Proceedings of the 2016 Winter Simulation Conference
T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, eds.

SIMULATION OPTIMIZATION VIA PROMISING REGION SEARCH AND SURROGATE MODEL APPROXIMATION

Qi Fan
Jiaqiao Hu

Department of Applied Mathematics and Statistics
State University of New York at Stony Brook
John S. Toll Drive
Stony Brook, NY 11794, USA

ABSTRACT

We propose a random search algorithm for solving simulation optimization problems with continuous decision variables. The algorithm combines ideas from promising area search, shrinking ball methods, and surrogate model optimization. We discuss the local convergence property of the algorithm and provide numerical examples to illustrate its performance.

1 INTRODUCTION

Many systems arising in engineering design and manufacturing require the use of simulation optimization (SO) techniques to improve their performance. In contrast to their deterministic counterparts, SO problems are typically more difficult to solve because additional simulation effort needs to be expended to deal with the noisy measurement of the objective function. For SO problems with little structure, one popular and effective method is to use random search. This encompasses a broad class of algorithms that use a sequence of randomly generated iterates (e.g., candidate solutions, probability models, promising subsets) to approximate the optimal solution. Examples of random search techniques include the stochastic ruler method (Yan and Mukai 1992), simulated annealing (Alrefaei and Andradóttir 1999), the nested partitions method (Shi and Ólafsson 2000), adaptive shrinking ball methods (Andradóttir and Prudius 2010), COMPASS (Hong and Nelson 2006, Xu et al. 2010), and model-based methods (Rubinstein and Kroese 2004, Hu et al. 2008). These algorithms primarily differ in the type of iterates an algorithm produces and the random strategy used to generate the iterates; see, e.g., Andradóttir (2014), Hu (2014) and Zabinsky (2014) for recent reviews of these techniques. When computer experiments are computationally expensive, it is often desirable to use surrogate (metamodels) to represent simulation input-output relations. This has motivated the use of surrogate-based or response surface methods (RSMs) for simulation or "black-box" optimization. A variety of RSMs have been proposed in the literature (e.g., Jones et al. 1998, Gutmann 2001, Nakayama et al. 2002, Sóbester et al. 2005, Regis and Shoemaker 2007), and their applications to (stochastic) simulation optimization can be found in e.g., Huang et al. (2006), Chang et al. (2013), and Kleijnen (2014).

In this paper, we propose an algorithm that integrates ideas from shrinking ball methods, promising area search, and surrogate model optimization for solving SO problems. The algorithm proceeds iteratively by constructing and optimizing a sequence of surrogate models, which are approximations of the objective function on promising subsets of the solution space. Each iteration of the algorithm consists of the following three basic steps: (1) Generate a set of candidate solutions by randomly sampling from the promising region constructed in the previous iteration, and use the shrinking ball technique (Andradóttir and Prudius 2010) to estimate the performance of the sampled solutions; (2) Based on candidate solutions generated in the current iteration, build a surrogate model of the objective function and optimize the fitted model; (3) A

new promising region containing an optimal solution to the surrogate model is then constructed by using an approach similar to that of Hong and Nelson (2006). Intuitively, the shrinking ball method reduces the simulation noise at a sampled solution by averaging observations at solutions that are close to it, thus avoiding the need to allocate multiple simulation replications to the same point. The use of promising area helps to concentrate the computational effort on subsets of the solution space. Additionally, the surrogate model uses sampling information to successively predict the response surface of the objective function, and when employed in conjunction with promising area search, may quickly direct the search to regions containing high quality solutions at no extra simulation effort. Under some appropriate conditions, we show that the algorithm converges with probability one to the set of locally optimal solutions.

The rest of this paper is organized as follows. We introduce the problem setting in Section 2 and provide a description of the algorithm in Section 3. The local convergence result is presented and its proof is outlined in Section 4. We conduct computational experiments in Section 5 to illustrate the algorithm and conclude in Section 6.

