Journal of Global Optimization (2006) 34: 441–466
DOI 10.1007/s10898-005-2454-3

Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models

D. HUANG[1], T.T. ALLEN[2], W.I. NOTZ[3] and N. ZENG[2]
[1]*Scientific Forming Technologies Corporation, 5038 Reed Road, Columbus, OH 43220, USA (e-mail:dhuang@deform.com)*
[2]*Department of Industrial, Welding, and Systems Engineering, The Ohio State University, 1971 Neil Avenue, Columbus, OH 43210, USA (e-mail:allen.515@osu.edu; zheng.481@osu.edu)*
[3]*Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, OH 43210, USA (e-mail:win@stat.ohio-state.edu)*

(Received 21 July 2004; accepted 21 August 2005)

Abstract. This paper proposes a new method that extends the efficient global optimization to address stochastic black-box systems. The method is based on a kriging meta-model that provides a global prediction of the objective values and a measure of prediction uncertainty at every point. The criterion for the infill sample selection is an augmented expected improvement function with desirable properties for stochastic responses. The method is empirically compared with the revised simplex search, the simultaneous perturbation stochastic approximation, and the DIRECT methods using six test problems from the literature. An application case study on an inventory system is also documented. The results suggest that the proposed method has excellent consistency and efficiency in finding global optimal solutions, and is particularly useful for expensive systems.

Key words: Efficient global optimization, expected improvement, kriging, stochastic black-box systems.

1. Introduction

Optimization methods for stochastic black-box systems have applications in many areas such as engineering and discrete-event system simulation. For example, in a metal-forming shop, the manufacturer may want to adjust certain process parameters, such as forming temperature and die geometry, to maximize the performance of the manufacturing system. The systems here are 'stochastic' because the measured outputs contain random errors. Thus, the objective function is the expected output. The manufacturing system is treated as a 'black box' because no closed-form formulation or gradient information of the objective function is available. In this paper, by the terms 'stochastic' or 'noisy,' we mean that these experiments, if repeated, will give different results. They are not to be confused with the experiments whose results deviate from the 'true' responses, but repeated

runs will give the same output, as the cases with many numerical models of physical phenomena.

Similar optimization problems may also be encountered in optimization of stochastic discrete-event systems, which is often referred to as the area of simulation optimization (Andradottir, 1998). Here a 'simulation' refers to a random observation on the system performance such as the time a customer waits. Although most of the efforts in simulation optimization have been devoted to exploiting the underlining system structure for gradient estimation, methods that treat the objective functions as black boxes continue to be useful due to their generality and ease of use (Fu, 1994).

In practice on the factory floor, a common method for finding the optimum is via 'one-shot' response surface methods (RSM). In one-shot RSM, the system responses are fit with a regression model using a classic experimental design, and the optimal solution is determined from this model. This method can be considered inefficient in that it attempts to accurately predict the response curve over the entire domain of feasibility, while we are more interested prediction in the neighborhood of the optimum. In addition, the regression models are often relatively simple and may not adequately fit certain complex systems over the entire feasible region.

Sequential RSM procedures have been applied for simulation optimization (Safizadeh, 1984; Neddermeijer et al., 2000; Angün et al., 2002). Instead of fitting the entire feasible region, small sub-regions (over which a low order polynomial response surface is thought to adequately approximate the function) are explored in succession, which leads to potential improvement, until the slope is 'approximately' zero. Usually, in the initial phase, the sub-regions are fitted with first-order regression models via fractional factorial designs. A move is taken towards the steepest descent direction with step size determined by a line search. In the final phase, a quadratic model with a central composite design is fit and the optimum determined. There is another framework that uses a similar strategy as the sequential RSM method. It is called sequential approximation optimization (SAO), which is adopted in the area of multi-discipline design optimization by Haftka et al. (1993) and Rodriguez et al. (2001). The basic concept in the SAO framework is to minimize a local RSM approximation of the objective function and constraints subjected to local move limits. The local approximations are often quadratic response surfaces that match zero and first order information, if available.

Additional optimization methods for stochastic black-box systems include the Nelder–Mead simplex search procedure. The Nelder–Mead method is a widely used non-derivative heuristic optimization method designed originally for deterministic problems. Recently, it has been revised to solve stochastic problems by Barton and Ivey (1996) and Humphrey and Wilson (2000). Neddermeijer et al. (1999) empirically compared Barton and Ivey's

approach with the sequential RSM procedure and concluded that, in general, the former performed more efficiently than the latter for the simulation model and test functions used in their study.

Another group of methods originated from the so-called stochastic approximation (SA) methods (Kushner and Clark, 1978). The SA methods are essentially gradient-based, with the gradients estimated either numerically or analytically. Among numerous gradient estimation approaches, here we only review the ones that treat the objectives as black boxes. The Kiefer–Wolfowitz (1952) algorithm uses the finite difference methods to estimate gradients. Spall (1992) proposed the simultaneous perturbation stochastic approximation (SPSA) method for gradient estimation, which requires only two evaluations per estimation, regardless of the dimension. In fact, the sub-region regression in the above-mentioned sequential RSM procedure is also a form of gradient estimation. In general, when random errors exist, the finite-differences-like gradient estimation can be challenging: for the bias to be small, it is necessary to have a small distance between points; but as the distance decreases, the variance of the estimators tends to increase.

In the area of global optimization, methods for stochastic black-box systems have also been studied. Kushner (1964) and Žilinskas (1980) developed alternatives to response surface methods using the Bayes' theorem. The method, so-called Bayesian global optimization, is based on a stochastic model of the objective function and is able to deal with noisy responses. However, Kushner (1964) used a heuristic optimization method to select the next point for sampling based on 'simplified' models; and Žilinskas (1980) only considered one-dimensional problems. Motivated by a modification to Lipschitzian optimization, Perttunen et al. (1993) and Gablonsky et al. (2001) developed the so-called DIRECT algorithm. The DIRECT method is a sampling algorithm, which requires no gradient information, and decides where to search next based on previously collected data. The name stands for '*Di*viding *rect*angles,' which describes the way the algorithm moves towards the optimum. The method performs a balance between local and global search within bound constraints.

Heuristic methods, such as the genetic algorithm, simulated annealing, and tabu search, can also be applied to solve stochastic black-box problems. A major advantage of these methods is their universal applicability, as they usually assume very little about the objective functions, not even the continuity or smoothness. Heuristic methods are particularly useful when the cost per evaluation is inexpensive. For non-linear problems that are continuous and smooth, it is generally believed that the heuristic methods require more evaluations than other classes of methods mentioned above.

In recent years, promising new methods with their roots in the Bayesian global optimization method were developed with a primary focus on

deterministic problems. Jones et al. (1998) proposed the efficient global optimization (EGO) algorithm, which utilizes a full kriging model to select sample points based on a rigorous search method. The new points, or 'infill samples,' are selected based on a criterion called 'expected improvement' that balances the need to exploit the approximating surface with the need to improve the approximation. The EGO algorithm and its improved versions have shown good generality and performance for various practical problems in engineering design (Sasena et al., 2001, 2002). An additional algorithm was developed by Sóbester et al. (2002), where gradient enhanced radial basis functions (GERBF) are used as the meta-models to determine the infill sampling points. This method showed advantages as compared to more traditional gradient-based and guided-random search methods.

Since EGO has its roots in the Bayesian global optimization, a method designed for stochastic responses, Jones et al. (1998) and Sasena (2002) commented that the EGO method would likely be adaptable to address stochastic systems. However, studies on this adaptation are lacking in the research literature. In this paper, we propose a formulation to extend the EGO scheme to stochastic systems, and compare it with alternative methods. We refer to our new method as sequential kriging optimization (SKO) to more clearly differentiate it from alternative approaches.

In Section 2, we describe the proposed method and the associated assumptions. The key ingredient in the method is the so-called 'augmented expected improvement' function that accommodates stochastic black-box function evaluations. Section 3 provides a simple numerical example to foster an intuitive understanding of the method. In Section 4, the performance of proposed method is compared empirically with alternative approaches from Spall (1992), Humphrey et al. (2000), and Gablonsky et al. (2001) using six test problems. Section 5 provides an application case study of SKO on an inventory system. Section 6 addresses the limitations and other relevant issues. Finally, Section 7 summarizes the conclusions and describes opportunities for future research.

2. Assumption and Formulation

2.1. THE OPTIMIZATION PROBLEM

The goal is to minimize the objective (or loss) function, $f(\mathbf{x})$, within the feasible region, χ, i.e.:

$$\min_{\mathbf{x}\in\chi} f(\mathbf{x}) \tag{1}$$

where $f(\mathbf{x})$ represents the expected performance of a system and $\mathbf{x}$ is a d-dimensional vector of parameters to be adjusted. We consider the system as

a black box that provides no information other than the measurements of system performance. We assume that the feasible region $\chi \subset \mathrm{R}^d$ is continuous, connected, and compact. The measurement Y of the objective function contains random error or noise:

$$Y = f(\mathbf{x}) + \varepsilon \tag{2}$$

In this paper, we assume that random errors from successive measurements are independent identically distributed (IID) normal deviates.

