model reliability.

1. If the relative error of prediction is Small then Increase the coverage coefficient.
2. If the relative error of prediction is Moderate then Maintain the coverage coefficient.
3. If the relative error of prediction is Large then Decrease the coverage coefficient.

Standard membership functions are used as illustrated in **Figure 5** for the three condition levels of relative prediction error: Small, Moderate, and Large, and for the three assessment levels of adjustment: Increase, Maintain, and Decrease. A simple centre average defuzzifier is applied to derive the adjusting factor (*AF*) for the coverage coefficient *C* as in (7). The adjustment factor will be between 0.8 and 1.2 depending on the prediction accuracy.

$$C_{i+1} = AF_i \times C_i\ . \tag{7}$$

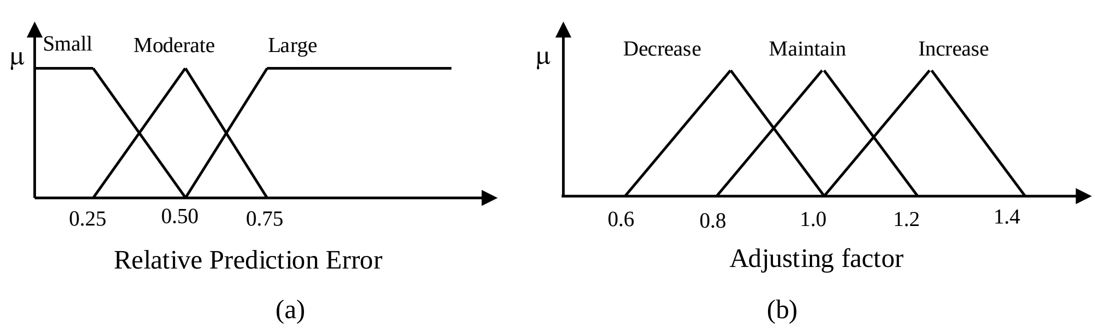
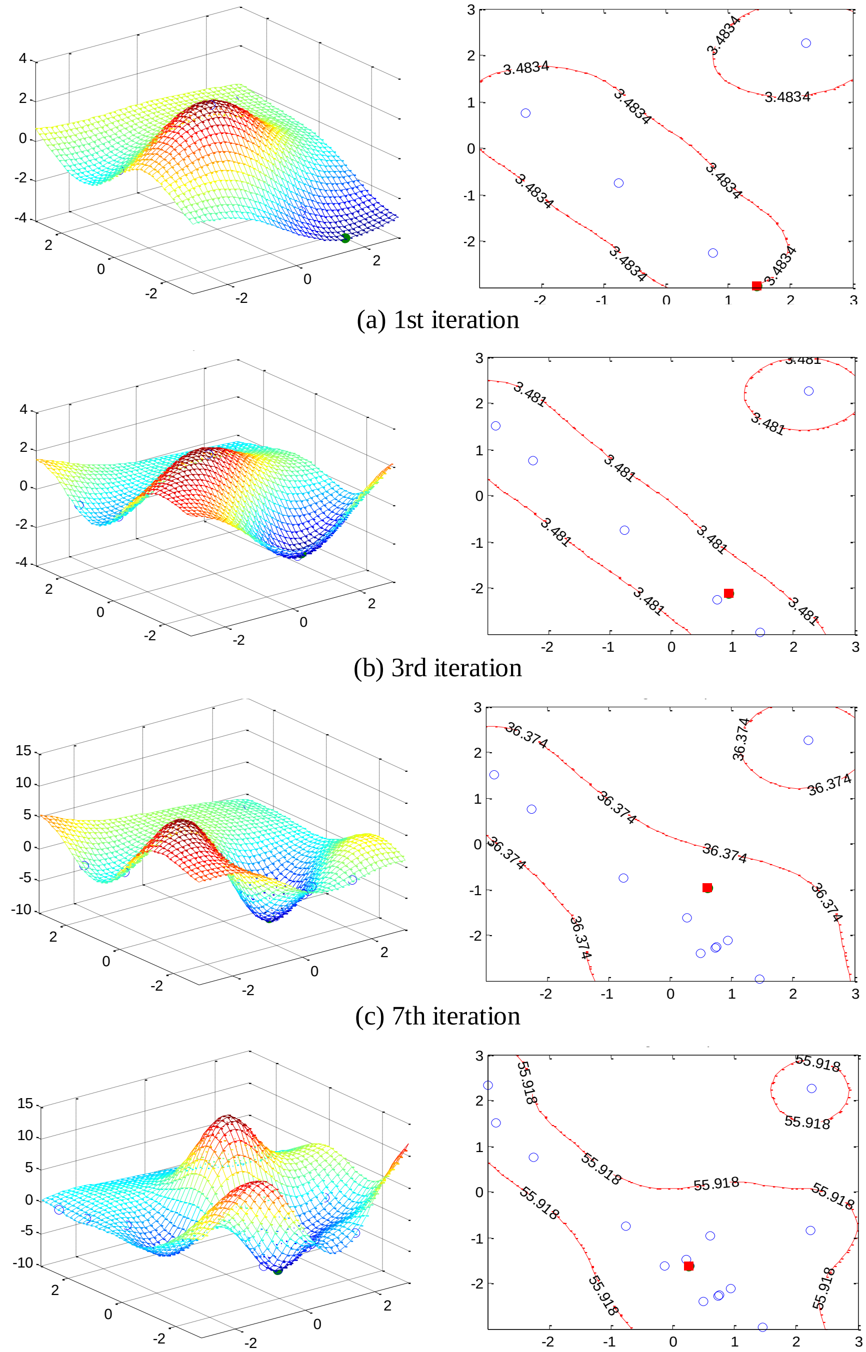



