Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

GLOBAL OPTIMIZATION FOR NOISY EXPENSIVE BLACK-BOX MULTI-MODAL FUNCTIONS VIA RADIAL BASIS FUNCTION SURROGATE

Yichi Shen

Institute of Operations Research and Analytics
National University of Singapore
3 Research Link
117602, SINGAPORE

Christine A. Shoemaker

Department of Industrial System Engineering and Management
National University of Singapore
1 Engineering Drive 2
117576, SINGAPORE

ABSTRACT

This study proposes a new surrogate global optimization algorithm that solves problems with expensive black-box multi-modal objective functions subject to homogeneous evaluation noise. Specifically, we propose a new radial basis function (RBF) surrogate to approximate noisy functions and extend the *Stochastic Response Surface* method, which was developed for deterministic problems, to optimize noisy functions. Instead of conducting multiple replications at each point to mitigate the influence of noise, we only do a single observation at every sampled point and regularize the RBF surrogate by penalizing the bumpiness. The proposed algorithm sequentially identifies a new point for the expensive function evaluation from a set of randomly generated candidate points using both *exploitation* and *exploration*. Numerical studies show that the proposed noisy RBF surrogate can produce reliable approximations for noisy functions, and the proposed algorithm is effective and competitive in solving the global optimization problems with noisy evaluations.

1 INTRODUCTION

In this research, we focus on finding the global minimum of expensive black-box multi-modal functions, which arise in different areas such as simulation optimization and manufacturing optimization (Huang et al. 2006; Quan et al. 2013; Chang et al. 2013; Jin et al. 2020; Yi et al. 2020), where an objective evaluation or simulation run may take up to several hours.

Since each function evaluation is very expensive, to increase the search efficiency, some algorithms sequentially determine the next evaluation points with the assistance of a surrogate (i.e., a metamodel or a response surface). Examples of these algorithm are Efficient Global Optimization (EGO) (Jones et al. 1998), Sequential Kriging Optimization (SKO) (Huang et al. 2006), Correlated Knowledge Gradient (Frazier et al. 2009), Two Stage Sequential Optimization (TSSO) (Quan et al. 2013), eTSSO (Liu et al. 2014), which are all based on Gaussian Process (GP) surrogate, and Metric Stochastic Response Surface (MSRS) (Regis and

Shoemaker 2007), DYCORS (Regis and Shoemaker 2013), which are based on the Radial Basis Function (RBF) surrogate.

One key challenge in surrogate optimization is to strike the balance between *exploitation* and *exploration* when sequentially selecting the evaluation points. Common acquisition functions that are developed to balance the exploitation and exploration include Lower Confidence Bound (LCB) (Cox and John 1992), Expected Improvement (EI) (Jones et al. 1998), Augmented EI (AEI) (Huang et al. 2006), Modified EI (MEI) (Quan et al. 2013), Minimal Quantile (MQ) (Picheny et al. 2013), Correlated Knowledge Gradient (CKG) (Frazier et al. 2009), Weighted Score (Regis and Shoemaker 2007), Weighted Improvement (Ji and Kim 2013).

In addition to an expensive objective function without available derivatives, another challenge we may face is the stochastic nature of the optimization problem. For example, in practice the outcome of a simulation model can be stochastic (i.e., the evaluation of $f(x)$ is not only expensive but also noisy). The references references above using GP surrogate have generalized the algorithms to the case of objective functions subject to evaluation noise. For example, the original EI function is augmented to handle noisy data in the SKO algorithm (Huang et al. 2006). In contrast, less work using RBF as a surrogate has been done in this area and most RBF-based algorithms focus on deterministic functions. One potential reason is that, due to probability framework of GP, it might be relatively easier to include observation noise in the GP model than in the RBF surrogate. However, we believe the RBF surrogate deserves to be further developed because its advantages as indicated in the previous studies (Gutmann 2001; Regis and Shoemaker 2007; Regis and Shoemaker 2013).

In this study, we will propose a Noisy RBF (NRBF) fitting framework algorithm for evaluations with homogeneous noise aiming at solving the problems with expensive black-box multi-modal objective functions subject to homogeneous evaluation noise.

In the following, a precise description of the mathematical model is given in Section 2. Introduction of the proposed algorithm is described in Section 3. Section 4 presents the details of the proposed noisy RBF surrogate. Comparison results are provided in Section 5. Section 6 concludes this study and provides the directions for future study.

2 PROBLEM DEFINITION

We consider the following minimization problem:

$$x^* = \arg\min_{x\in\mathbb{X}} f(x),$$

where $\mathbb{X} = [lb, ub]^d$ is a d-dimensional hypercube in $\mathbb{R}^d$. However, the explicit analytical form of $f(x)$ or the derivative of $f(x)$ is not available, and $f(x)$ can only be evaluated through an expensive simulation model. The output of a function evaluation is stochastic:

$$y(x_i) = f(x_i) + \varepsilon_i, \forall x_i \in \mathbb{X},$$

where ε_i represents the evaluation noise at x_i and is a realization of a zero-mean random variable with variance $0 < \sigma^2 < \infty$.

We assume that $f(x)$ is a continuous multi-modal black-box function on $\mathbb{X}$ and the evaluation of $f(x)$ is very expensive. An efficient optimization algorithm should be able to return an x_N^* such that $f(x_N^*)$ is close to the true global optimum $f(x^*)$ using a limited number (N) of function evaluations. Note that $f(x)$ can have multiple local minima, and we are trying to find the global minimum. Therefore, an optimization algorithm is supposed to be able to explore the whole feasible domain and does not only focus on local information.

We assume that for any two sample points x_i and x_j, the evaluation noises ε_i and ε_j are independent and identically distributed (IID). In other words, we consider homogeneous noise structure for the function evaluations. In the presence of noise, a natural way to mitigate its influence is to conduct multiple observations at the same point and take the sample average. This type of approach might work, but its performance or efficiency is not satisfying. Since each evaluation is expensive and the total budget is very limited, doing replications always reduce the ability to explore the whole feasible domain to find the global optimum.