2.2. OVERVIEW OF THE PROCEDURE

The outline for the proposed SKO method is identical to that of EGO by Jones et al. (1998). For the matter of completeness, we review that framework:

Step 1. Build an initial kriging meta-model of the objective function.
Step 2. Use cross validation to ensure that the kriging prediction and measure of uncertainty are satisfactory.
Step 3. Find the location that maximizes the Expected Improvement (EI) function. If the maximal EI is sufficiently small, stop.
Step 4. Add an evaluation at the location where the EI is maximized. Update the kriging meta-model using the new data point. Go to Step 3.

The proposed SKO methods differ from the original EGO methods in the implementation of the kriging meta-model in Step 1 and in the formula for the EI function in Step 3. These differences are motivated largely by the need to accommodate random errors, as described in Sections (2.3)–(2.8). For Step 2, as in the original EGO, we generate a prediction with one data point excluded from the data set. Then we check whether that data point falls within a certain confidence interval for the prediction. If the test fails, appropriate transformations such as log or inverse may be applied to the response values. For example, when the response values vary greatly in magnitudes, a log transformation can be helpful for generating a better kriging prediction.

2.3. KRIGING META-MODELING WITH RANDOM ERRORS

In kriging meta-modeling, the response is assumed to be the sum of a linear model, a term representing the systematic departure (bias) from the linear model, and noise (Cressie, 1993):

$$Y(\mathbf{x}) = \sum_{i=1}^{k} \beta_i h_i(\mathbf{x}) + Z(\mathbf{x}) + \varepsilon \tag{3}$$

where h and β are basis functions and their coefficients, respectively. Z is the systematic departure and ε is the random error. The basis functions are usually polynomials, and often only one term, i.e., the constant term, is sufficient for generating good kriging meta-models (Sacks et al., 1989a,b).

The kriging meta-model derives from an estimation process in which the systematic departure from the linear model, Z, is assumed to be a realization of a stationary Gaussian stochastic process. We use the following formula to describe the covariance of systematic errors between outputs at two points $\mathbf{t}=(t_1,\ldots,t_d)$ and $\mathbf{u}=(u_1,\ldots,u_d)$:

$$\operatorname{cov}[Z(\mathbf{t}),Z(\mathbf{u})]=\sigma_Z^2 R_Z(\mathbf{t},\mathbf{u})=\sigma_Z^2\exp\left[-\sum_{j=1}^{d}\theta_j(t_j-u_j)^2\right] \tag{4}$$

where σ_Z^2 is the variance of the stochastic process, R_Z is the correlation function (sometimes referred to as the Gaussian correlation function), and θ_j is a scale parameter associated with dimension j. A larger θ_j implies a higher 'activity,' i.e., lower correlation within the dimension j. Under this covariance structure, the correlation between points decreases as the distance between them increases, which is a desirable property. Other forms of covariance structures have also been used in modeling computer experiments, a detailed summary of which can be found in Santner et al. (2003).

As mentioned previously, in this study the random errors are assumed to be IID. We denote by σ_ε^2 the variance of the random error, and by $Y_1, Y_2, \ldots, Y_n$ the data drawn from an n-point design $\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n\}$. To describe the kriging model predictor, we introduce the following notation:

$$\begin{aligned}
&\mathbf{h}_{\mathbf{x}}'=[h_1(\mathbf{x}),\ldots,h_n(\mathbf{x})]\\
&\mathbf{V}=[\operatorname{cov}(Y_i,Y_j)]_{1\leqslant i,j\leqslant n}=[\operatorname{cov}(Z(\mathbf{x}_i),Z(\mathbf{x}_j))]_{1\leqslant i,j\leqslant n}+[\sigma_\varepsilon^2\delta_{ij}]_{1\leqslant i,j\leqslant n}\\
&\mathbf{v}_{\mathbf{x}}'=[\operatorname{cov}(Z(\mathbf{x}_1),Z(\mathbf{x})),\ldots,\operatorname{cov}(Z(\mathbf{x}_n),Z(\mathbf{x}))]\\
&\mathbf{y}'=[Y_1,\ldots,Y_n]\\
&\mathbf{F}=[h_l(\mathbf{x}_i)]_{1\leqslant i\leqslant n}
\end{aligned}$$

where $'$ denotes the transpose and $\delta_{ij}=1$ for $i=j$, and $\delta_{ij}=0$ for $i\neq j$. Note that the covariance matrix, $\mathbf{V}$, includes contributions from the random error, which was not considered in the original EGO method. The best linear predictor (BLP) of Y is:

$$\hat{Y}(\mathbf{x})=\mathbf{h}_{\mathbf{x}}'\hat{\boldsymbol{\beta}}+\mathbf{v}_{\mathbf{x}}'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{F}\hat{\boldsymbol{\beta}}) \tag{5}$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{F}'\mathbf{V}^{-1}\mathbf{F})^{-1}\mathbf{F}'\mathbf{V}^{-1}\mathbf{y}$ is the generalized least squares estimate of $\boldsymbol{\beta}$. And the mean squared error (MSE) of prediction can be obtained as:

$$s^2(\mathbf{x}) = \sigma_z^2 - [\mathbf{h}_\mathbf{x}', \mathbf{v}_\mathbf{x}'] \begin{bmatrix} \mathbf{0} & \mathbf{F}' \\ \mathbf{F} & \mathbf{V} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{h}_\mathbf{x} \\ \mathbf{v}_\mathbf{x} \end{bmatrix} \tag{6}$$

2.4. MAXIMUM LIKELIHOOD ESTIMATION OF THE PARAMETERS

In this paper, we focus on maximum likelihood estimation (MLE) of the parameters $\sigma_Z^2, \sigma_\varepsilon^2$, and θ_i (for $i = 1, \ldots, d$) in the formulations (4)–(6). Denote by $\mathbf{R}$ the correlation matrix between data $Y_1, Y_{2,}, \ldots, Y_n$, i.e., $\mathbf{R} = \mathbf{V}/(\sigma_Z^2 + \sigma_\varepsilon^2)$. Thus, for the ijth component of $\mathbf{R}$, we have:

$$R_{ij} = \begin{cases} 1 & (i = j) \\ g R_Z(\mathbf{x}_i, \mathbf{x}_j) & (i \neq j) \end{cases} \tag{7}$$

where $g = \sigma_Z^2/(\sigma_Z^2 + \sigma_\varepsilon^2)$. Note that here $\mathbf{R}$ depends only on the scale parameters, θ_i (for $i = 1, \ldots, d$) and the parameter g. Also, the value $(1 - g)$ is referred to as the 'nugget' in geo-statistics (Cressie, 1993).

Sacks (1989a, b) derived the likelihood (omitting constants):

$$p(\mathbf{y}|\mathbf{R}) \sim \frac{1}{(\det R)^{1/n}\hat{\sigma}^2}, \quad \text{where } \hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{F}\hat{\boldsymbol{\beta}})'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{F}\hat{\boldsymbol{\beta}}) \tag{8}$$

Maximizing (8), we can obtain estimates for θ_i (for $i = 1, \ldots, d$) and g, and then compute $\hat{\sigma}^2$. As $\hat{\sigma}^2 = \hat{\sigma}_Z^2 + \hat{\sigma}_\varepsilon^2$, with g known, we can also get the estimates for σ_Z^2 and σ_ε^2.

2.5. THE DESIGN FOR INITIAL FIT

Step 1 described in Section 2.2 involves an experimental design for the initial kriging fit. Design problems of this type have been studied in a wealth of the literature on design and analysis of computer experiments (DACE). There are two main design strategies: space-filling and criterion-based. Space-filling designs include Latin hypercube designs (McKay et al., 1979; Stein, 1987), uniform designs (Fang et al., 2000), and Monte Carlo designs (Niederreiter, 1992). Design criteria that have been proposed include entropy (Lindley, 1956), MSE (Sacks et al., 1989a, b), and maximin and minimax distance (Johnson et al., 1990). For summaries on this area, see Santner et al. (2003) and Koehler, et al. (1996).

In this research, we follow Jones et al. (1998) and use Latin hypercube designs that maximize the minimum distance between design points. The MATLAB® Statistics toolbox subroutine 'lhd' was used to generate

these designs. The number of points for the initial design may depend on the user's prior information about the objective function. That is, if it is believed that the objective contains many non-negligible fine features (very 'bumpy'), more points should be used. Otherwise, fewer points may be adequate. In this study, we adopt the 'rule of thumb' by Jones et al. (1998) for the default number of points. This number is $10 \times d$, where d is the dimension of the input space. In addition, to estimate the random errors, after the first $10 \times d$ points are evaluated, we add one replicate at each of those locations, where the 1st-dth best responses are found (thus there are $11 \times d$ total points in the initial design).

For one-dimensional cases, the Latin hypercube design has evenly spaced points. In this case, when the number of points is few (usually <8), the maximum likelihood estimates (MLE) sometimes overestimate θ_1 and hence can generate a low-quality prediction. This phenomenon was also reported by Sasena (2002). However, this issue does not appear to pose a serious threat to our optimization scheme, because subsequent samples break up the even spacing and improve the prediction. In addition, Santner et al. (2003) suggested that restricted maximum likelihood (REML) can be used to mitigate-related problems.