2 PROBLEM SETTING

We consider the optimization problem

$$\min_{x\in\mathbb{X}} H(x) = \mathbb{E}_{\phi}[h(x,\phi)], \tag{1}$$

where $\mathbb{X}$ is a continuous compact subset of $\mathbb{R}^d$, H is a deterministic, real-valued function, and ϕ is a random variable representing the stochastic uncertainty of the system. For a given solution x, we assume that the expected performance $H(x)$ cannot be evaluated exactly, but its noisy estimate $h(x,\phi)$ can be obtained through computer simulation. To simplify notation, we will sometimes suppress the dependency of h on the sample path ϕ and simply write $h(x)$ instead of $h(x,\phi)$.

3 ALGORITHM DESCRIPTION

We begin by introducing some notation. Let N_k be the number of candidate solutions sampled at the kth iteration of the algorithm and Λ_k be the set of sampled solutions. Let $\{r_k\}_{k\geq 1}$ be a sequence of deterministic positive real numbers. For a given $x\in\mathbb{X}$, $B(x,r)=\{y\in\mathbb{X}: d(x,y)<r\}$ denotes the ball with radius r and center x, where d is the Euclidean distance. At iteration k, the true performance $H(x)$ at x is estimated by $\widetilde{H}_k(x)$, the average of the objective function observations collected at all sampled candidate solutions in $B(x,r_k)\cap\Lambda_k$. Denote the number of elements in this set as $N_k(x,r_k)$ for every point $x\in\Lambda_k$ at the kth iteration. Let $S_k(x)$ and $P_k\subseteq\mathbb{X}$ be the surrogate model fitted and the promising region constructed at the kth iteration of the algorithm. The detailed algorithmic steps are given below.

Simulation Optimization via Promising Region Search and Surrogate Model Approximation (SOPS)

Step 0: Set the iteration counter $k=0$, and $P_0=\mathbb{X}$. Specify a small positive constant $\delta>0$ and a shrinking ball strategy $\{r_k\}_{k\geq 1}$.

Step 1: Let $k=k+1$. Uniformly sample a set of N_k candidate solution $\Lambda_k=\{x_{k_1},x_{k_2},\cdots,x_{k_{N_k}}\}$ from the current promising region P_{k-1}. Obtain sample performance at each $x\in\Lambda_k$ and use the shrinking ball method to construct performance estimates $\widetilde{H}_k(x)$ for all $x\in\Lambda_k$.

Step 2: Construct a surrogate model $S_k(x)$ that interpolates the objective function estimates $\widetilde{H}_k(x)$ at the set of sampled points Λ_k.

Step 3: Optimize the surrogate model $S_k(x)$ on P_{k-1} to get a local minimizer x_k^*. Construct a promising area P_k based on x_k^* as follows:

$$P_k=\left\{x\in\mathbb{X}:\ d\left(x,\ x_k^*\right)\leq d\left(x,\ y+2(y-x_k^*)\frac{\delta}{d(x_k^*,y)}\right),\ \forall\ y\in\Lambda_k\right\}.$$

Reiterate from Step 1 until a stopping condition is satisfied.

Note that at Step 1, any sampling measure F_k can be used to sample candidate solutions from the promising region P_{k-1}, provided that $F_k(A) > 0$ for any measurable set $A \subseteq P_{k-1}$ with positive Lebesgue measure. We have used a uniform distribution for simplicity. Step 2 requires the use of an interpolation-based fitting strategy, which ensures that $S_k(x) = \widetilde{H}_k(x)$ at all $x \in \Lambda_k$. The optimization at Step 3 can be carried out using any algorithm for deterministic optimization. The underlying assumption is that simulation experiments are much more expensive to run than evaluating the surrogate model, so that $S_k(x)$ can be optimized relatively efficiently without any additional simulation effort. The optimizer x_k^* of $S_k(x)$ is then used to construct a new promising region P_k of the solution space. This is conducted in a way that is very similar to the approach proposed in Hong and Nelson (2006), with the difference being that now there is an additional positive constant $\delta > 0$ involved to prevent P_k from degenerating into a single point when the set of sampled points becomes dense in the neighborhood of x_k^*.