3 PROPOSED ALGORITHM

To solve the above optimization problem, we employ the surrogate method and sequentially determine the new evaluation points to strike the trade-off between *exploitation* and *exploration*, which are necessary for global optimization problem. In the following, we will propose the Noisy Radial Basis Function (NRBF) algorithm for the optimization of expensive black-box multi-modal functions subject to evaluation noise.

The NRBF algorithm is developed based on the Stochastic Response Surface (SRS) framework in Regis and Shoemaker (2007), which was developed to solve deterministic problems. The outline of the SRS framework is as follows:

1) Fit a response surface based on initial evaluations.
2) Randomly generate a fixed number of candidate points for the next expensive function evaluation.
3) Select a new point from the candidate points based on the acquisition function.
4) Evaluate the new point and update the surrogate.
5) If there is remaining computing budget, go to Step 2); Otherwise, return the current best solution.

Note that most surrogate-based optimization algorithms comply with the SRS framework. Different approaches may adopt different surrogates (e.g., RBF or GP) and have different acquisition functions (e.g., weighted score or EI) to address the trade-off between exploration and exploitation.

Details of the proposed NRBF algorithm are given in Algorithm 1. For the generation of candidate points and definition of weighted score in Step 2 and Step 3 within the iteration loop, we employ the DYCORS rule developed in Regis and Shoemaker (2013), which has been shown to have very good performance according to extensive case studies, especially for high-dimensional problems. Details of the perturbation rule for generating candidate points and weighted score for selecting a candidate point for the next expensive evaluation of $f(x)$ are given in Appendix A.

Algorithm 1: The NRBF Algorithm

Input: Objective function $f(x)$, total evaluation budget N, number of initial points N_0;
Output: x_N^*, $\hat{f}(x_N^*)$
Initialization: Set $n = N_0$, evaluate the N_0 initial points in $\mathscr{F}_n = \{x_1, x_2, ..., x_{N_0}\}$ and record the evaluations as $\mathscr{Y}_n = \{y(x_1), y(x_2), ..., y(x_{N_0})\}$;
while $n < N$ **do**
    Step 1. Build an NRBF surrogate $\hat{f}_n(x)$ based on $\mathscr{F}_n$ and $\mathscr{Y}_n$;
    Step 2. Generate t candidate points $\Omega_n = \{x_{n,1}, x_{n,2}, ..., x_{n,t}\}$ by perturbing the current best solution $x_n^* = \arg\min_{x \in \mathscr{F}_n} \hat{f}_n(x)$;
    Step 3. Determine the next evaluation point $x_{n+1} = \arg\min_{x \in \Omega_n} w_n(x)$, where $w_n(x)$ is the weighted score;
    Step 4. Update information: $\mathscr{F}_{n+1} = \mathscr{F}_n \cup \{x_{n+1}\}, \mathscr{Y}_{n+1} = \mathscr{Y}_n \cup \{y(x_{n+1})\}$. Set $n = n+1$;
end
Return the best solution: $x_N^* = \arg\min_{x \in \mathscr{F}_N} \hat{f}_N(x)$, and $\hat{f}(x_N^*) = \min_{x \in \mathscr{F}_N} \hat{f}_N(x)$.

Although the NRBF algorithm shares a similar framework with the algorithms in Regis and Shoemaker (2007) and Regis and Shoemaker (2013), they are different in the following ways. First, the RBF surrogate in these earlier papers is used to fit deterministic functions. In the present study, we generalize the RBF surrogate to stochastic settings, which is able to estimate the underlying black-box function $f(x)$ based on noisy data with only single observation at each point. Details of the surrogate for noisy evaluations will be given in the Section 4. Second, the identification of best solution is different. In the stochastic environment, it is not clear to say which point x is the "best" based on the noisy data of $f(x)$. Instead of using the point with the best sample observation which may have great bias, after each iteration we will identify the best point as the one with the best surrogate prediction value (i.e., $x_n^* = \arg\min_{x\in\mathscr{F}_n} \hat{f}_n(x)$), where $\mathscr{F}_n$ is defined in Algorithm 1 Step 1.

4 RBF SURROGATE

4.1 RBF Surrogate for Deterministic Evaluations

In this section, we introduce the original RBF surrogate for deterministic evaluations. We refer to the work of Gutmann (2001) for the details of RBF surrogate in general. For better readability, we briefly summarize the key steps of fitting an RBF surrogate in the following.

The RBF is a real-valued function whose value depends only on the distance between the input and some fixed points. It can be used to approximate given functions by a linear combination of the basis functions $\phi(\cdot)$, where the fixed points are the locations at which $f(x)$ have been evaluated.

Given a set of sampling points $x_1, x_2, \ldots, x_n \in \mathbb{R}^d$, and the function values $f(x_1), f(x_2), \ldots, f(x_n)$, the approximation can be obtained by:

$$\hat{f}(x) = \sum_{i=1}^{n} \lambda_i \phi(\| x - x_i \|) + p(x) \tag{1}$$

where $\| \cdot \|$ is the Euclidean norm, $\lambda_i \in \mathbb{R}^d$ and $p(x)$ is from Π_m, the space of polynomials of degree less than or equal to m. The radial basis functions $\phi(\cdot)$ can have different forms, e.g., linear ($\phi(r) = r$), cubic ($\phi(r) = r^3$), thin plate ($\phi(r) = r^2 \log r$), multi-quadratic ($\sqrt{r^2+\gamma^2}$), or Gaussian ($e^{-\gamma r^2}$). In this research, we adopt the cubic RBF ($\phi(r) = r^3$) with linear tail (i.e., linear $p(x) \in \Pi_1$) because of its excellent performance as indicated in previous studies (Regis and Shoemaker 2013). If $x = (x^1, x^2, ..., x^d))$, then $p(x) = c_0 + \sum_{j=1}^{j=d} c_j x^j$, and $c = (c_0, c_1, \ldots, c_d) \in \mathbb{R}^{(d+1)}$ is the coefficient vector of the polynomial $p(x)$.