2.6. ALTERNATIVE EXPECTED IMPROVEMENT FUNCTIONS

As mentioned in Section 2.2, Step 3, EI is the criterion for selecting the location of the subsequent infill point. For deterministic problems, Jones et al. (1998) used the following EI function:

$$E[I(\mathbf{x})] \equiv E\left\{\max\left\lfloor f(\mathbf{x}^*) - Y_p(\mathbf{x}), 0\right\rfloor\right\}. \tag{9}$$

where $\mathbf{x}^*$ is the current best solution and where

$$Y_p(\mathbf{x}) \sim N\left\lfloor \hat{Y}(\mathbf{x}), s(\mathbf{x})^2 \right\rfloor. \tag{10}$$

Note that $Y_p(\mathbf{x})$ can be interpreted as the Bayesian estimate for the posterior distribution of an unknown function [for detailed Bayesian interpretation of the kriging model, refer to the Bayesian approaches for modeling computer experiments by Currin et al. (1991) and O'Hagan (1989)]. However, in the context of stochastic function evaluations, as the true objective values at sampled locations are not known for sure, the applicability of Equation (9) is unclear. Two difficulties are involved here: (1) the current best solution, $\mathbf{x}^*$, is not well-defined, and (2) the prediction uncertainty associated with $f(\mathbf{x}^*)$ is not accounted for.

In a different but related context, Williams et al. (2000) considered a sequential design of computer experiments to minimize the integrated

response function over a set of uncontrollable 'environmental' variables. From this work, the following EI formulation is implied:

$$E[I(\mathbf{x})] \equiv E\left\lfloor \max\left(Y_p(\mathbf{x}^*) - Y_p(\mathbf{x}), 0\right)\right\rfloor. \tag{11}$$

This formulation addresses the prediction uncertainty associated with the current best solution. However, it leads to $E[I(\mathbf{x})] \to 0$, when $\mathbf{x} \to \mathbf{x}^*$, meaning that no replicate would ever be run at the location that is considered the current best. This property is not desirable because there is benefit from a replicate at the current best solution as it reduces the prediction error. In addition, computation of the expectation is relatively expensive.

2.7. THE PROPOSED AUGMENTED EI FUNCTION

As indicated above in Equation (11), one route towards developing an infill selection criterion for stochastic evaluations would be to redefine the improvement function, $I(\mathbf{x})$, and calculate the expectation of it. In this paper, we adopt a different approach that modifies the original EI to provide certain desirable properties. We propose the following augmented EI function:

$$EI(\mathbf{x}) \equiv E\left[\max\left(\hat{Y}(\mathbf{x}^{**}) - Y_p(\mathbf{x}), 0\right)\right] \cdot \left(1 - \frac{\sigma_\varepsilon}{\sqrt{s^2(\mathbf{x}) + \sigma_\varepsilon^2}}\right) \tag{12}$$

where $\mathbf{x}^{**}$ stands for the current 'effective best solution,' which is determined as explained below. In Equation (12), the expectation is conditional given the past data and given estimates of the correlation parameters. Therefore, the expectation is computed by integrating over the distribution of $Y_p(\mathrm{x})$ in Equation (9), with $\hat{Y}(\mathbf{x}^{**})$ a fixed value. Based on results in Jones et al. (1998), the expectation can be calculated analytically as follows:

$$\begin{aligned} E\left[\max\left(\hat{Y}(\mathbf{x}^{**}) - Y_p(\mathbf{x}), 0\right)\right] = {} & \left(\hat{Y}(\mathbf{x}^{**}) - \hat{Y}(\mathbf{x})\right) \Phi\left(\frac{\hat{Y}(\mathbf{x}^{**}) - \hat{Y}(\mathbf{x})}{s(\mathbf{x})}\right) \\ & + s(\mathbf{x}) \phi\left(\frac{\hat{Y}(\mathbf{x}^{**}) - \hat{Y}(\mathbf{x})}{s(\mathbf{x})}\right) \end{aligned} \tag{13}$$

where Φ and ϕ are the standard normal probability density and cumulative distribution functions, respectively. Note that Equation (13) is non-zero when $\mathbf{x} = \mathbf{x}^{**}$, which is a good property because replicates at the effective best solution reduce the prediction uncertainty.

To determine the effective best solution, $\mathbf{x}^{**}$, we introduce a utility function, denoted by $u(\mathbf{x})$, to account for the uncertainty associated with the

predicted objective values. In general, the form of the utility function may be selected according to the user's preference. In this study, we adopt the following formula:

$$u(\mathbf{x}) = -\hat{Y}(\mathbf{x}) - cs(\mathbf{x}) \tag{14}$$

where c is a constant that can reflect the degree of risk aversion. We select $c = 1.0$ as our default, which implies a willingness to trade 1 unit of the predicted objective value for 1 unit of the standard deviation of prediction uncertainty.

The effective best solution, $\mathbf{x}^{**}$, can be derived by maximizing $u(\mathbf{x})$ over the entire feasible region. This approach, however, can be computationally costly. As the locations that have never been observed presumably contain larger uncertainty, with little loss but noticeable computational benefit, here we maximize $u(\mathbf{x})$ over previously observed locations, $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_n$, i. e.:

$$\mathbf{x}^{**} = \underset{\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_{1n}}{\arg\max} \left[u(\mathbf{x})\right]. \tag{15}$$

Thus, the first term (the expectation term) in Equation (12) represents how much we expect the objective value at $\mathbf{x}$ to be better than $\hat{Y}(\mathbf{x}^{**})$, which is the prediction at the current best solution. The term reduces to the original EI function of Jones et al. (1998) in equation (9), when σ_ε goes to zero. However, unlike in Equation (11), $\hat{Y}(\mathbf{x}^{**})$ is not included in the expectation, so this term is not strictly how much the objective value at $\mathbf{x}$ is expected to be better the objective value at $\mathbf{x}^{**}$.

The second term of the product in Equation (12) is a factor designed to account for the diminishing return of additional replicates as the prediction becomes more accurate. Note that at (and only at) $\mathbf{x}^{**}$, the expectation term in Equation (12) is proportional to the posterior standard deviation of the objective value. Therefore, inspired by Bayesian concepts, this factor is equal to the ratio of the reduction in the posterior standard deviation after a new replicate is added. This factor approaches one when the variance of the random errors approaches zero.

In summary, we feel that the augmented EI criterion in Equation (12) is consistent with our intuition about the 'usefulness' of the additional sample at a particular location. This partially intuition-based infill sampling criterion is further justified by the performance of the algorithm as shown in later sections.

From Section 2.2, Step 3, the location of the next evaluation, $\mathbf{x}_{n+1}$, is selected by maximizing EI, i.e.:

$$\mathbf{x}_{n+1} = \underset{\mathbf{x}}{\arg\max}\, EI(\mathbf{x}). \tag{16}$$

Here the EI maximization problem is solved using a Nelder–Mead simplex approach.

2.8. THE STOPPING CRITERION

As mentioned in Section 2.2, the optimization scheme stops when

$$\max_{\mathbf{x}} EI(\mathbf{x}) < \Delta_s, \tag{17}$$

where Δ_s is the stopping tolerance. To accommodate stochastic evaluations, we generally require (17) to be satisfied a number of times consecutively before the final stopping. In this paper, we use $(d+1)$ for this number of times.

In practice, we also found sometimes it is convenient to use a variation of (17):

$$\frac{\max_{\mathbf{x}} EI(\mathbf{x})}{\max(Y_1, Y_1, \ldots, Y_n) - \min(Y_1, Y_1, \ldots, Y_n)} < \Delta_r \tag{18}$$

where the left-hand side is a ratio between the maximal EI and the 'active span' of the responses, which we also refer to as the maximal 'relative EI.' An advantage of this stopping criterion is that the user can set the 'relative' tolerance, Δ_r, without having to know the magnitudes of the problem responses.

3. An Illustrative Example

In this section, we use the two-dimensional six-hump camel back function (Branin, 1972) to illustrate search patterns of the proposed SKO optimization method. The formula for the function is available in Table 1, and a surface plot of the function is displayed in Figure 1. The function is symmetric about the origin and has three pairs of local minima, which include a pair of global minima. To simulate measurement errors, we add independent random errors that are $\sim \mathrm{N}(0, 0.12^2)$.