It is easy to observe that in SOPS, if there is no surrogate model and the promising region is taken to be the entire feasible region in all iterations, then the algorithm is identical to the deterministic shrinking ball algorithm discussed in Andradóttir and Prudius (2010). On the other hand, if the solution space is (discrete) integer-ordered, then since each ball $B(x, r_k)$ will only contain x itself (when r_k is small enough), the shrinking ball strategy reduces to the usual sample average approximation. Thus, the algorithm (without the surrogate model) becomes the COMPASS algorithm of Hong and Nelson (2006). In this respect, SOPS can be viewed as the extension of COMPASS to the continuous domain.

4 LOCAL CONVERGENCE OF SOPS

In this section, we show that the sequence of surrogate model optimizers $\{x_k^*\}_{k\geq 1}$ converges to the set of locally optimal solutions of (1) with probability one (w.p.1). Throughout this paper, a sequence a_k is said to be $\Omega(k^n)$ if $\exists\ c > 0$ and $k_0 > 0,\ s.t.\ \forall\ k \geq k_0,\ a_k \geq ck^n$ and to be $\Theta(k^n)$ if $\exists\ c_1, c_2 > 0,\ \exists\ k_0 > 0,\ s.t.\ \forall\ k > k_0,\ c_1 k^n \leq a_k \leq c_2 k^n$. The following assumptions are needed in our analysis:

A1: The objective function $H(x)$ is Lipschitz continuous with Lipschitz constant L_1.

A2: $\exists\ l \geq 2,\ R \in \mathbb{R}^+, s.t.\ E[(h(x) - H(x))^{2l}] \leq R,\ \forall\ x \in \mathbb{X}$, where $h(x)$ is the sample performance at x.

A3: The surrogate model $S_k(x)$ satisfies $S_k(x) = \widetilde{H}_k(x),\ \forall x \in \Lambda_k$. Moreover, all $S_k(x)$ are twice differentiable with their Lipschitz constants uniformly bounded by L_2 for all k w.p.1.

A4: $\exists\ \delta_0 > 0$ such that $P(B(x_k^*, \delta_0) \nsubseteq int(P_{k-1})\ i.o.) = 0$, where $int(P_{k-1})$ is the interior of P_{k-1}.

A5: $\exists\ \kappa > 0,\ s.t.\ \inf_{x \in B(x_k^*, \delta_0)} \lambda_{min}(\mathbf{H}_{S_k}(x)) > \kappa$ for all k w.p.1, where $\mathbf{H}_{S_k}(x)$ is the Hessian matrix of S_k at x and $\lambda_{min}(\mathbf{H}_{S_k}(x))$ indicates its smallest eigenvalue.

A6: $N_k = \Theta(k^s)$ and $r_k = \Omega(k^{-p/d})$ with $\lim_{k\to\infty} r_k = 0$, where $1 - p \leq s < l - 1$ and $p < 1 - (s+1)/l$.

A1 and A2 are the respective assumptions on the true objective function and the simulation noise. A3 is the condition imposed on the fitting strategy. A4 requires the local minimizer x_k^* to lie in the interior of the promising area and A5 states that the Hessian matrix of the surrogate model remains strictly positive definite within a small neighborhood of the local minimizer. A6 controls the sample size N_k used at each iteration and the decreasing speed of the shrinking ball radius.

Let M be the set of all local minimizers of (1). Our main convergence result shows that the sequence of the δ_0-neighborhood of the surrogate model minimizers $B(x_k^*, \delta_0)$ will contain M infinitely often (i.o.) w.p.1. By construction, it is easy to observe that the δ-neighborhood of x_k^*, $B(x_k^*, \delta) \subseteq P_k$ for all k. The next result shows that the collection of sampled solutions will eventually become dense in P_{k-1}.