**Figure 5** Membership functions for the fuzzy inference of reliable region (a) prediction error (b) adjusting factor

Published in ENGINEERING OPTIMIZATION, 51(2): 247-264, 2019
https://doi.org/10.1080/0305215X.2018.1446084




(a) 1st iteration

(b) 3rd iteration

(c) 7th iteration

(d) 13th iteration

**Figure 6** Sequential surrogate models and the corresponding evolution of reliable regions for Peaks function

The self-learning mechanism of the reliable region using the fuzzy inference will automatically adjust the size of reliable regions and guide the searching range of optimizer according to the surrogate generality, and reduces the sensitivity of the initial selection of *C*. The initial selection of the coverage coefficient arbitrarily assumes 0.2 in this study. The response surfaces in **Figure 6** show how the proposed scheme progressively refined the

Published in ENGINEERING OPTIMIZATION, 51(2): 247-264, 2019
https://doi.org/10.1080/0305215X.2018.1446084

Kriging surrogate by introducing additional learning sample, and the contour plots in **Figure 6** show how the fuzzy inference dynamically updated the reliable regions.

## *2.4. Hybrid infilling sampling strategy*

Global generality of prediction accuracy is not necessary for surrogate based optimization. Since sampling efficiency is crucial to expensive optimization, additional samples are preferred in the promising areas of optimum. This study suggests a smaller set of initial training samples and distributes additional samples identified during the iteration of optimization to progressively and selectively refine the surrogate. The proposed optimization scheme, EORKS constrains the optimizer search in the reliable regions to improve the prediction accuracy of the quasi-optimum from the possibly inadequate surrogate. Additional training samples from the true function calls or experiments of the derived quasi optimums are commonly used to exploit promising regions. However, if the additional samples adopt only the quasi-optimums which are constrained to the vicinity of training samples, the search is liable to be trapped to a local optimum. On the other hand, using an exploration approach only might lead to incorrect convergence due to smooth out the best problem. A hybrid infilling sampling strategy considering both exploitation and exploration is introduced in this study. The exploitation approach using verified quasi-optimum is combined with exploration approach using maximum expected improvement criteria.

From the evolutionary search of the reliable region, calls to the true function for the quasi-optimum is added to the training samples to refine the surrogate. If the predicted quasi-optimum is converged to current samples, an alternative sample site with the maximum expected improvement of the current surrogate will replace the quasi-optimum in the next verification and retraining process. Kriging model assumes the prediction $y(\boldsymbol{x})$ is a stochastic distribution centred on $\hat{y}(\boldsymbol{x})$ with a variance of $s^2(\boldsymbol{x})$. The definition of expected improvement for a search of minimum is illustrated in **Figure 7**. If the best observed objective of the current samples is $y_{\min}$, considering the Kriging prediction is normally distributed, the probability of improvement $I(\boldsymbol{x}) = y_{\min} - y(\boldsymbol{x})$ is the area enclosed by the Gaussian distribution below $y_{\min}$ as in **Figure 7**. The expected improvement, $E[I(\boldsymbol{x})]$, can be estimated as follows.

$$E[I(\boldsymbol{x})] = \begin{bmatrix} \frac{1}{s\sqrt{2\pi}} \int_{-\infty}^{y_{min}} [y_{min} - y(x)] \times e^{\frac{-[y(\boldsymbol{x})-\hat{y}(\boldsymbol{x})]^2}{2s^2}} dy & if\ s > 0 \\ 0 & if\ s = 0 \end{bmatrix} \quad (8)$$

$E[I(\boldsymbol{x})]$ will be zero at all sample sites due to the nature of Kriging standard error. In case the maximum expected improvement criterion is activated, the coverage coefficient will maintain the value in the last iteration.