In Figure 2, the search pattern of the SKO method is displayed together with the contour of the objective function. There are a total of 44 evaluation points, consisting of three groups. The initial 20 points are based on a Latin hypercube design, the following two replicates are placed on the two best points from the initial design, and thereafter infill points are selected by maximizing the EI criterion. For the infill points, the sequence of the evaluations is marked in numbers. Note that the EI criterion balances the need to exploit the meta-model (local search) and the need to improve the meta-model (global search). Thus, some points are

Table 1. Test functions

Name	Function description	Source
Six-hump camel back	$d=2$ $f(x)=4x_1^2-2.1x_1^4+1/3x_1^6+x_1x_2+4x_2^2+4x_2^4$ $-1.6\leqslant x_1\leqslant 2.4,\ \ -0.8\leqslant x_2\leqslant 1.2$ $N_{\text{local}}=6, N_{\text{global}}=2$ $\mathbf{x}^*=(0.089,-0.713)$ and $(-0.089, 0.713)$, $f^*=-1.03$	Branin (1972)
'Tilted' Branin	$d=2$ $f(x)=\left(x_2-\frac{5.1}{4\pi^2}x_1^2+\frac{5}{\pi}x_1-6\right)^2+10\left(1-\frac{1}{8\pi}\right)\cos x_1+10+0.5x_1$ $-5\leqslant x_1\leqslant 10, 0\leqslant x_2\leqslant 15$ $N_{\text{local}}=3, N_{\text{global}}=1$ $\mathbf{x}^*=(-3.2, 12.3)$, $f^*=-1.17$	Modified based on Branin et al. (1972)
Hartman 3	$d=3$ $f(x)=-\sum_{i=1}^{4}c_i\exp\left[-\sum_{j=1}^{3}\alpha_{ij}(x_j-p_{ij})^2\right]$ where $\alpha_{ij}=\begin{bmatrix}3&10&30\\0.1&10&35\\3&10&30\\0.1&10&35\end{bmatrix}$ $c_i=\begin{bmatrix}1\\1.2\\3\\3.2\end{bmatrix}$ $p_{ij}=\begin{bmatrix}0.3689&0.1170&0.2673\\0.4699&0.4387&0.7470\\0.1091&0.8732&0.5547\\0.03815&0.5743&0.8828\end{bmatrix}$ $0\leqslant x_i\leqslant 1.0$, for $i=1, 2, 3$, and $j=1,\ldots,4$ $N_{\text{local}}>1, N_{\text{global}}=1$ $\mathbf{x}^*=(0.114, 0.556, 0.852)$, $f^*=-3.86$	Hartman (1973)
Ackley 5 $([-32.8, 32.8]^d)^\dagger$	$d=5$ $f(x)=-a\exp\left[-b\sqrt{\frac{1}{n}\sum_{i=1}^{d}x_i^2}\right]-\exp\left[\frac{1}{n}\sum_{i=1}^{d}\cos(cx_i)\right]+a+\exp(1),$ $a=20; b=0.2; c=2\pi$ $-32.8\leqslant x_i\leqslant 32.8$, for $i=1,\ldots,$ d $N_{\text{local}}>1, N_{\text{global}}=1$ $\mathbf{x}^*=(0., 0., 0.)$, $f^*=0.0$	Ackley (1987)
Ackley 5$([-2, 2]^d)^\ddagger$	Same as above except $-2\leqslant x_i\leqslant 2$	Ackley (1987)

For DIRECT, Ackley 5s region of interest is 'shifted' by 10%, such that the starting point does not coincide with the global optimum at (0,0), i.e., †: $-26.1\leqslant x_i\leqslant 39.6$;‡: $-1.6\leqslant x_i\leqslant 2.4$.

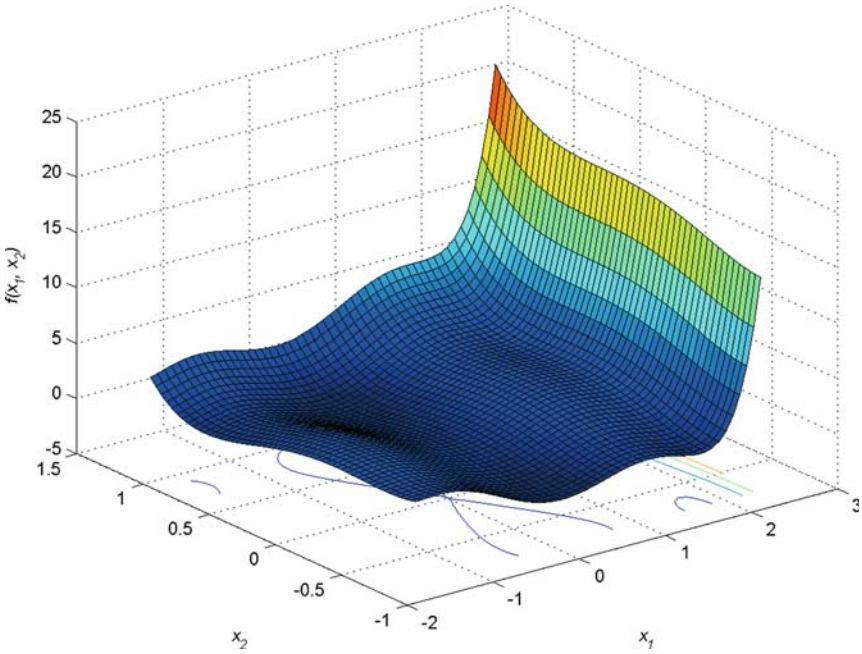

Figure 1. Surface of the six-hump camel back function (Branin, 1972).

placed near the global minima, while others are placed in 'unexplored' territory away from the global minima. It is random which of the two global minima will eventually be selected as the best solution.

As compared to most conventional methods, an additional advantage of the SKO method is that at the end it provides a global meta-model of the responses. Figure 3 displays the final kriging prediction as well as the different between the prediction and the true objective function. Note that the prediction gives a good match in general, while it provides more accurate information on areas that are near the global optima. This may be particularly helpful when some local minima are nearly as good as global minima and considerations not captured in the objective function are relevant to decision-makers.

4. Empirical Comparison with Alternative Methods

In this section, the proposed SKO optimization method is empirically compared with three alternative approaches from the literature. The first alternative considered is the SPSA from Spall (1992, 1998). The second approach is the revised simplex search (RSS) procedure, a variant of the Nelder–Mead method, proposed by Humphrey and Wilson (2000). As the original RSS procedure was not designed to handle constraints, we modify

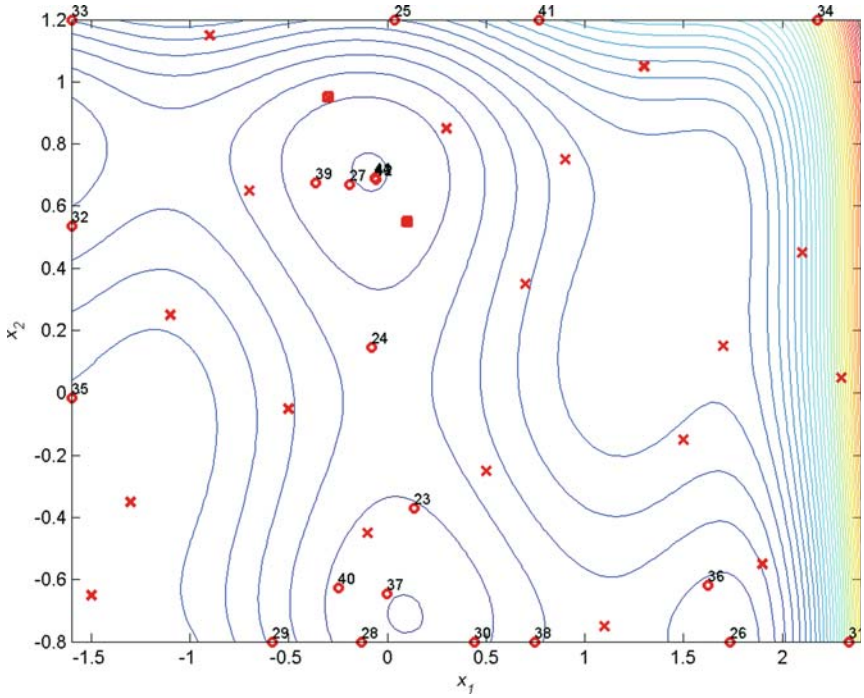

Figure 2. Evaluation points and contour of the six-hump camel back function. ×: initial-fit design (20 points); □: replicates (2 points); ◯: infill samples (22 points, with numbers indicating the sequence).

it so that whenever a point is infeasible, the algorithm selects the nearest feasible point instead. The third approach is the DIRECT method developed by Gablonsky and Kelley (2001).

Five test functions are used to compare these optimization methods. They are listed in Table 1, where we document the number of dimensions, objective function formula, region of interest, number of local minima (N_{local}), number of global minima (N_{global}), and global minima $\mathbf{x}^*$, f^*. The dimensions of these test functions range from two to five. The 'Tilted' Branin function was created by adding a term $0.5x_1$ to the original Branin function (Branin and Hoo, 1972), so that not all local minima are global minima. We also included two versions of a five-dimensional test problem from Ackley (1987). The first problem uses the traditional feasible region $[-32.8, 32.8]^5$ and has hundreds of local minima, causing all of the alternative methods to fail to find the global minimum in our experiments. The second version involves a relatively small number of local minima with a restricted region of interest $[-2, 2]^5$.