Lemma 1. *For any $\varepsilon > 0$ and $x_k \in P_{k-1}$, define $A_k(x_k, \varepsilon) = \{\exists\ x \in \Lambda_k,\ d(x_k, x) < \varepsilon\}$. If A1 and A6 hold, then $P(\bar{A}_k(x_k, \varepsilon)\ i.o.) = 0$.*

Proof. Define $\mathscr{F}_k = \sigma\{\Lambda_1, h(\Lambda_1), \Lambda_2, h(\Lambda_2), \cdots, \Lambda_k, h(\Lambda_k)\}$, $k = 1, 2, \cdots$ as the sequence of increasing σ-fields generated by the set of all sampled solutions and their performance estimates obtained up to

iteration k. Note that given $\mathscr{F}_k$, x_k^* is completely determined. Now we claim that for any $x_k \in P_{k-1}$, $\exists\ P_\varepsilon > 0,\ s.t.\ F_k(B(x_k,\varepsilon)) \geq P_\varepsilon$, where recall that F_k is the uniform sampling measure used at the kth iteration. To see this, we let $\eta = \min\{\frac{\delta}{2}, \frac{\varepsilon}{2}\}$ and $z = x_k + (x_{k-1}^* - x_k)\frac{\eta}{d(x_{k-1}^*, x_k)}$. Then for any $z' \in B(z,\eta)$, we have

$$d(x_{k-1}^*, z') \leq d(x_{k-1}^*, z) + d(z, z') \leq \max(\eta, \delta - \eta) + \eta \leq \delta \implies z' \in B(x_{k-1}^*, \delta).$$
$$d(x_k, z') \leq d(x_k, z) + d(z, z') \leq \eta + \eta \leq \varepsilon \implies z' \in B(x_k, \varepsilon).$$

Therefore, $B(z,\eta) \subseteq B(x_{k-1}^*, \delta) \cap B(x_k, \varepsilon)$, which implies that $B(z,\eta) \subseteq P_{k-1} \cap B(x_k,\varepsilon)$. Since F_k is uniform, we have $F_k(B(x_k,\varepsilon)) \geq F_k(B(z,\eta)) \geq P_\varepsilon > 0$, where $P_\varepsilon := \frac{Vol(B_\eta)}{Vol(\mathbb{X})}$ with $Vol(B_\eta)$ being the volume of any ball with radius η.

Consequently, we have $P\left(\bar{A}_k(x_k,\varepsilon)\right) = E\left[P(\bar{A}_k(x_k,\varepsilon)\middle|\mathscr{F}_{k-1})\right] \leq (1-P_\varepsilon)^{N_k}$ for each $k \geq 1$, and by A6, $\sum_{k=0}^{\infty} P\left(\bar{A}_k(x_k,\varepsilon)\right) \leq \sum_{k=0}^{\infty}(1-P_\varepsilon)^{N_k} < \infty$. By the Borel-Cantelli lemma, we obtain $P\left(\bar{A}_k(x_k,\varepsilon)\ i.o.\right) = 0$. □

Lemma 2 below states that on the promising area P_{k-1}, the true objective function $H(x)$ can be closely approximated by the surrogate model $S_k(x)$.

Lemma 2. *For any $\varepsilon > 0$ and $x_k \in P_{k-1}$, let $\Delta_k(x_k,\varepsilon) = \{|S_k(x_k) - H(x_k)| > \varepsilon\}$. If A1-A3 and A6 hold, then $P(\Delta_k(x_k,\varepsilon)\ i.o.) = 0$.*

Proof. Define $D_k = \{\forall\ x_k \in \Lambda_k,\ N_k(x_k,r_k) \geq L_k\}$, where $N_k(x_k,r_k)$ is the number of elements in the set $B(x_k,r_k) \cap \Lambda_k$, as defined before, and $L_k = \Theta(k^q)$, where q satisfies $p + q < 1$ and $ql - s > 1$. It is not hard to see when A6 is satisfied, this q exists. Since $r_k \to 0$ as $k \to \infty$, $N_k = \Theta(k^s)$, then $\exists\ k' \in \mathbb{N},\ c_1,\ c_2 \in \mathbb{R}^+,\ s.t.\ \forall\ k \geq k',\ r_k \leq \frac{\varepsilon}{4L_1},\ L_k \geq c_1 k^q,\ N_k \leq c_2 k^s$, and also let $\varepsilon' = \frac{\varepsilon}{2(L_1+L_2)}$. So for any $k > k',\ \forall\ x_k \in P_{k-1}$,