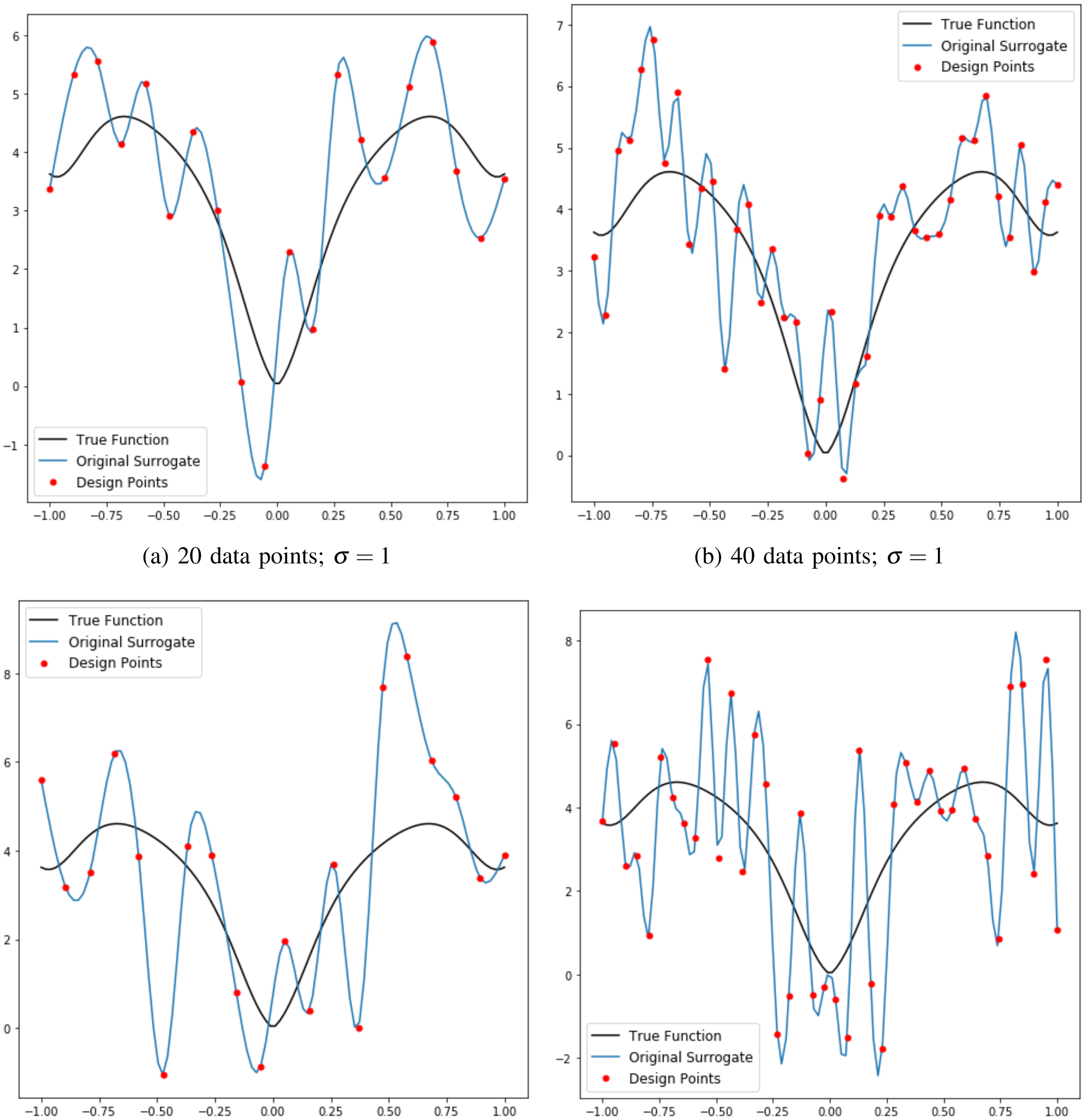
To estimate the parameters λ and c in Equation (1), we define the matrix $\Phi \in \mathbb{R}^{n\times n} : \Phi_{ij} = \phi(\| x_i - x_j \|)$, $i, j = 1, 2, \ldots, n$, and matrix $P \in \mathbb{R}^{n\times(d+1)}$. The vector in the ith row of P is $[1, x_i]$. Then the parameters in Equation (1) can be obtained by solving the following equation for λ and c:

$$\begin{bmatrix} \Phi & P \\ P^T & 0_{(d+1)\times(d+1)} \end{bmatrix} \begin{bmatrix} \lambda \\ c \end{bmatrix} = \begin{bmatrix} F \\ 0_{(d+1)\times 1} \end{bmatrix} \tag{2}$$

where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n$, $F = (f(x_1), f(x_2), \ldots, f(x_n))^T$, $0_{(d+1)\times(d+1)} \in \mathbb{R}^{(d+1)\times(d+1)}$ is a matrix of zeros, $c = (c_0, c_1, \ldots, c_d) \in \mathbb{R}^{(d+1)}$ is the coefficient vector of the polynomial $p(x)$, and $0_{(d+1)} \in \mathbb{R}^{(d+1)}$ is a vector of zeros. Powell (1992) showed that the coefficient matrix in (2) is invertible if and only if $\text{rank}(P) = \hat{m}$, where P is the matrix defined above, and $\hat{m} = \dim(\Pi_m)$. Note that the deterministic RBF surrogate exactly interpolates the observed values. Furthermore, as the number of samples goes to infinity, the RBF surrogate will converge to the underlying deterministic continuous function $f(x)$.