Published in ENGINEERING OPTIMIZATION, 51(2): 247-264, 2019
https://doi.org/10.1080/0305215X.2018.1446084

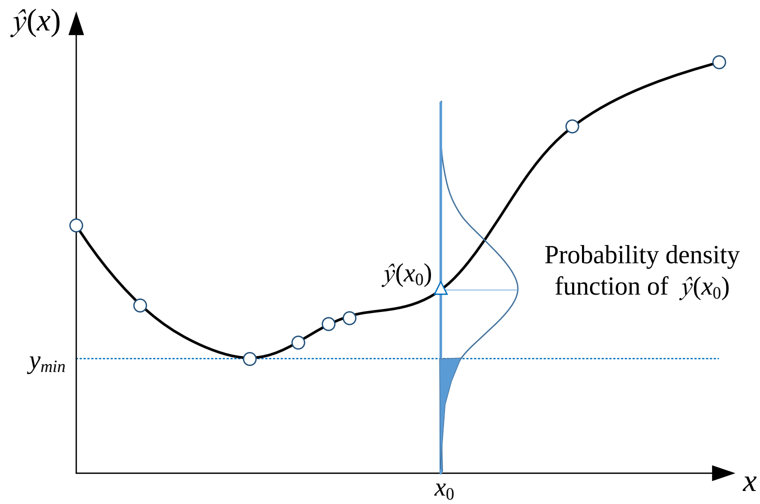



**Figure 7** Probability of improvement for Kriging prediction

### *2.5. Iteration for optimum*

To cope with sparse and ill distribution of learning samples in expensive optimization, the proposed scheme establishes a reliable regional Kriging surrogate based on the prediction error of Kriging modelling. The quasi-optimum searched in the reliable regions will provide a more reliable improvement from the inadequate surrogate. To reduce the cost of additional training samples, only one additional sample from the hybrid infilling sampling strategy is introduced to the learning samples in each iteration. The augmented learning samples progressively evolve the reliable regional Kriging surrogate based on the fuzzy inference of the prediction generality. The accuracy of the surrogate model will then improve in the promising area of optimum, which ensures the sample efficiency and the robustness of searched optimum. The stopping criteria include the convergence of the quasi-optimum for three consecutive times and the maximum expected improvement less than 1% of the current optimum. Sometimes, premature stop is possible in engineering applications due to cost constraint and insignificant increment of improvement within tolerance.

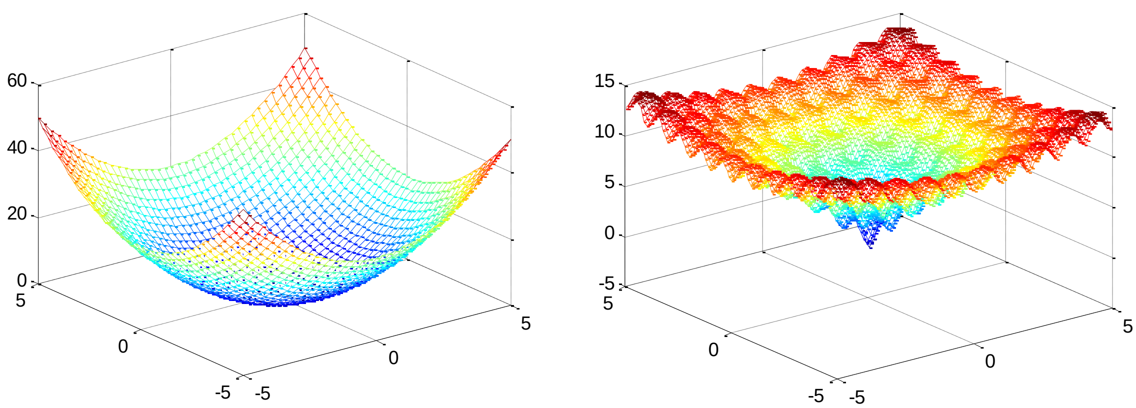
## 3. Benchmark Test Functions

Six benchmark functions are applied to compare the sampling efficiency and the searched results using the proposed optimization algorithm, EORKS, with those using the surrogate assisted evolutionary algorithms in literature  (Jones, Schonlau, and Welch 1998). Branin and Goldstein-Price functions are two-dimensional. Hartman 3-D and Hartman 6-D are 3-dimensional and 6-dimensional functions respectively. For the numerical problems in higher dimensions, the optimization of two extreme benchmark test functions, the Sphere function and the Ackley function in 20-dimension, are applied to compare with the literature results. The Sphere function, as shown in **Figure 8** (a) of a two-dimensional

Published in ENGINEERING OPTIMIZATION, 51(2): 247-264, 2019
https://doi.org/10.1080/0305215X.2018.1446084

example, is a simple and smooth contour, which has only one single global optimum and no local optimum. On the other hand, the Ackley function is a 'noisy' function, as shown in **Figure 8** (b) of a two-dimensional example, which has multiple steep valleys of one global optimum and numerous local optimums. Though these response surfaces are unlikely to happen in real world optimization problems, these two extreme examples are used to test the robustness of the proposed methods.