$$\begin{aligned}
&P\left(|S_k(x_k) - H(x_k)| > \varepsilon\ \middle|\mathscr{F}_{k-1}\right)\\
&\leq P\left(|S_k(x_k) - H(x_k)| > \varepsilon,\ A_k(x_k,\varepsilon')\ \middle|\mathscr{F}_{k-1}\right) + P\left(\bar{A}_k(x_k,\varepsilon')\ \middle|\mathscr{F}_{k-1}\right)\\
&= P\left(|S_k(x_k) - H(x_k)| > \varepsilon,\ \cup_{x_k' \in \Lambda_k} d(x_k,x_k') < \varepsilon'\ \middle|\mathscr{F}_{k-1}\right) + P\left(\bar{A}_k(x_k,\varepsilon')\ \middle|\mathscr{F}_{k-1}\right)\\
&\leq |\Lambda_k| P\left(|S_k(x_k) - H(x_k)| > \varepsilon,\ d(x_k,x_k') < \varepsilon'\ \middle|\mathscr{F}_{k-1}\right) + P\left(\bar{A}_k(x_k,\varepsilon')\ \middle|\mathscr{F}_{k-1}\right)\\
&\leq N_k P\left(\left|S_k(x_k) - S_k(x_k')\right| + \left|S_k(x_k') - H(x_k')\right| + \left|H(x_k') - H(x_k)\right| > \varepsilon,\ d(x_k,x_k') < \varepsilon'\ \middle|\mathscr{F}_{k-1}\right)\\
&\qquad + P\left(\bar{A}_k(x_k,\varepsilon')\ \middle|\mathscr{F}_{k-1}\right)\\
&\leq c_2 k^s P\left(\left|S_k(x_k') - H(x_k')\right| > \varepsilon/2\ \middle|\mathscr{F}_{k-1}\right) + P\left(\bar{A}_k(x_k,\varepsilon')\ \middle|\mathscr{F}_{k-1}\right)\\
&\leq c_2 k^s P\left(\left|S_k(x_k') - H(x_k')\right| > \varepsilon/2,\ D_k\ \middle|\mathscr{F}_{k-1}\right) + P\left(\bar{A}_k(x_k,\varepsilon')\ \middle|\mathscr{F}_{k-1}\right) + c_2 k^s P\left(\bar{D}_k\ \middle|\mathscr{F}_{k-1}\right)\\
&= c_2 k^s P\left(\left|\sum_{i=1}^{N_k(x_k',r_k)} \left(h(x_i) - H(x_k')\right)/N_k\left(x_k',r_k\right)\right| > \varepsilon/2,\ D_k\ \middle|\mathscr{F}_{k-1}\right)\\
&\qquad + P\left(\bar{A}_k(x_k,\varepsilon')\ \middle|\mathscr{F}_{k-1}\right) + c_2 k^s P\left(\bar{D}_k\ \middle|\mathscr{F}_{k-1}\right)\\
&\leq c_2 k^s P\left(\sum_{i=1}^{N_k(x_k',r_k)} \left(|h(x_i) - H(x_i)| + \left|H(x_i) - H(x_k')\right|\right)/N_k\left(x_k',r_k\right) > \varepsilon/2,\ D_k\right)\\
&\qquad + P\left(\bar{A}_k(x_k,\varepsilon')\right) + c_2 k^s P\left(\bar{D}_k\ \middle|\mathscr{F}_{k-1}\right)\\
&\leq c_2 k^s P\left(\left|\sum_{i=1}^{N_k(x_k',r_k)} \left(h(x_i) - H(x_i)\right)/N_k\left(x_k',r_k\right)\right| > \varepsilon/4,\ D_k\ \middle|\mathscr{F}_{k-1}\right)\\
&\qquad + P\left(\bar{A}_k(x_k,\varepsilon')\ \middle|\mathscr{F}_{k-1}\right) + c_2 k^s P\left(\bar{D}_k\ \middle|\mathscr{F}_{k-1}\right)
\end{aligned}$$