4.2 RBF Surrogate for Noisy Evaluations

In this section, we generalize the original RBF surrogate to fit the noisy data. One may think "Is it possible to ignore the observation noise and exactly interpolate the data points as in the previous section?" In Figure 1, we employ the original RBF surrogate to fit noisy data points from a 1-dimensional Ackley function with 20 or 40 data points and $\sigma = 1$ or $\sigma = 2$. The descriptions for Ackley function (Ackley 2012) are given in Appendix B. Unfortunately, as shown by Figure 1, the presence of noise will incredibly increase the bumpiness of fitting curve and make the approximation unreliable, especially when there are more points or the noise level is high. Apparently, this surrogate will produce misleading information and probably



(a) 20 data points; $\sigma = 1$

(b) 40 data points; $\sigma = 1$

(c) 20 data points; $\sigma = 2$

(d) 40 data points; $\sigma = 2$

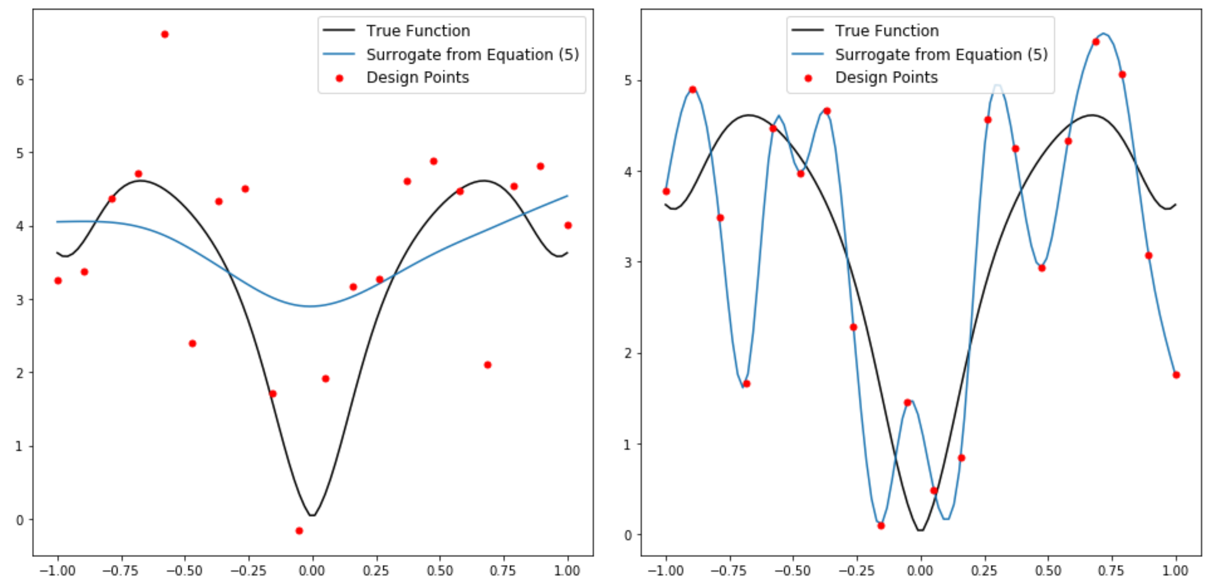
Figure 1: Examples of the interpolating curve using deterministic RBF for 20 or 40 samples of the noisy 1-dimensional Ackley test function with $\sigma = 1$ or $\sigma = 2$.

cannot facilitate the search for the global minimum.

One possible approach to deal with noise is to add a diagonal matrix to the coefficient in Equation (2), and we have:

$$\begin{bmatrix} \Phi + \eta I_{n \times n} & P \\ P^T & 0_{(d+1)\times(d+1)} + \eta I_{(d+1)\times(d+1)} \end{bmatrix} \begin{bmatrix} \lambda \\ c \end{bmatrix} = \begin{bmatrix} F \\ 0_{(d+1)\times 1} \end{bmatrix}, \tag{3}$$

where I denotes the identity matrix (Eriksson et al. 2019). However, the surrogate performance is very sensitive to the choice of η, and there is no general rule to determine its value. In practical applications, the optimal value of η is often adaptively determined according to Leave-one-out cross-validation (LOOCV). Figure 2 shows two instances of the RBF surrogate based on Equation (3) for the 1-dimensional Ackley function. Although the value of η is adaptively determined according to LOOCV, the surrogate performance is still not stable. It may fail to recognize the noise or reveal the underlying structure of the actual function $f(x)$.



(a) 20 data points; $\sigma = 1$

(b) 20 data points; $\sigma = 1$

Figure 2: Examples of the fitting curve based on Equation (3) for 20 samples of the noisy 1-dimensional Ackley test function with $\sigma = 1$.

A more successful approach to deal with noise in the regression references is to include a regularization term to avoid over-fitting. Instead of solving Equation (2) or (3) directly, we include a loss function of the parameters, and determine λ and c by solving the following:

$$\min_{\lambda,c} \left\| \begin{bmatrix} \Phi & P \\ P^T & 0_{(d+1)\times(d+1)} \end{bmatrix} \begin{bmatrix} \lambda \\ c \end{bmatrix} - \begin{bmatrix} F \\ 0_{(d+1)\times 1} \end{bmatrix} \right\|_2^2 + L(\lambda, c), \tag{4}$$

where the first term is the Squared Error over the evaluated points, and L is the loss function corresponding to the parameters λ and c. The loss function can have many different forms (e.g., the Tikhonov Regulator), and it significantly influences the performance of the surrogate. Intuitively, the loss function should well reflect the jumpiness or bumpiness of the surrogate caused by noise as shown in Figure 1.

Powell (1981) showed that the interpolant $\hat{f}(x)$ defined by the Equation (2) minimizes $\int_R [g''(x)]^2 dx$ among all functions $g: R \to R$ that satisfy the interpolation conditions $g(x_i) = f(x_i)$, $i = 1, 2, ..., n$ for which $\int_R [g''(x)]^2 dx$ exists and is finite. Therefore, $\int_R [\hat{f}''(x)]^2 dx$ is a reasonable measure for the bumpiness. Furthermore, when the cubic RBF and linear tail are adopted, $\int_R [\hat{f}''(x)]^2 dx = 12\lambda^T \Phi \lambda$ (Gutmann 2001).