The noises associated with the response are normally distributed in the tests. The size of the noise was selected so that the standard deviation was

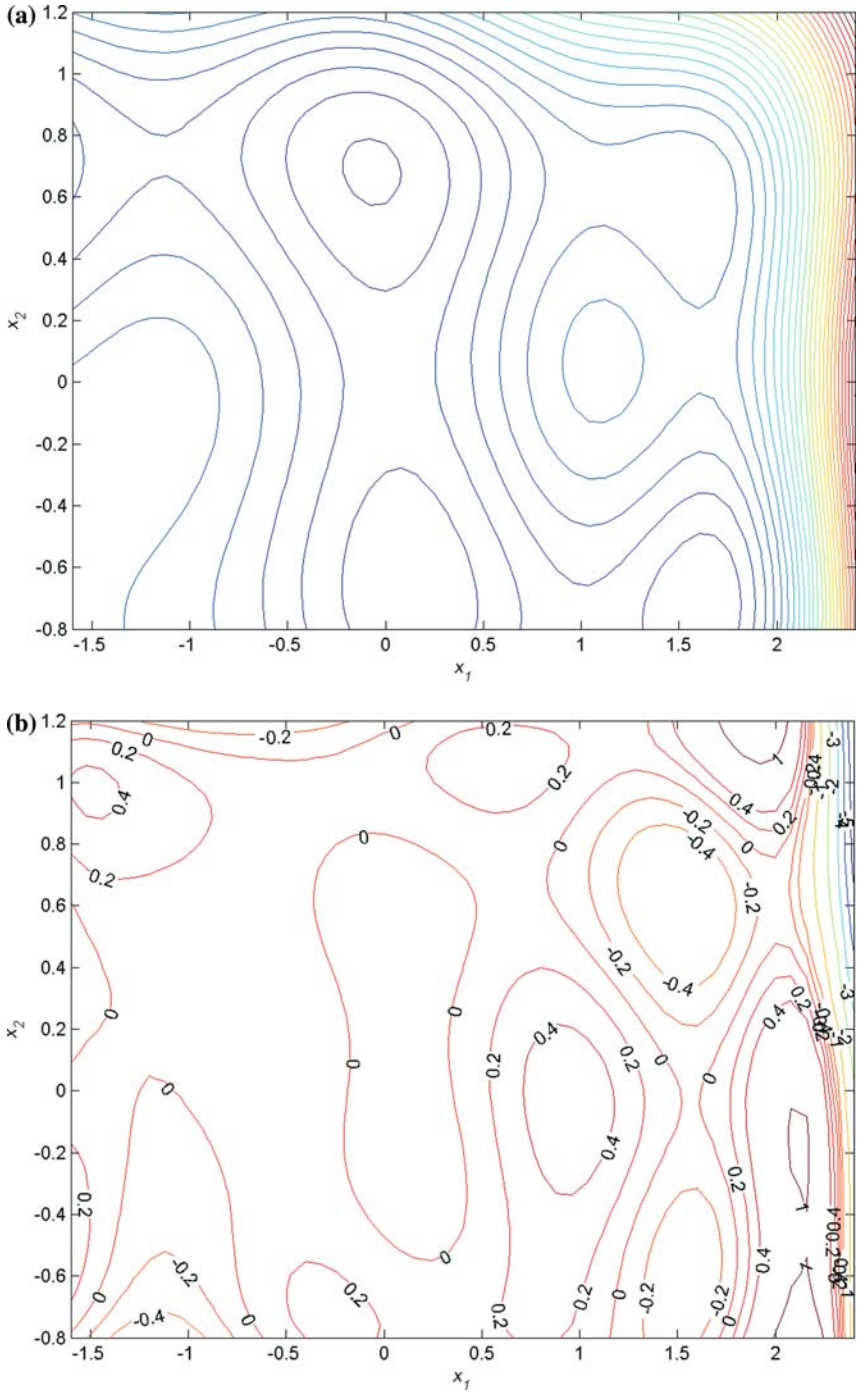

Figure 3. (a) The final kriging prediction. (b) The difference between the prediction and the true function.

approximately $0.5-4\%$ of the range of the responses in the region of interest. Each optimization run is independently randomized. For the SPSA and RSS methods, each run has a random starting point. For the SKO method, the initial-fit designs undergo random symmetry operations. Unfortunately, for the DIRECT method, the starting points are always fixed, which often causes all runs in a test to have similar paths. This fact limits our ability to draw more general conclusions on the DIRECT method using a small number of test functions.

We documented the histories of the true objective values corresponding to the current best solutions in solving the six-hump camel back function with noise $\sim \mathrm{N}(0, 0.12^2)$. In Figure 4, the results of 10 runs are displayed for every method. Note that, due to the random errors, sometimes optimization algorithms can be 'mislead' and the objective function values may become worse. Visual examination of this graph suggests that the SKO and DIRECT methods consistently find the global minima, while the other two methods sometimes converge to non-global local minima. For the runs that reach global minima, Figure 5 shows the average, the best 10%, and the worst 10% performance out of 50 random runs. In this test, SKO and DIRECT seem to have better efficiency that that of SPSA and RSS. Also, SPSA and RSS have relatively larger variance in the objective histories.

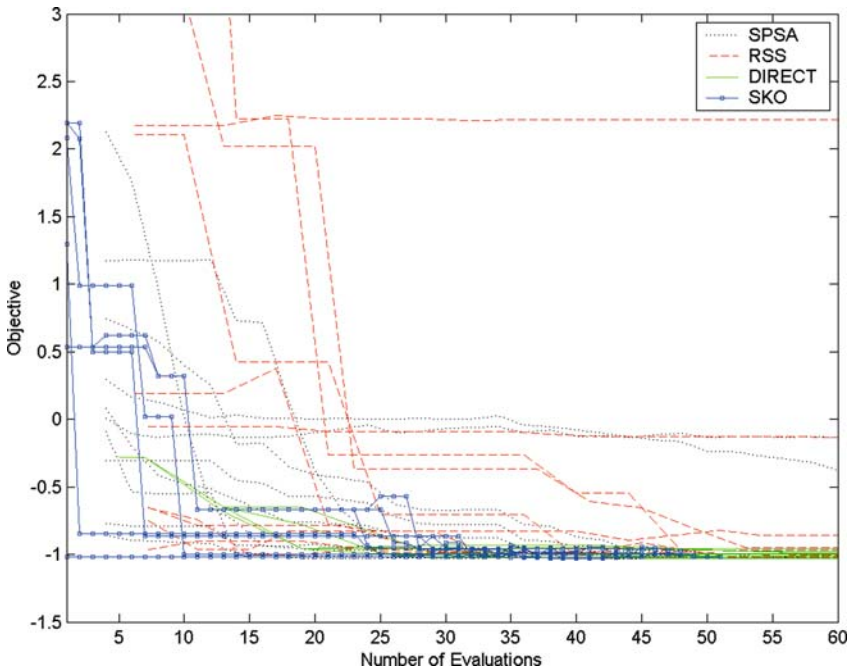

Figure 4. Objective histories in solving the six-hump camel back function (10 random runs each method).

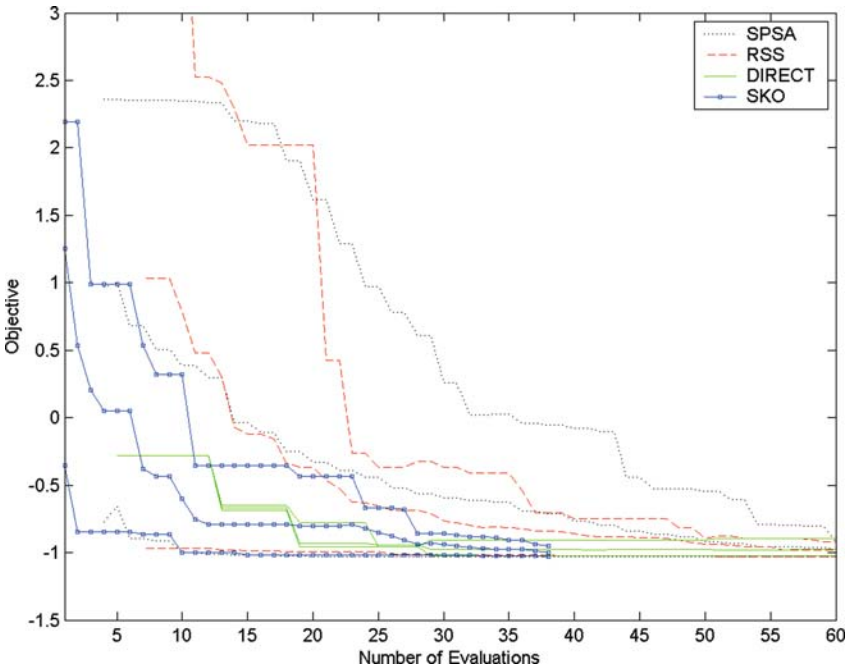

Figure 5. The average, the best 10%, and the worst 10% performance of the runs that reach global optima.

Next, we consider the efficiency of the optimization algorithms by investigating the number of function evaluations needed to come within a certain relative distance of the true global minima. Following Barton (1984), we define the following measure for iteration i:

$$G_i = \frac{f(\mathbf{x}^{(1)}) - f(\mathbf{x}^{(i)})}{f(\mathbf{x}^{(1)}) - f^*} \tag{19}$$

where f^* stands for the true global optima. Therefore, $f(\mathbf{x}^{(1)}) - f^*$ is the gap between the starting value and the global minima, and G_i describes the reduction of this gap that is achieved after i iterations of the optimization run (for SKO, where the initial design consists of several points that are all starting values, we choose $f(\mathbf{x}^{(1)})$ equal to the median of these values). In this paper, we evaluate the efficiency of the algorithms using $S_{0.99}$, which is the cumulative number of evaluations performed until $G \geqslant 0.99$.