The optimization of Branin, Goldstein-Price, Hartman 3-D, and Hartman 6-D functions comparison are compared with the results using EGO (Jones, Schonlau, and Welch 1998). For Goldstein-Price and Hartman 6-D functions, the log-transformed forms are used as suggested in the reference. EGO is a surrogate-based optimization approach using Kriging model and the maximum expected improvement as the sequential sampling strategy. The initial size of samples used in EGO adopts 10 time of the dimension of the problem, while EORKS started from a smaller initial sample of 2 times of the dimension of the problem. The comparison in **Table 3** shows that EORKS required much smaller sets of samples than EGO did to reach the optimum with less than 1% error for all the four benchmark test functions. In average, EORKS reduces the number of samples by 39% in comparison with EGO.




(a) (b)

**Figure 8** Benchmark test functions (a) 2-D Sphere function (b) 2-D Ackley function.

**Table 3** Comparison of the sampling efficiency for EGO and EORKS algorithms

| Test function | Method | No. of Initial samples | No. of sequential samples required to derive the optimum with <1% error | Total no. of samples |
|---|---|---|---|---|
| Branin | EGO* | 21 | 28 | 49 |
| | EORKS | 4 | 36 | 40 |
| Goldstein-Price | EGO* | 21 | 32 | 53 |
| | EORKS | 4 | 26 | 30 |
| Hartmann 3-D | EGO* | 33 | 35 | 68 |
| | EORKS | 6 | 37 | 43 |
| Hartmann 6-D | EGO* | 65 | 121 | 186 |
| | EORKS | 12 | 68 | 80 |

* Optimization results in (Jones, Schonlau, and Welch 1998)

Published in ENGINEERING OPTIMIZATION, 51(2): 247-264, 2019
https://doi.org/10.1080/0305215X.2018.1446084

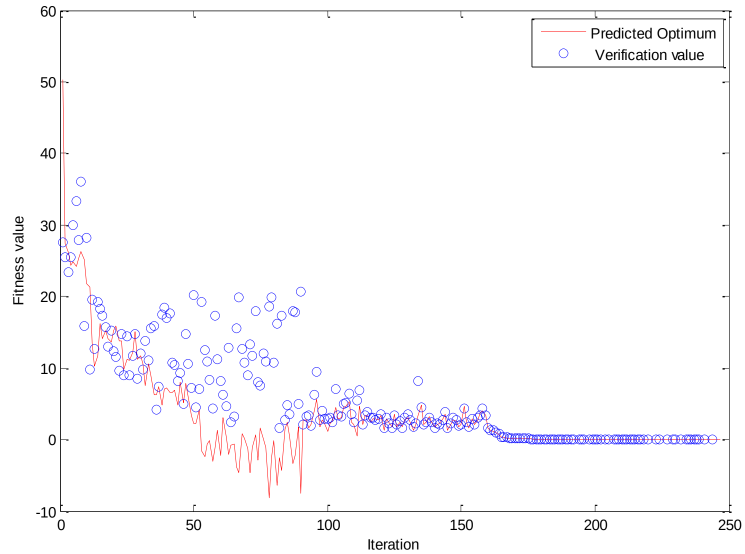



**Figure 9** Iteration results of the 20-D Sphere function

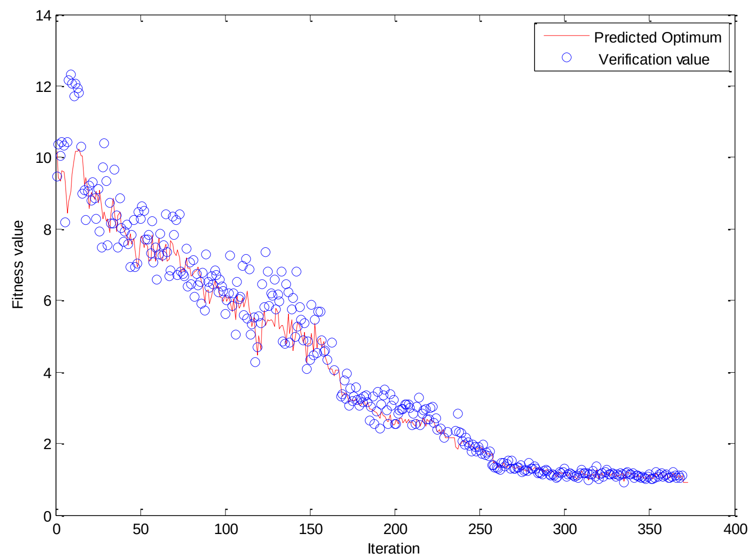



**Figure 10** Iteration results of the 20-D Ackley function

**Figure 9** and **Figure 10** show the iteration results of the Sphere function and Ackley function of 20-dimension using EORKS. The initial surrogates are established from -LHS with the number two times of design variables which are 40 samples in these examples. Continuous solid line represents the iteration result of predicted quasi-optima, while the hollow dots are the verified responses of the quasi-optima. The quasi-optima have larger prediction errors in the early iterations due to lack of generality of the initial surrogate model. However, the discrepancy stabilized due to the use of reliable regions. The quasi-optimum converges smoothly to the global optimum. The iteration curve of the Ackley

Published in ENGINEERING OPTIMIZATION, 51(2): 247-264, 2019
https://doi.org/10.1080/0305215X.2018.1446084

function is more severe than the Sphere function as expected, and requires more samples to converge because the response surface of Ackley function contains many sharp local optima around the global optimum.