Intuitively, a sensible loss function $L(\lambda, c)$ should be positively correlated with the bumpiness to avoid over-fitting. In this research, we propose the following loss function:

$$L(\lambda, c) := \frac{1}{n} \lambda^T \Phi \lambda = \frac{1}{n} \begin{bmatrix} \lambda \\ c \end{bmatrix}^T \begin{bmatrix} \Phi & 0_{n\times(d+1)} \\ 0_{(d+1)\times n} & 0_{(d+1)\times(d+1)} \end{bmatrix} \begin{bmatrix} \lambda \\ c \end{bmatrix}, \tag{5}$$

where n is number of sample points. Note that we propose the above loss function (5) based on numerical investigations. It has good performance when applied to different test problems and is robust to different noise levels.

Combining Equation (5), the optimization problem (4) can be rewritten in a generalized Tikhonov regularization form (Tikhonov et al. 2013):

$$\min_b ||Ab - z||_2^2 + ||b||_Q^2, \tag{6}$$

where $A = \begin{bmatrix} \Phi & P \\ P^T & 0_{(d+1)\times(d+1)} \end{bmatrix}$, $b = \begin{bmatrix} \lambda \\ c \end{bmatrix}$, $z = \begin{bmatrix} F \\ 0_{(d+1)\times 1} \end{bmatrix}$, $Q = \frac{1}{n}\begin{bmatrix} \Phi & 0_{n\times(d+1)} \\ 0_{(d+1)\times n} & 0_{(d+1)\times(d+1)} \end{bmatrix}$, and $||b||_Q^2 = b^T Q b$ is a weighted norm.

Note that, the solution to the optimization problem (6) can be given by

$$b = \begin{bmatrix} \lambda \\ c \end{bmatrix} = (A^T A + Q)^{-1}(A^T z). \tag{7}$$

In summary, in the presence of noise, the Noisy Radial Basis Function (NRBF) surrogate is given by $\hat{f}(x) = \sum_{i=1}^n \lambda_i \phi(\| x - x_i \|) + p(x)$, where the parameters λ and c can be computed from the system (7).

To examine the performance of the proposed NRBF surrogate, we again conduct a curve fitting for the data points from the 1-dimensional Ackley function. As shown by Figure 3, the proposed new NRBF surrogate based on the RBF parameters computed in (7) has significantly better performance in discovering the underlying structure of the black-box function $f(x)$ compared with Figure 1 and Figure 2. It is also robust even with higher noise and tends to be more accurate when there are more data points. These features are exactly what we expect and they provide a foundation for generalizing the RBF-based global optimization algorithms to stochastic settings.

5 COMPARISON RESULTS

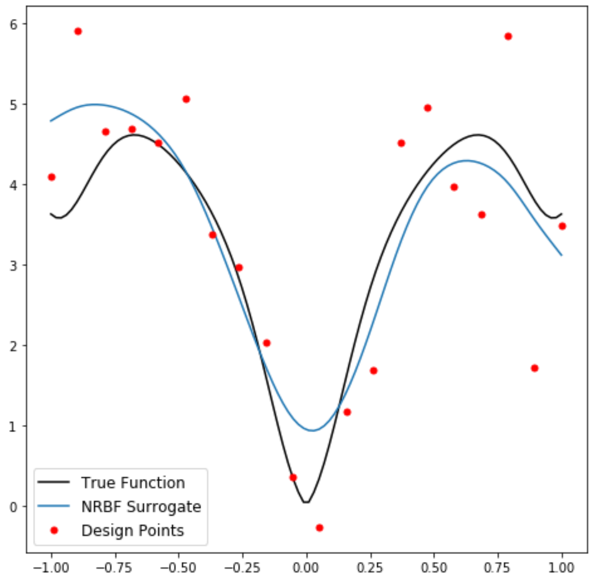
In this section, we examine the performance of NRBF algorithm for functions with homogeneous noise. We compare the performance with SKO (Huang et al. 2006) and KGCP (Scott et al. 2011) over three test functions: 2-dimensional Six-Hump Camel Back function (Branin 1972), 3-dimensional Hartman function (Hartman 1973), and 5-dimensional Ackley function (Ackley 2012). Details of the functions are given in the Appendix B. To observe the response of the algorithms to different noise settings, we consider Gaussian noise with three variance levels: Low: $\sigma^2 = 0.1$, Medium: $\sigma^2 = 1$, and High: $\sigma^2 = 10$. Note that the variance of noise is unknown to the optimization algorithms. For each experiment, $2(d+1)$ initial points are generated by the Latin Hypercube Sampling (LHS) experiment design. 500 trials are conducted for each case to compare the statistical performance. The NRBF algorithm is coded based on the open source pySOT package (Eriksson et al. 2019).

For comparison purposes, we employ the Opportunity Cost (OC) as the performance measure:

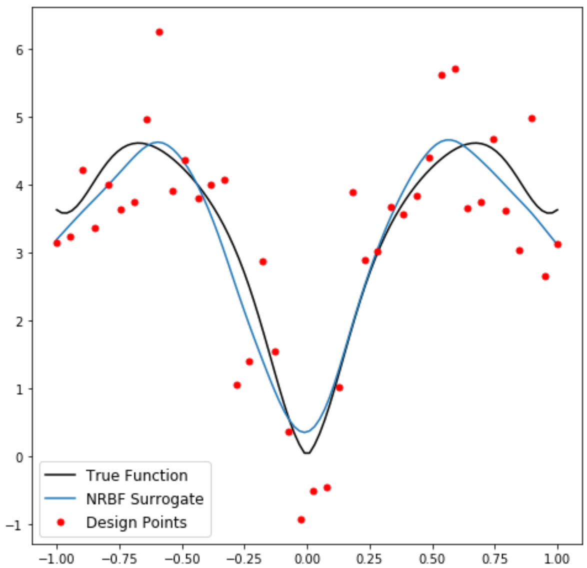
$$OC = f(x_N^*) - f(x^*),$$

where x_N^* is the solution returned by each algorithm when all evaluation budget are consumed, and x^* is the actual global optimum.