Each optimization method has different parameters to control the stopping. Here we adjusted the stopping criteria to give every method a reasonable chance of finding the global optimum. In other words, these criteria are 'tight', such that $G \geqslant 0.99$ can usually be achieved if the search is not 'trapped' in the local optima. The stopping criteria are listed in Table 2. Note that the DIRECT method does not have a specific parameter

Table 2. Stopping criteria

SPSA	Step size is less than 10^{-8} in the normalized input space $[0, 1]^d$.
RSS	Simplex size is less than 10^{-8} in the normalized input space $[0, 1]^d$.
DIRECT	Number of evaluations exceeds $200 \times d$ (d: dimension).
SKO	Maximal relative EI is less than 0.0005.

Table 3. Percentage of runs that reach $G \geqslant 0.99$, average of $S_{0.99}$, and standard deviation of $S_{0.99}$ (in parentheses)

No.	Test functions	Noise	SPSA	RSS	DIRECT	SKO
1	Six-hump camel back	$\sim$ N (0, 0.12^2)	70%, 40.7 (28.2)	86%, 51.3 (29.7)	100%, 22.3 (14.6)	100%, 29.2 (5.7)
2	Six-hump camel back	$\sim$ N (0, 0.24^2)	72%, 57.5(37.3)	82%, 65.3 (53.9)	96%, 39.2 (37.2)	94%, 29.4 (6.6)
3	'Tilted' Branin	$\sim$ N (0, 2.0^2)	20%, 51.8 (42.5)	22%, 58.8 (43.5)	100%, 37.2 (33.6)	98%, 28.4 (5.3)
4	Hartman 3	$\sim$ N (0, 0.08^2)	44%, 127.5 (145.2)	42%, 87.5 (64.1)	54%, 213.6 (132.4)	96%, 45.4 (7.9)
5	Ackley 5 ($[-2, 2]^d$)	$\sim$ N (0, 0.06^2)	36%, 856.1 (468.6)	38%, 310.4 (132.6)	100%, 248.4 (84.3)	94%, 98.9 (5.6)

for stopping, so a limit on the number of evaluations is imposed instead (in theory, it is possible that infinitely many evaluations may be needed to determine $S_{0.99}$, but to run the methods indefinitely is not practical). For each test, 50 random runs are conducted for each method. The percentages of the runs that reach $G \geqslant 0.99$ are recorded. And among these runs, averages and standard deviations of $S_{0.99}$ are computed. The results of the tests are listed in Table 3.

Based on the results in Table 3, we have the following findings. (1) As we expected, in general the global methods, DIRECT and SKO, have considerably higher percentages of runs that find the global optima than the local methods, SPSA and RSS. Specifically, the SKO method has good consistency in all test cases. The DIRECT method has high percentages in most cases, except for the Hartman 3 function. The exception may be because the number of evaluations, $200 \times d$, is not enough in this case. (2) Considering only the runs that do find global minima, the number of evaluations ($S_{0.99}$) needed by SKO is the least on average for the majority of the test cases, except test No. 1, where DIRECT is the most efficient method. But DIRECT's efficiency is poor in the Hartman 3 case. The fact that SKO, being a global method, is consistently more efficient than the local methods is impressive, because resources are required to ensure the globality of the solution. (3) Probably more importantly, the standard deviations of $S_{0.99}$

for SKO are significantly less than those for SPSA, RSS, and DIRECT. (4) In addition, data from test Nos. 1 and 2 suggest that the performance of SKO is less sensitive to increased noise levels than other methods. Observations (3) and (4) are presumably because SKO is based on a global meta-model of the expected outputs, which 'smoothes out' the affect of noises. (5) At last, the efficiency advantages of SKO seem to become more pronounced as the number of dimensions increases.

The Ackley 5 function with the traditional region of interest $[-32.8, 32.8]^5$ is a special problem here, because it is very 'bumpy,' containing hundreds of local minima. Using the default experimental design for initial fit, we find that the kriging models developed during specific test runs have large apparent prediction inaccuracies (this is not surprising considering that the number of local minima greatly exceeds the number of samples). As the kriging approximation is poor, we observe that the search pattern by SKO appears not better than a random search. And we expect that the proposed SKO method may not perform well.

As mentioned previously, for the Ackley 5 function in $[-32.8, 32.8]^5$, with noise $\sim \mathrm{N}(0, 0.06^2)$, none of the four compared methods was able to find the global minimum in any of our test runs Figure 6. Thus, here we compare the objective values of the best solutions found after 150 evaluations. The results of 25 independent runs for each method are displayed in Figure 4. As evaluated using the mean objective value achieved, the DIRECT method seems to provide the best performance in this test. However, as mentioned previously, because the DIRECT method uses a fixed set of starting points, all runs lead to very similar results. Interestingly, SKO also has the relatively good performance in general, followed by RSS, and then SPSA. We speculate that, although the kriging approximation is not ideal, the global search component in the EI criterion provides some heuristic-like mechanism to look into regions not previously visited (i.e., for which there is large uncertainty about the predicted objective value) for the global optimum. Also, the kriging meta-modeling approach combined with the utility maximization selection methods together may provide added assurance that the best solutions encountered are not lost. We suggest that additional investigation of situations in which the initial kriging model is highly inaccurate merit further study.

5. Application Example: Inventory System

In this section, we apply the SKO method to optimize the ordering policy of an inventory system, which is evaluated using a simulation program provided by Law and Kelton (2000). In this problem, there are two decision variables: the reorder point (x_1) and the maximal holding quantity (x_2). The objective function is the expected total cost per month. The

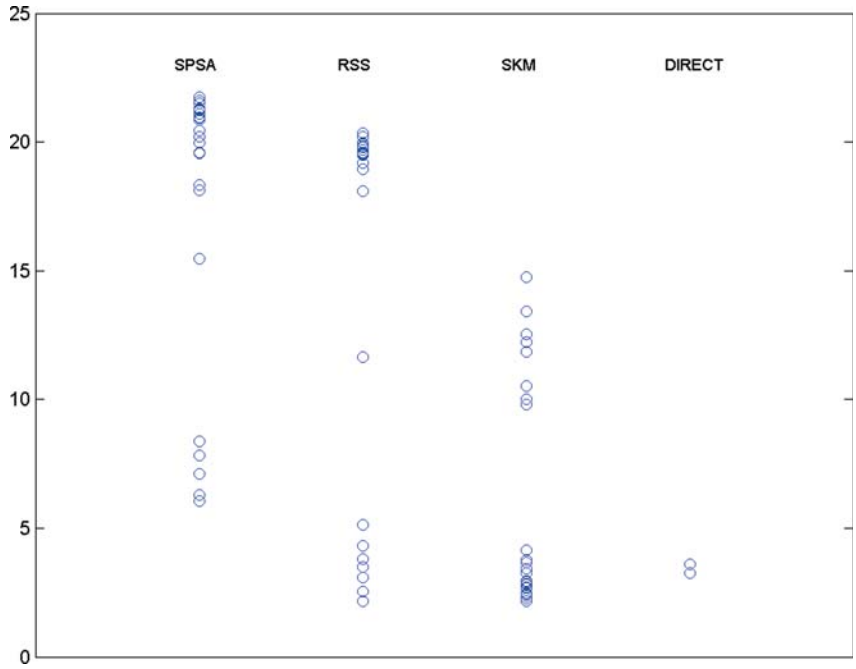

Figure 6. The best objective values after 150 evaluations for Ackley 5 function in $[-32.8, 32.8]^5$ (the range of responses in the feasible region is approximately [0, 22]).

ordering quantity, Z, is decided as follows:

$$Z = \begin{cases} x_2 - I & \text{if } I < x_2 \\ 0 & \text{if } I \geqslant x_2 \end{cases}$$

where I is the inventory level at the beginning of each month. The model is representative of an actual inventory system, which involves ordering cost, holding cost, and shortage cost. The time between demands and the sizes of the demands are generated as random variables; and the inventory system is simulated for a period of 120 months. As the outputs of replicated simulations may vary considerably, we use the average output of 10 replicated simulations as a single function evaluation.