The Sphere function is very smooth, and has a global optimum at the origin with the response of zero. Bhattacharya (Bhattacharya 2010) used support vector machine to build the surrogate model for the 20-dimensional (20-D) Sphere function, and applied the preference learning technique with genetic algorithm for the sequential training and search. The results shown in **Table 4** are the average performance and the average number of function calls of three optimization runs. EORKS provides a superior optimum using 57% less number of the true function calls in average compared with the literature result.

Similar advantages can be observed for the optimization of the 20-D Ackley function in comparison with the result in the reference (Jin, Olhofer, and Sendhoff 2002) that adopted artificial neural network surrogate with weighted online learning using the covariance matrix in combination with an adaptive genetic algorithm search. **Table 4** compares the performance and the required function calls of ten runs using EORKS with the results in the cited reference. The average best fitness is 0.81 and the average number of function calls is 1483 in the reference. On the other hand, the optimal fitness provide by EORKS is 0.331 that outperforms the literature result using only 44% of the function calls.

**Table 4** Comparison of the optimization results for 20-dimensional Sphere function and Ackley function

| Test Function | Optimization Method | Best Fitness | | Function calls | |
|---|---|---|---|---|---|
| | | Average | Standard Deviation | Average | Standard Deviation |
| Spherical 20-D* | Literature Results | 1.013E-07 | n/a | 2750 | n/a |
| | EORKS | 9.820E-10 | 4.211E-10 | 1172 | 39 |
| Ackley 20-D** | Literature Results | 0.81 | 0.32 | 1483 | 77.8 |
| | EORKS | 0.33 | 0.37 | 654 | 10.1 |

* Optimization result based on 3 runs and compare with in (Bhattacharya 2010)
** Optimization result based on 10 runs in (Jin, Olhofer, and Sendhoff 2002)

## 4. Truss optimization

The optimization of a truss structure (Shao and Krishnamurty 2008) is used to illustrate the engineering application. **Figure 11** shows the schematic sketch of the initial design of the truss structure. The truss is 12 meter long with a maximum height of 3 meter, and is constructed of 21 straight and uniform steel bars pinned together. The circular cross sectional area of all truss members is 25 cm$^2$. The design objective is to maximize the truss efficiency, which is defined as the ratio of the maximum load ($F_{max}$) that the truss can bear to the self-weight of the truss ($W$). The load $F$ reaches the maximum value when any one of the 21 bars reach the following constraints:

(1) The induced maximum tensile stress $\sigma_t$ > 200 MPa,
(2) The induced maximum compression stress $\sigma_c$ > 200 MPa, and,
(3) The stability ratio RBK of the maximum compression stress ($\sigma_{c,i}$) to the associated critical buckling stress ($S_{cr,i}$) of bar $i$ is greater than 1.

Published in ENGINEERING OPTIMIZATION, 51(2): 247-264, 2019
https://doi.org/10.1080/0305215X.2018.1446084

Here, $S_{cr,i} = \pi^2 E/(L_{e-i}/\rho)^2$, whereas $E$ is Young's modulus of the truss material, $\rho$ is the radius of gyration of the truss bar, and $L_{e-i}$ are the equivalent length of $i$-th truss member, $i$ =1, 2, ..., 21, which is equal to the length of the corresponding pinned truss member (Juvinall and Marshek 2006, 228). The truss is fixed at its left end support and simply supported at its right end. Since the truss is symmetric, the seven locations of its joints are considered as design variables including the horizontal positions of joints 1 ($x_1$) and 2 ($x_2$), the vertical and horizontal positions of joints 3 ($x_3$, $y_3$) and 4 ($x_4$, $y_4$), and the vertical position of joint 6 ($y_6$). The optimization problem can be formulated as follows;