$$\leq \frac{C}{k^{ql-s}} + P\left(\bar{A}_k(x_k,\varepsilon') \,\middle|\, \mathscr{F}_{k-1}\right) + c_2 k^s P\left(\bar{D}_k \,\middle|\, \mathscr{F}_{k-1}\right), \tag{2}$$

where $x'_k \in \Lambda_k$ is a sampled point during iteration k that is at most ε' away from x_k, and $\{x_i\}_{i=1}^{N_k(x'_k,r_k)}$ are those points in $B(x'_k,r_k)\cap\Lambda_k$. The fourth inequality is true because of $d(x_k,x'_k)<\varepsilon'$ and the Lipschitz continuous property of the original function and the surrogate model. The seventh inequality holds since every x_i is within radius r_k of x'_k and $r_k \leq \frac{\varepsilon}{4L_1}$. The first term in the last inequality is a direct result of lemma 1 in Andradóttir and Prudius (2010). So (2) implies $P\left(|S_k(x_k)-H(x_k)|>\varepsilon\right) \leq \frac{C}{k^{ql-s}} + P\left(\bar{A}_k(x_k,\varepsilon')\right) + c_2 k^s P\left(\bar{D}_k\right)$ when $k>k'$.

As a result, we can obtain $\sum_{k=1}^{\infty} P\left(|S_k(x_k)-H(x_k)|>\varepsilon\right)<\infty$ by noticing that $ql-s>1$ and applying the result of lemma 1 and a slightly modification of lemma 2 in Andradóttir and Prudius (2010) with the simple size on the order of $\Theta(k^s)$ ($s \geq 1-p$). The result then follows from the Borel-Cantelli lemma. □

Let $cl\left(B(x_k^*,\delta_0)\right)$ be the closure of $B(x_k^*,\delta_0)$. Since $cl\left(B(x_k^*,\delta_0)\right)$ is compact, by using an argument similar to the proof of lemma 4.1 in Hu et al. (2012), it can be seen that the result in lemma 2 holds uniformly on $cl\left(B(x_k^*,\delta_0)\right)$.

Proposition 1. *For any $\varepsilon>0$, if A1-A4 and A6 hold, then $P\left(\max_{x\in cl\left(B(x_k^*,\delta_0)\right)} |S_k(x)-H(x)|>\varepsilon \ i.o.\right)=0$.*

The main result is shown in Theorem 1 below.

Theorem 1 If A1-A6 hold, then for any $0<\delta'\leq\delta_0$, $P\left(B(x_k^*,\delta')\cap M=\varnothing \ i.o.\right)=0$.

Proof. For any $0<\delta'\leq\delta_0$, let $x_k = argmin_{x\in cl(B(x_k^*,\delta'))}H(x)$. If $B(x_k^*,\delta')\cap M=\varnothing$, then x_k must lie on the boundary of $B(x_k^*,\delta')$ and $d(x_k,x_k^*)=\delta'$. Consider the Taylor expansion of $S_k(x)$ in the neighborhood of x_k^*,

$$S_k(x_k) = S_k(x_k^*) + \nabla^T S_k(x_k^*)(x_k-x_k^*) + \frac{1}{2}(x_k-x_k^*)^T \mathbf{H}_{S_k}(\bar{x})(x_k-x_k^*),$$

where $\bar{x}\in B(x_k^*,\delta')$. If $B(x_k^*,\delta')\subseteq int(P_{k-1})$, since $x_k^* = argmin_{x\in B(x_k^*,\delta')}S_k(x)$, we have $\nabla^T S_k(x_k^*)=0$, and by A5, $(x_k-x_k^*)^T\mathbf{H}_{S_k}(\bar{x})(x_k-x_k^*) \geq \lambda_{min}\left(\mathbf{H}_{S_k}(\bar{x})\right) d(x_k,x_k^*)^2 \geq \kappa\delta'^2$. This implies that $|S_k(x_k)-S_k(x_k^*)| \geq \kappa\delta'^2/2$. On the other hand,