The region of interest is $20 \leqslant x_1 \leqslant 60$ and $40 \leqslant x_2 \leqslant 100$. The initial-fit points and the sequential infill points are indicated in Figure 7. The stopping criterion is when the maximal relative EI is less than 0.001, which was reached in a total of 47 evaluations. The optimal solution is found at (25.05, 62.16), where the expected total cost per month equals 116.83. In addition, kriging global meta-models are created as by-products of the optimization, which provides useful visualization for understanding the simulation outputs in this case. Figure 8 (a) shows the contour of the final kriging prediction. Also the variance of the evaluation random errors is

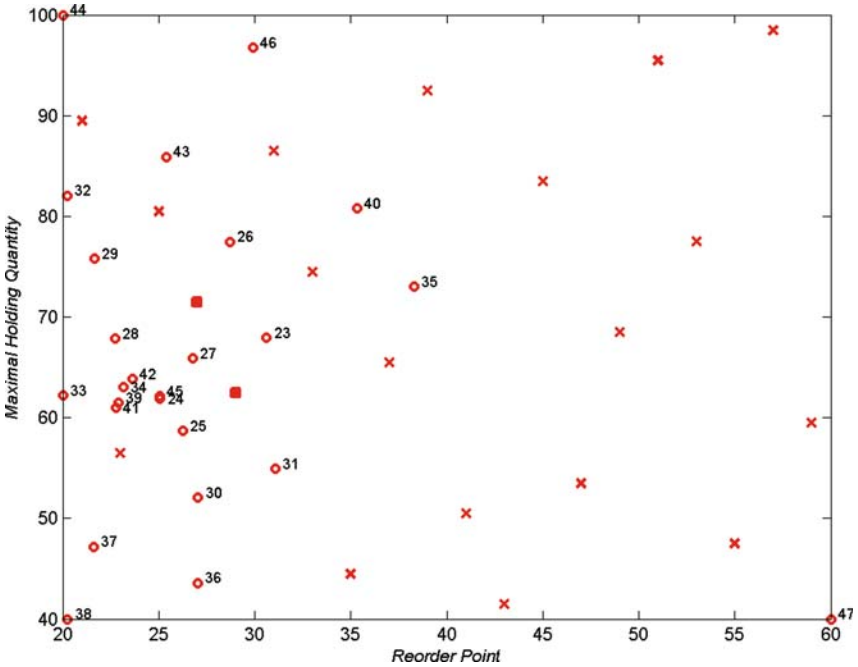

Figure 7. Evaluation points for solving the inventory system problem. ×: initial-fit design (20 points); □: replicates (2 points); ◯: infill samples (25 points, with numbers indicating the sequence).

estimated to be 0.31. Furthermore, from Equation (6), we can compute the MSE of the prediction, which is displayed in Figure 8 (b). Note that this figure is different from Figure 3 (b), where the actual discrepancy between the prediction and the true function is shown (here the true function is not known). As we expected, the uncertainty of the prediction is low where more evaluation points are available and high, where evaluation points are sparse.

We also used alternative methods, RSS, SPSA, and DIRECT, to solve this simulation optimization problem and their performances are compared. In this application study, as the true objective function is non-analytical, we know neither the true optimum value, nor exactly how good our current solution is. Therefore, a rigorous comparison is relatively difficult. Here in Figure 9 we display the history of the 'best response observed so far' of an arbitrary run for every method (note that Figure 9 is unlike Figures 4 and 5 which display the true objective values corresponding to the current best solution). Figure 9 shows that the local methods, SPSA and RSS, and the global methods, DIRECT and SKO, converged to the same solution. This is not a surprise, because from the meta-model [Figure 8 (a)], it seems that the problem has only single local optimum. In this

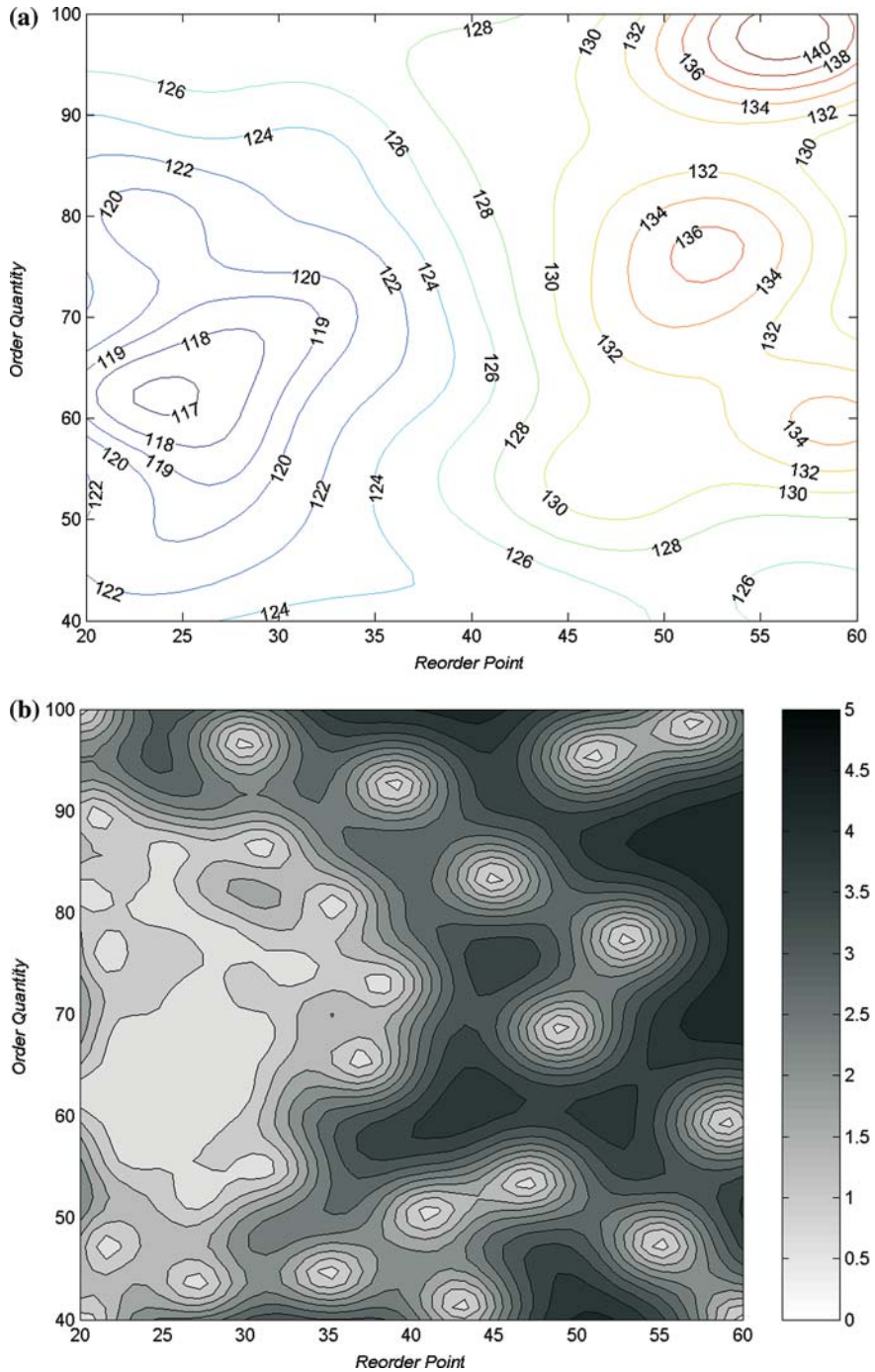

Figure 8. (a) The final kriging prediction of the expected cost per month. (b) MSEs of the final kriging prediction.

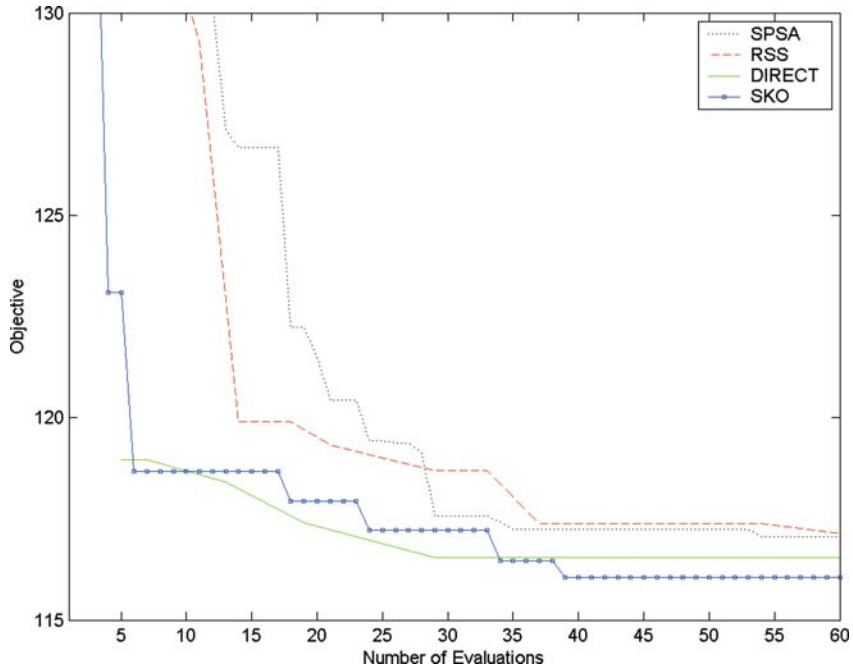

Figure 9. The histories of the best response observed in solving the inventory system problem.

case, DIRECT and SKO demonstrate comparable efficiency, and they outperform RSS and SPSA by only a moderate margin.

6. Limitations and Other Relevant Issues

We find that a limitation of SKO (and any EGO type method) can be due to the algorithm's overhead costs, which include fitting the kriging meta-model and maximizing the EI function. For example, when the number of samples is 150, the computing time per iteration reaches about 60 sec on a Pentium III 1.2G processor. Therefore, the SKO method is suitable only for 'expensive' systems. In particular, probably one should consider SKO only when the cost of a function evaluation is much higher than the cost of fitting a kriging meta-model.

In addition, the cost of fitting kriging meta-models increases as the number of samples increases. And to generate useful predictions, a larger number of samples are needed when the input space dimensionality becomes higher. Therefore, the algorithm may be impractical when the dimensionality is too high. In our studies, the maximum number of dimensions tried was 10.