Maximize $F_{max}/W = f(x_1, x_2, x_3, y_3, x_4, y_4, y_6)$

Subject to $\{\sigma_t, \sigma_c\} \leq \sigma_a$ (= 200 MPa)

$$\frac{\sigma_{c,i}}{S_{cr,i}} \leq 1$$

$$0.8 \leq \{x_1, x_3, y_3, y_4, y_6\} \leq 3.0 \text{ m}$$

$$3.3 \leq \{x_2, x_4\} \leq 5.5 \text{ m}$$

There is no analytical solution for the truss efficiency, and the maximum load has to be obtained using finite element analysis (ANSYS) which is a computationally expensive problem. Shao and Krishnamurty proposed a clustering-based multilocation search (CMLS) to solve the problem using 70 initial sampling points and add 25 extra samples after five iterations to converge to an optimum. The proposed method starts from a rough Kriging model with 14 initial samples distributed using LHS method as shown in **Table 5**. EORKS converged in 73 iterations. The maximum load of optimum design is 231.74 kN while its weight is 9022.4 N. Maximum tensile stress and maximum compressive stress are 83.47 MPa and -107.49 MPa respectively. The largest stability ratio, $R_{BK,}$ due to critical buckling stress is 0.99 which also satisfies the design constraint. The optimization results for truss efficiency and cost of computation are compared with the best design of initial samples and the result in the literature as listed in **Table 6** and illustrated in **Figure 12.** The truss efficiency increases by 8.4% and reduces the number of samples by 12.6% using EORKS compared with the result in the literature.

**Table 5** LHS Initial Samples for the Truss Optimization

| Sample | Design variables | | | | | | | $F_{max}/W$ |
|---|---|---|---|---|---|---|---|---|
| | $x_1$ | $x_2$ | $x_3$ | $y_3$ | $x_4$ | $y_4$ | $y_6$ | |
| 1 | 0.879 | 5.421 | 1.821 | 1.350 | 3.379 | 2.607 | 2.450 | 4.511 |
| 2 | 1.821 | 4.321 | 2.607 | 1.193 | 4.793 | 1.821 | 1.664 | 8.615 |
| 3 | 1.350 | 4.636 | 1.507 | 1.979 | 5.421 | 2.921 | 0.879 | 4.705 |
| 4 | 2.921 | 4.007 | 2.921 | 1.664 | 3.850 | 2.293 | 1.193 | 4.962 |
| 5 | 2.293 | 3.693 | 1.193 | 2.293 | 3.536 | 2.764 | 2.293 | 11.224 |
| 6 | 1.193 | 3.379 | 2.293 | 1.036 | 5.107 | 1.979 | 2.607 | 2.975 |
| 7 | 2.136 | 3.850 | 0.879 | 1.507 | 4.479 | 1.507 | 2.136 | 7.148 |
| 8 | 2.607 | 4.950 | 1.979 | 2.136 | 4.321 | 1.036 | 2.921 | 17.035 |
| 9 | 2.764 | 4.479 | 2.136 | 1.821 | 5.264 | 2.136 | 2.764 | 12.059 |
| 10 | 1.979 | 5.264 | 2.450 | 2.921 | 4.636 | 2.450 | 1.821 | 7.288 |
| 11 | 1.036 | 4.164 | 2.764 | 2.450 | 4.007 | 0.879 | 1.507 | 5.939 |
| 12 | 1.664 | 3.536 | 1.664 | 2.607 | 3.693 | 1.664 | 1.350 | 9.053 |
| 13 | 2.450 | 5.107 | 1.350 | 0.879 | 4.164 | 1.350 | 1.036 | 9.550 |
| 14 | 1.507 | 4.793 | 1.036 | 2.764 | 4.950 | 1.193 | 1.979 | 6.330 |

Published in ENGINEERING OPTIMIZATION, 51(2): 247-264, 2019
https://doi.org/10.1080/0305215X.2018.1446084

**Table 6** Optimal designs of the truss example

| Result | Design Variables | | | | | | | $F_{max}/W$ | No. of Sample |
|---|---|---|---|---|---|---|---|---|---|
| | $x_1$ | $x_2$ | $x_3$ | $y_3$ | $x_4$ | $y_4$ | $y_6$ | | |
| Initial | 2.607 | 4.950 | 1.979 | 2.136 | 4.321 | 1.036 | 2.921 | 17.04 | 14 |
| CMLS | 2.370 | 4.174 | 1.877 | 1.721 | 4.024 | 1.438 | 2.186 | 22.79 | 95 |
| EORKS | 3.000 | 4.209 | 1.902 | 2.285 | 4.222 | 1.709 | 2.644 | 25.66 | 87 |