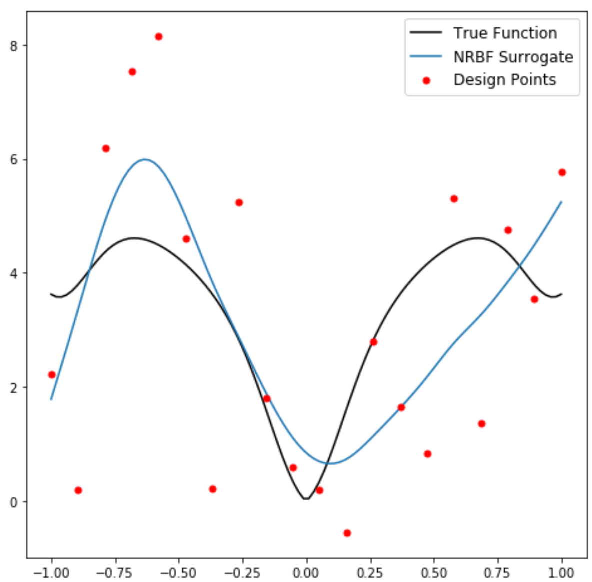
Tables 1, 2 and 3 show the mean of OC ($E[OC]$) and its standard deviation ($STD[OC]$) for each algorithm from 500 trials after 50 iterations (i.e., the total evaluation budget is $50 + 2(d+1)$). The significantly better values of the mean of OC ($E[OC]$) are presented in bold font. Note that the results of SKO and KGCP in the tables are directly retrieved from Table 6.1 in Scott et al. (2011). As shown by the results, the NRBF algorithm has the best performance for 6 out of the 9 cases. For the Six Hump Function with high noise, SKO seems to produce the best performance. For the remaining 2 cases, the performance of NRBF has no significant difference from KGCP, but is better than SKO. This indicates that, the proposed



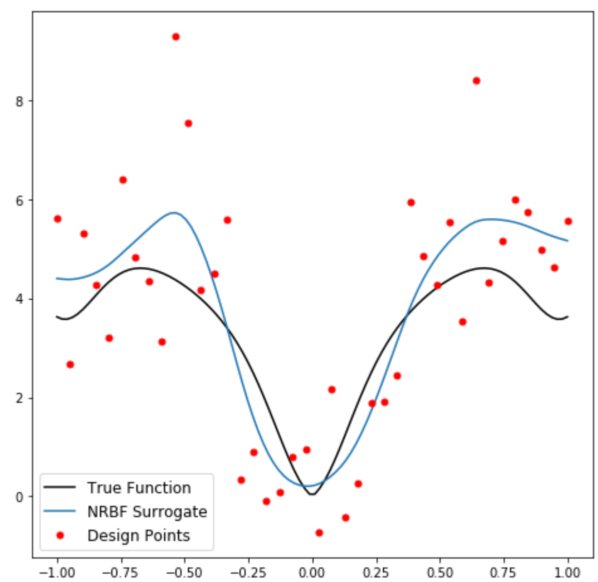
(a) 20 data points; $\sigma = 1$



(b) 40 data points; $\sigma = 1$



(c) 20 data points; $\sigma = 2$



(d) 40 data points; $\sigma = 2$

Figure 3: Examples of the fitting curve using NRBF for 20 or 40 samples of the noisy 1-dimensional Ackley test function with $\sigma = 1$ or $\sigma = 2$.

NRBF algorithm is a very competitive approach for the optimization of expensive black-box multi-modal functions subject to evaluation noise.

Compared with the traditional GP-based algorithms, the proposed NRBF does not need to solve an optimization sub-problem (i.e., optimize the AEI or CKG) at each iteration, which is actually hard to solve and has no exact closed-form solution. Hence, one potential reason the NRBF algorithm tends to outperform others is that the weighted score as given in Appendix A.2 can better balance the trade-off between exploration and exploitation.

Noise Levels	NRBF		SKO		KGCP	
	$E[OC]$	$STD(E[OC])$	$E[OC]$	$STD(E[OC])$	$E[OC]$	$STD(E[OC])$
Low: $\sigma^2 = 0.1$	**0.0548**	0.0025	0.1112	0.0059	0.0714	0.0087
Medium: $\sigma^2 = 1$	**0.2968**	0.0142	0.3597	0.0156	0.3208	0.0192
High: $\sigma^2 = 10$	1.0153	0.0383	**0.8488**	0.0370	1.0264	0.0391

Table 1: Comparison results of Opportunity Cost (OC) for the 2-dimensional Six-Hump Camel Back function under three different noise levels with 500 trials.

Noise Levels	NRBF		SKO		KGCP	
	$E[OC]$	$STD(E[OC])$	$E[OC]$	$STD(E[OC])$	$E[OC]$	$STD(E[OC])$
Low: $\sigma^2 = 0.1$	0.0669	0.0035	0.1079	0.0075	0.0690	0.0063
Medium: $\sigma^2 = 1$	**0.3295**	0.0135	0.5012	0.0216	0.5336	0.0296
High: $\sigma^2 = 10$	**1.5742**	0.0465	1.8370	0.0510	1.8200	0.0541

Table 2: Comparison results of Opportunity Cost (OC) for the 3-dimensional Hartman function under three different noise levels with 500 trials.

Noise Levels	NRBF		SKO		KGCP	
	$E[OC]$	$STD(E[OC])$	$E[OC]$	$STD(E[OC])$	$E[OC]$	$STD(E[OC])$
Low: $\sigma^2 = 0.1$	**2.8873**	0.0630	7.8130	0.1802	5.7304	0.1874
Medium: $\sigma^2 = 1$	**7.5714**	0.2300	12.6346	0.2088	10.8315	0.2413
High: $\sigma^2 = 10$	17.6634	0.1488	18.1126	0.1156	17.3670	0.1477

Table 3: Comparison results of Opportunity Cost (OC) for the 5-dimensional Ackley function under three different noise levels with 500 trials.