The problems presented in this paper are all simple bound constrained problems. How ever, the SKO method can be derived to address more

complex constrained problems. The potential approaches include the penalty method, the infeasibility probability method, and the constrained maximization of EI method. For more detailed information, please refer to the review work by Sasena (2002).

7. Conclusions and Future Work

In this paper, we proposed the SKO method as an extension of EGO of Jones et al. (1998) to address stochastic black-box systems. We compared SKO with three relevant alternatives from the literature using six test problems. The proposed SKO method compared favorably with alternatives in terms of consistency in finding global optima and efficiency as measured by number of evaluations. Also, in the presence of noise, the augmented EI function for infill sample selection appears to achieve the desired balance between the need for global and local searches.

Kriging meta-models are very flexible and can approximate a wide variety of functions. However, if there are not sufficient samples or the true objective function is not smooth and lacks any systematic trend, the kriging approximation may be poor and the efficiency of SKO could conceivably drop to the level of a purely random search. Also, an additional limitation of the SKO method is the relatively high overhead cost incurred in fitting the kriging model.

Ongoing efforts are addressing the overhead cost issue by re-utilizing information from the previous iteration. In addition, theoretical analysis on the probability and rate of finding global optima is important for further development of the method. Finally, the performance of SKO under higher levels of noise, for constrained problems, and for very bumpy objective functions needs further investigation.

Acknowledgments

This research was partially funded by scientific forming technologies corporation. We would like to thank Allen Miller and Thomas Santner for ideas, references, and encouragement. In addition, we thank Matthias Schonlau, Daniel Finkel, Michael Sasena, and Ofelia Marin for providing source code, documentation, and discussions. Finally, we thank the anonymous reviewers for valuable suggestions that helped us increase the generality of the proposed method and our method comparison.

References

1. Ackley, D.H. (1987), *A Connectionist Machine for Genetic Hill-climbing*, Kluwer Academic Publishers, Boston.

2. Andradóttir, S. (1998), Simulation optimization. In: *Handbook on Simulation*, Banks, J. (ed.), Chapter 9, Wiley, New York.
3. Angün, E., Kleijnen, J.P., Hertog, D.D. and Gürkan, G. (2002), Response surface methodology revised. In: *Proceedings of the 2002 Winter Simulation Conference*, pp. 377–383.
4. Barton, R.R. (1984), Minimization algorithms for functions with random noise. *American Journal of Mathematical and Management Sciences* 4, 109–138.
5. Barton, R.R. and Ivey, J.S. (1996), Nelder–Mead simplex modifications for simulation optimization, *Management Science* 42, 954–973.
6. Branin, F.H. (1972), Widely convergent methods for finding multiple solutions of simultaneous nonlinear equations, *IBM Journal of Research Developments* 16, 504–522.
7. Branin, F.H. and Hoo, S.K. (1972), A method for finding multiple extrema of a function of n variables. In: Lootsma, F.A. (ed.), *Numerical Methods of Nonlinear Optimization*, Academic Press, London, pp. 231–237.
8. Cressie, N.A.C. (1993), *Statistics for Spatial Data,* revised edition, Wiley, New York.
9. Currin, C., Mitchell, M., Morris, M. and Ylvisaker, D. (1991), Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments, *Journal of American Statistics Association* 86, 953–963.
10. Fang, K.-T., Lin, D.K., Winker, P. and Zhang, Y. (2000), Uniform design: theory and application, *Technometrics* 42, 237–248.
11. Fu, M.C. (1994), Optimization via simulation: a review, *Annals of Operations Research* 53, 199–247.
12. Gablonsky, J.M. and Kelley, C.T. (2001), A locally-biased form of the direct algorithm, *Journal of Global Optimization* 21, 27–37.
13. Haftka, R.T. and G¨urdal, Z. (1993), *Elements of Structural Optimization*, 3rd edition, Kluwer, Boston, MA.
14. Hartman, J.K. (1973), Some experiments in global optimization, *Naval Research Logistics Quarterly* 20, 569–576.
15. Humphrey, D.G. and Wilson, J.R. (2000), A revised simplex search procedure for stochastic simulation response-surface optimization, *INFORMS Journal on Computing* 12, 272–283.
16. Johnson, M.E., Moore, L.M. and Ylvisaker, D. (1990), Minimax and maximin distance designs, *Journal of Statistical Planning and Inference* 26, 131–148.
17. Jones, D., Schonlau, M. and Welch, W. (1998), Efficient global optimization of expensive black-box functions, *Journal of Global Optimization* 13, 455–492.
18. Kiefer, J. and Wolfowitz, J. (1952), Stochastic estimation of a regression function, *Annals of Mathematical Statistics* 23, 462–466.
19. Koehler, J.R. and Owen, A.B. (1996), Computer experiments. In: Ghosh, S. and Rao, C.R. (eds.), *Handbook of Statistics*, Vol. 13, Elsevier Science B.V., Amsterdam, the Netherlands.
20. Kushner, H.J. (1964), A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise, *Journal of Basic Engineering* 86, 97–106.
21. Kushner, H.J. and Clark, D.C. (1978), *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Springer, New York.
22. Law, A.M. and Kelton, W.D. (2000), *Simulation Modeling and Analysis*, third edition, McGraw-Hill, Boston, pp. 60–80.
23. Lindley, D.V. (1956), On a measure of information provided by an experiment, *Annals of Mathematical Statistics* 27, 986–1005.
24. McKay, M.D., Beckman, R.J. and Conover, W.J. (1979), A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics* 21, 239–245.

25. Neddermeijer, H.G., Piersma, N., van Oortmarssen, G.J., Habbema, J.D.F. and Dekker, R. (1999), Comparison of response surface methodology and the Nelder and Mead simplex method for optimization in microsimulation models, *Econometric Institute Report EI-9924/A*, Erasmus University Rotterdam, The Netherlands.
26. Neddermeijer, H.G., van Oortmarssen, G.J., Piersma N. and Dekker R. (2000), A framework for response surface methodology for simulation optimizaton. In: *Proceedings of the 2000 Winter Simulation Conference*, pp. 129–136.
27. Niederreiter, H. (1992), Random number generation and quasi-Monte-Carlo methods, SAIM, Philadelphia.
28. O'Hagan, A. (1989), Comment: design and analysis of computer experiments, *Statistic Science* 4, 430–432.
29. Perttunen, C.D., Jones, D.R. and Stuckman, B.E. (1993), Lipschitzian optimization without the lipschitz constant, *Journal of Optimization Theory and Application* 79(1), 157–181.
30. Rodriguez, J.F., Perez, V.M., Padmanabhan, D. and Renaud, J.E. (2001), Sequential approximate optimization using variable fidelity response surface approximation, *Structure Multidiscipline Optimization* 22, 23–34 (Springer-Verlag).
31. Sacks, J., Welch, W.J., Mitchell, T.J. and Wynn, H.P. (1989a), Design and analysis of computer experiments (with discussion), *Statistical Science* 4, 409–435.
32. Sacks, J., Schiller, S.B. and Welch, W. (1989b), Design for computer experiments, *Technometrics* 31, 41–47.
33. Safizadeh, M.H. and Thornton, B.M. (1984), Optimization in simulation experiments using response surface methodology, *Comparative Industrial Engineering* 8, 11–27.
34. Santner, T.J., Williams, B.J. and Notz, W.I. (2003), *The Design and Analysis of Computer Experiments*, Springer, New York.
35. Sasena, M.J., Papalambros, P.Y. and Goovaerts, P. (2001), The use of surrogate modeling algorithms to exploit disparities in function computation time within simulation-based optimization. In: *Proceedings of the 4th Congress on Structural and Multidisciplinary Optimization*, Dalian, China, June 4–8, 2001.
36. Sasena, M.J., Papalambros, P.Y. and Goovaerts, P. (2002), Exploration of metamodeling sampling criteria for constrained global optimization, *Engineering Optimization* 34, 263–278.
37. Sasena, M.J. (2002), *Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations*, Ph.D. dissertation, University of Michigan.
38. Sóbester, A., Leary, S. and Keane, A.J. (2002), A parallel updating scheme for approximating and optimizing high fidelity computer simulation. In: *The Third ISSMO/AIAA Internet Conference on Approximations in Optimization*, October 14–18, 2002.
39. Spall, J.C. (1992), Multivariate stochastic approximation using a simultaneous perturbation gradient approximation, *IEEE Transactions on Automatic Control* 37, 332–341.
40. Spall, J.C. (1998), Implementation of the simultaneous perturbation algorithm for stochastic optimization, *IEEE Transactions on Aerospace and Electronics Systems* 34, 817–823.
41. Stein, M. (1987), Large sample properties of simulation using Latin hypercube sampling, *Technometrics* 29, 143–151.
42. Williams, B.J., Santner, T.J. and Notz, W.I. (2000), Sequential design of computer experiments to minimize integrated response functions, *Statistica Sinica* 10, 1133–1152.
43. Žilinskas, A. (1980), MINUN – Optimization of one-dimensional multimodal functions in the presence of noise, algorithms 44, *Aplikace Matematiky* 25, 392–402.