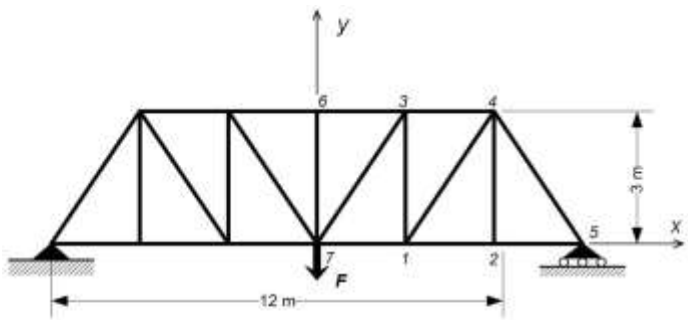



**Figure 11** Initial design of the truss optimization problem

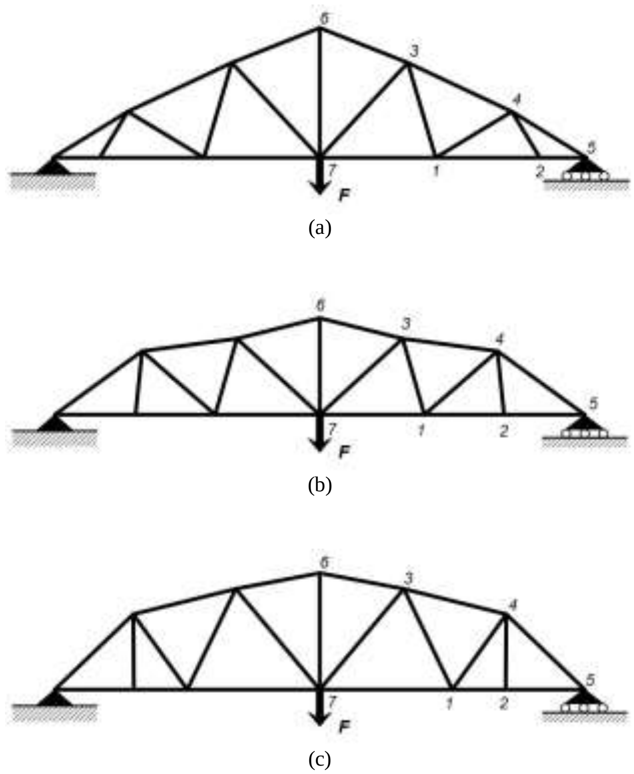



**Figure 12** Truss optimization results (a) The best design from initial samples, (b) Optimum design from CMLS method, (c) Optimum design from EORKS method

Published in ENGINEERING OPTIMIZATION, 51(2): 247-264, 2019
https://doi.org/10.1080/0305215X.2018.1446084

## 5. Conclusions

The proposed optimization algorithm, EORKS, starts from a Kriging based surrogate model using a small number of samples, and constraints the optimizer search within the evolving reliable region defined by a dynamic allowable prediction error. The constrained search provides a reliable quasi-optimum even the surrogate model is inadequate in global generality due to sparse samples. The reliable region evolves dynamically using the fuzzy inference with the addition of new sample and the verification of predication accuracy of the surrogate model. The reliable regional Kriging surrogate adaptive and the infilling sampling strategy combining the exploitation strategy and the maximum expected improvement criteria have shown superior sampling efficiency and searching robustness in various benchmark test functions. EORKS provides superior results even in high-dimensional benchmark functions for both a smooth function such as 20-D Sphere function and a noisy function such as 20-D Ackley function at a much less sampling cost. A satisfactory quasi-optimum at a lower sampling cost is preferred considering the tolerances of variables in engineering applications. The truss optimization example demonstrates the optimality and sampling efficiency of the proposed algorithm, which makes EORKS attractive in expensive optimization applications.

## 6. Acknowledgment

This work was supported by Ministry of Science and Technology, TAIWAN, ROC, under Grants 104-2221-E-327-025 and 105-2221-E-327-020-MY2; and Indonesian DIKTI Scholarships 2013 under Grant No. 145.17/E4.4/2013

## 7. References

1. Abdel-Malek, Hany L, Shaimaa EK Ebid, and Ahmed SA Mohamed. 2017. "A trust region approach with multivariate Padé model for optimal circuit design." *Engineering Optimization*:1-17.
2. Alexandrov, Natalia M, John E Dennis Jr, Robert Michael Lewis, and Virginia Torczon. 1998. "A trust-region framework for managing the use of approximation models in optimization." *Structural optimization* 15 (1):16-23.
3. Bhattacharya, Maumita. 2010. "An investigation on two surrogate-based EAs." *Australian Journal of Intelligent Information Processing Systems, ISSN*:1321-2133.
4. Eason, John, and Selen Cremaschi. 2014. "Adaptive sequential sampling for surrogate model generation with artificial neural networks." *Computers & Chemical Engineering* 68:220-232.
5. Eberhart, Russell, and James Kennedy. 1995. "A new optimizer using particle swarm theory." Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on.
6. Elsayed, Khairy, and Chris Lacor. 2014. "Robust parameter design optimization using Kriging, RBF and RBFNN with gradient-based and evolutionary optimization techniques." *Applied Mathematics and Computation* 236:325-344.
7. Forrester, A. I. J., and A. J. Keane. 2009. "Recent advances in surrogate-based optimization." *Progress in Aerospace Sciences* 45 (1-3):50-79. doi: 10.1016/j.paerosci.2008.11.001.
8. Giunta, Anthony A, Steven F Wojtkiewicz, and Michael S Eldred. 2003. "Overview of modern design of experiments methods for computational simulations." Proceedings of the 41st AIAA aerospace sciences meeting and exhibit, AIAA-2003-0649.
9. Gupta, Deepti, and Shabina Ghafir. 2012. "An overview of methods maintaining diversity in genetic algorithms." *International journal of emerging technology and advanced engineering* 2 (5):56-60.