6 CONCLUSION

This study has generalized the RBF surrogate method to solve expensive black-box multi-modal functions with homogeneous noise. Specifically, we develop a new noisy RBF fitting framework to deal with noisy data by penalizing the bumpiness. Through the regularization, the NRBF is able to provide a reliable inference for the noisy black-box function with only single observation at each point. Based on the proposed new surrogate, the NRBF algorithm is developed and has been shown to have better performance compared with two benchmark algorithms: SKO and KGCP.

The proposed algorithm has the potential to be applied to the optimization of practical large-scale systems, and this is one of the future research directions. Furthermore, we will also try to analyze the convergence property of the algorithm in future.

This research focuses on evaluations with homogeneous noise. In some practical cases, the levels of noise at different points might be different and it makes the surrogate fitting more difficult. Generalizing the approach to the heterogeneous noise setting is a direction for future research.

ACKNOWLEDGMENTS

This research was supported by the Institute for Operations Research Analytics (IORA) at National University of Singapore (NUS); by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (E2S2- CREATE) programme; and by an NUS grant to Prof. Shoemaker.

A APPENDIX

A.1 Generating of Candidate Points

At Step 2 of Algorithm 1, t candidate points are generated by randomly perturbing on some or all the coordinates of the best solution found so far (x_n^*). For each coordinate of x_n^*, the perturbing probability $\varphi(n)$ is given by:

$$\varphi(n) = \varphi_0 \cdot \frac{1 - ln(n - N_0 + 1)}{ln(N - N_0)}, \quad N_0 \leq n \leq N - 1,$$

where $\varphi_0 = \min(20/d, 1)$, and the perturbation noise is a random number following Normal distribution $\mathcal{N}(0, \delta^2)$. Note that $\varphi(n)$ will be smaller as n increases, which means DYCORS perturbs more coordinates at the beginning but perturbs less dimensions at the end of the optimization process.

A.2 Weighted Score

At Step 3 of Algorithm 1, the next evaluation point is given by $x_{n+1} = \arg\min_{x \in \Omega_n} w_n(x)$, where $w_n(x)$ is the weighted score that balances exploration and exploitation. Following Regis and Shoemaker (2013), for $x \in \Omega_n$, we have

$$w_n(x) = \omega \cdot V_n^R(x) + (1 - \omega) \cdot V_n^D(x),$$

$$V_n^R(x) = (\hat{f}_n(x) - \min_{x \in \Omega_n} \hat{f}_n(x)) / (\max_{x \in \Omega_n} \hat{f}_n(x) - \min_{x \in \Omega_n} \hat{f}_n(x)),$$

$$V_n^D(x) = (\max_{x \in \Omega_n} \Delta_n(x) - \Delta_n(x)) / (\max_{x \in \Omega_n} \Delta_n(x) - \min_{x \in \Omega_n} \Delta_n(x)),$$

and

$$\Delta_n(x) = \min_{x' \in \mathscr{F}_n} ||x - x'||_2.$$

Note that ω is a given weight, $\Delta_n(x)$ is the distance from x to the evaluated points, $V_n^R(x)$ is the score for exploitation, and $V_n^D(x)$ is the score for exploration.

B APPENDIX

Six-Hump Camel Back Function
Dimension: $d = 2$;
Expression: $f(x) = 4x_1^2 - 2.1x_1^4 + 1/3x_1^6 + x_1x_2 + 4x_2^2 + 4x_2^4$;
Domain: $x_1 \in [-1.6, 2.4], x_2 \in [-0.8, 1.2]$;
Optimum: $x^* = (0.089, -0.713)$, or $x^* = (-0.089, 0.713)$; $f(x^*) = -1.0316$.
Hartman-3 Function
Dimension: $d = 3$;
Expression: $f(x) = -\sum_{i=1}^4 \alpha_i \exp - \sum_{j=1}^3 A_{ij}(x_j - P_{ij})^2)$, where $\alpha = [1.0, 1.2, 3.0, 3.2]^T$,

$$A = \begin{bmatrix} 3.0 & 10 & 30 \\ 0.1 & 10 & 35 \\ 3.0 & 10 & 30 \\ 0.1 & 10 & 35 \end{bmatrix}, \text{ and } P = 10^{-4} \begin{bmatrix} 3689 & 1170 & 2673 \\ 4699 & 4387 & 7470 \\ 1091 & 8732 & 5547 \\ 381 & 5743 & 8828 \end{bmatrix};$$

Domain:$x_i \in [0, 1]$, $i = 1, 2, 3$;
Optimum: $x^* = (0.1146, 0.5556, 0.8525)$; $f(x^*) = -3.8628$.
Ackely Function
Dimension: d;
Expression: $f(x) = -20\exp(-0.2\sqrt{1/d\sum_{i=1}^d x_i^2}) - \exp(1/d\sum_{i=1}^d \cos(2\pi x_i)) + 20 + \exp(1)$;
Domain: $x_i \in [-15, 30]$, $i = 1, 2, ..., d$;
Optimum: $x^* = (0, 0, ..., 0)$, $f(x^*) = 0$.

REFERENCES

Ackley, D. 2012. *A Connectionist Machine for Genetic Hillclimbing*. Boston: Springer Science & Business Media.

Branin, F. H. 1972. "Widely Convergent Method for Finding Multiple Solutions of Simultaneous Nonlinear Equations". *IBM Journal of Research and Development* 16(5):504–522.

Chang, K.-H., L. J. Hong, and H. Wan. 2013. "Stochastic Trust-Region Response-Surface Method (STRONG) - A New Response-Surface Framework for Simulation Optimization". *INFORMS Journal on Computing* 25(2):230–243.