10. Holland, John H. 1973. "Genetic algorithms and the optimal allocation of trials." *SIAM Journal on Computing* 2 (2):88-105.
11. Hu, Jiexiang, Qi Zhou, Ping Jiang, Xinyu Shao, and Tingli Xie. 2017. "An adaptive sampling method for variable-fidelity surrogate models using improved hierarchical kriging." *Engineering Optimization*:1-19.
12. Jin, Yaochu, Markus Olhofer, and Bernhard Sendhoff. 2002. "A framework for evolutionary optimization with approximate fitness functions." *Evolutionary Computation, IEEE Transactions on* 6 (5):481-494.
13. Jones, Donald R. 2001. "A taxonomy of global optimization methods based on response surfaces." *Journal of global optimization* 21 (4):345-383.
14. Jones, Donald R, Matthias Schonlau, and William J Welch. 1998. "Efficient global optimization of expensive black-box functions." *Journal of Global optimization* 13 (4):455-492.
15. Juvinall, Robert C, and Kurt M Marshek. 2006. *Fundamentals of machine component design*. Vol. 83: John Wiley & Sons New York.
16. Kanakasabai, Pugazhendhi, and Anoop K Dhingra. 2014. "Reliability-based design optimization with progressive surrogate models." *Engineering Optimization* 46 (12):1609-1627.
17. Kitayama, Satoshi, Masao Arakawa, and Koetsu Yamazaki. 2011. "Sequential approximate optimization using radial basis function network for engineering optimization." *Optimization and Engineering* 12 (4):535-557.
18. Le, Minh Nghia, Yew Soon Ong, Stefan Menzel, Yaochu Jin, and Bernhard Sendhoff. 2013. "Evolution by adapting surrogates." *Evolutionary computation* 21 (2):313-340.
19. Liu, Heping, and Saeed Maghsoodloo. 2011. "Simulation optimization based on Taylor Kriging and evolutionary algorithm." *Applied Soft Computing* 11 (4):3451-3462.
20. Long, Teng, Di Wu, Xin Chen, Xiaosong Guo, and Li Liu. 2016. "A deterministic sequential maximin Latin hypercube design method using successive local enumeration for metamodel-based optimization." *Engineering Optimization* 48 (6):1019-1036.
21. Parr, J. M., A. J. Keane, A. I. J. Forrester, and C. M. E. Holden. 2012. "Infill sampling criteria for surrogate-based optimization with constraint handling." *Engineering Optimization* 44 (10):1147-1166. doi: 10.1080/0305215x.2011.637556.
22. Rezaee Jordehi, A, and Jasronita Jasni. 2013. "Parameter selection in particle swarm optimisation: a survey." *Journal of Experimental & Theoretical Artificial Intelligence* 25 (4):527-542.
23. Sacks, Jerome, William J Welch, Toby J Mitchell, and Henry P Wynn. 1989. "Design and analysis of computer experiments." *Statistical science*:409-423.
24. Shao, Tiefu, and Sundar Krishnamurty. 2008. "A clustering-based surrogate model updating approach to simulation-based engineering design." *Journal of Mechanical Design* 130 (4):041101.
25. Suprayitno, Yu, Jyh-Cheng. 2016. "Evolutionary algorithm using progressive Kriging model and dynamic reliable region for expensive optimization problems." Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on.
26. Tenne, Yoel. 2012. "A computational intelligence algorithm for expensive engineering optimization problems." *Engineering Applications of Artificial Intelligence* 25 (5):1009-1021.
27. Xia, Bin, Ziyan Ren, and Chang Seop Koh. 2017. "A novel reliability-based optimal design of electromagnetic devices based on adaptive dynamic Taylor Kriging." *IEEE Transactions on Magnetics*.
28. Younis, Adel, and Zuomin Dong. 2010. "Metamodelling and search using space exploration and unimodal region elimination for design optimization." *Engineering Optimization* 42 (6):517-533.
29. Yu, Jyh-Cheng, Zhi-Fu Liang, and Tsung-Ren Hung. 2014. "Evolutionary regional network modeling for efficient engineering optimization." Evolutionary Computation (CEC), 2014 IEEE Congress on.
30. Yu, Jyh- Cheng, and Jyh- Yeong Juang. 2010. "Design optimization of extrusion- blow- molded parts using prediction- reliability- guided search of evolving network modeling." *Journal of applied polymer science* 117 (1):222-234.
31. Zang, Hongnian, Shujun Zhang, and Kevin Hapeshi. 2010. "A review of nature-inspired algorithms." *Journal of Bionic Engineering* 7:S232-S237.
32. Zhan, Zhi-Hui, Jun Zhang, Yun Li, and Henry Shu-Hung Chung. 2009. "Adaptive particle swarm optimization." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 39 (6):1362-1381.