Cox, D. D., and S. John. 1992. "A statistical method for global optimization". In *Proceedings of the 1992 IEEE International Conference on Systems, Man, and Cybernetics*, 1241–1246. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Eriksson, D., D. Bindel, and C. A. Shoemaker. 2019. *pySOT and POAP: An Event-Driven Asynchronous Framework for Surrogate Optimization*. arXiv Preprint. https://arxiv.org/abs/1908.00420.

Frazier, P., W. Powell, and S. Dayanik. 2009. "The Knowledge-Gradient Policy for Correlated Normal Beliefs". *INFORMS Journal on Computing* 21(4):599–613.

Gutmann, H.-M. 2001. "A Radial Basis Function Method for Global Optimization". *Journal of Global Optimization* 19(3):201–227.

Hartman, J. K. 1973. "Some Experiments in Global Optimization". *Naval Research Logistics Quarterly* 20(3):569–576.

Huang, D., T. T. Allen, W. I. Notz, and N. Zheng. 2006. "Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models". *Journal of Global Optimization* 34(3):441–466.

Ji, Y., and S. Kim. 2013. "An Adaptive Radial Basis Function Method Using Weighted Improvement". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S. H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 957–968. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Jin, X., Y. Shen, L. H. Lee, E. P. Chew, and C. A. Shoemaker. 2020. "A Hybrid of Shrinking Ball Method and Optimal Large Deviation Rate Estimation in Continuous Contextual Simulation Optimization with Single Observation". In *Proceedings of the 2020 Winter Simulation Conference*, edited by K. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, Accepted. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization of Expensive Black-Box Functions". *Journal of Global Optimization* 13(4):455–492.

Liu, C., G. Pedrielli, and S. H. Ng. 2014. *eTSSO: Adaptive Search Method for Stochastic Global Optimization under Finite Budget*. Semantic Scholar. https://pdfs.semanticscholar.org/ff70/adfeeb9cea3b877e92853dfcb80704a4acc8.pdf?_ga=2.15305590.1525711594.1597891339-1017363533.1597891339.

Picheny, V., T. Wagner, and D. Ginsbourger. 2013. "A Benchmark of Kriging-Based Infill Criteria for Noisy Optimization". *Structural and Multidisciplinary Optimization* 48(3):607–626.

Powell, M. 1981. *Approximation Theory and Methods*. Cambridge: Cambridge University Press.

Powell, M. J. 1992. "The Theory of Radial Basis Function Approximation in 1990". *Advances in Numerical Analysis*:105–210.

Quan, N., J. Yin, S. H. Ng, and L. H. Lee. 2013. "Simulation Optimization via Kriging: A Sequential Search Using Expected Improvement with Computing Budget Constraints". *IIE Transactions* 45(7):763–780.

Regis, R. G., and C. A. Shoemaker. 2007. "A Stochastic Radial Basis Function Method for the Global Optimization of Expensive Functions". *INFORMS Journal on Computing* 19(4):497–509.

Regis, R. G., and C. A. Shoemaker. 2013. "Combining Radial Basis Function Surrogates and Dynamic Coordinate Search in High-Dimensional Expensive Black-Box Optimization". *Engineering Optimization* 45(5):529–555.

Scott, W., P. Frazier, and W. Powell. 2011. "The Correlated Knowledge Gradient for Simulation Optimization of Continuous Parameters Using Gaussian Process Regression". *SIAM Journal on Optimization* 21(3):996–1026.

Tikhonov, A. N., A. Goncharsky, V. Stepanov, and A. G. Yagola. 2013. *Numerical Methods for the Solution of Ill-Posed Problems*, Volume 328. Dordrecht: Springer Science & Business Media.

Yi, J., Y. Shen, and C. A. Shoemaker. 2020. "A Multi-Fidelity RBF Surrogate-Based Optimization Framework for Computationally Expensive Multi-Modal Problems with Application to Capacity Planning of Manufacturing Systems". *Structural and Multidisciplinary Optimization*:1–21.

AUTHOR BIOGRAPHIES

YICHI SHEN is currently a postdoctoral research fellow in the Institute of Operations Research and Analytics, National University of Singapore. He obtained his Ph.D. in Industrial and System Management from Nanyang Technological University, Singapore and holds the M.Sc. and B.Sc. in Mathematics from Nanjing University, China. His current research focuses on developing efficient and cost-effective algorithms for the computational optimization of stochastic expensive black-box functions and he is passionate about applying the research to solving real life problems. His research interests also cover Operations Research, Data Analysis, Stochastic Models and Smart Manufacturing. His e-mail address is orasy@nus.edu.sg.

CHRISTINE A. SHOEMAKER is Distinguished Professor in the Industrial Systems Engineering and Management Department and Civil and Environmental Engineering Department at the National University of Singapore. She is Joseph P. Ripley Professor Emerita at Cornell University in USA. Prof Shoemaker is Member of the USA National Academy of Engineering, and she is a (professional society elected) Fellow in SIAM (applied mathematics), a Fellow in INFORMS (operations research), a Fellow in AGU (geophysics) and a Fellow and Distinguished Member of ASCE (Civil and Environmental Engineering). Prof. Shoemaker received her PhD in mathematics (on dynamic programming applied to environmental problems) supervised by Richard Bellman. Her current research has two main components which are: 1) development of computationally efficient optimization algorithms that can be used to solve problems where the objective function is multi-modal (i.e. has multiple local minima) and is expensive to compute (so relatively few objective functions can be made) and no derivatives are available. In this research a radial basis function surrogate is used and both serial and parallel algorithms have been developed for continuous and integer variables and for single and multi-objective problems. The current submission in this area extends her prior work to noisy single objective continuous functions. 2) application of optimization methods to difficult real problems described by nonlinear simulation models (deterministic or stochastic, PDE or discrete time dynamic models)) that describe environmental problems, especially in the area of water quality or water resources (including lakes, watersheds, and groundwater). The methods developed in I) above are applied to these problems and more recently to hyperparameter optimization in deep learning. Her e-mail address is shoemaker@nus.edu.sg.