$$\begin{aligned}
&|S_k(x_k)-S_k(x_k^*)| \leq |S_k(x_k)-H(x_k)| + |H(x_k)-S_k(x_k^*)| \\
&\leq \max_{x\in cl\left(B(x_k^*,\delta')\right)} |S_k(x)-H(x)| + |\min_{x\in cl\left(B(x_k^*,\delta')\right)} H(x) - \min_{x\in cl\left(B(x_k^*,\delta')\right)} S_k(x)| \\
&\leq 2\max_{x\in cl\left(B(x_k^*,\delta')\right)} |S_k(x)-H(x)|
\end{aligned}$$

Consequently, $B(x_k^*,\delta')\cap M=\varnothing$ and $B(x_k^*,\delta')\subseteq int(P_{k-1})$ implies $\max_{x\in cl\left(B(x_k^*,\delta')\right)} |S_k(x)-H(x)| \geq \kappa\delta'^2/4$. It follows that

$$P\Big(\{B(x_k^*,\delta')\cap M=\varnothing\}\cap\{B(x_k^*,\delta')\subseteq int(P_{k-1})\}\Big) \leq P\left(\max_{x\in cl\left(B(x_k^*,\delta')\right)} |S_k(x)-H(x)| \geq \kappa\delta'^2/4\right)$$

Hence, by directly applying Proposition 1, we obtain

$$\begin{aligned}
&P\Big(\{B(x_k^*,\delta')\cap M=\varnothing\}\cap\{B(x_k^*,\delta')\subseteq int(P_{k-1})\} \ i.o.\Big) \\
&\leq P\left(\max_{x\in cl\left(B(x_k^*,\delta')\right)} |S_k(x_k)-H(x_k)| \geq \kappa\delta'^2/4 \ i.o.\right) \\
&= 0,
\end{aligned}$$

If follows that if $B(x_k^*,\delta')\cap M=\varnothing$ occurs infinitely often, then $B(x_k^*,\delta')\nsubseteq int(P_{k-1})$ must also occur infinitely often. This in turn implies that $P\left(B(x_k^*,\delta')\cap M=\varnothing \ i.o.\right)=0$ by Assumption A4. □

5 NUMERICAL EXAMPLES

In this section, we illustrate the performance of SOPS on several continuous stochastic optimization problems. In our implementation, we have used a MCMC technique discussed in Smith (1984) to sample candidate solutions from the promising region at each iteration. It has been shown that this method asymptotically generates uniform points in a bounded region. The surrogate model is constructed using the radial basis function (RBF) approximation (Bishop 1995), which has been successfully used as a curve fitting tool in surrogate model-based optimization. The specific approximator considered here is a linear combination of RBFs of the following form: $S_k(x) = \sum_{i=1}^{N_k} w_i \phi(\|x - x_i\|)$, where $\phi(r) = r^3$, N_k is the total number of solutions sampled during iteration k, x_i's are the sampled solutions, and w_i's are the weights that can be obtained by solving a system of linear equations. In all testing cases, we set $\delta = 0.1$. Since the shrinking ball radius is on the order $\Omega(k^{-p/d})$ and the choice of the sample size N_k is on the order of $\Theta(k^s)$, where $1 - p \le s \le l - 1,\ p < 1 - (s+1)/l$ and d is the dimension of the problem, we simply choose the radius in the form of $\frac{a}{k^{0.5/d}}$ and the sample size to be $\max(\sqrt{k}, 4)$, with the value of a taken to be 5% of the length of the domain in each direction.

The following four test functions with additive noise are used in our experiments.

(1) Goldstein-Price function with additive noise

$$\begin{aligned} h_1(x,\phi_1) =& (1 + (x_1 + x_2 + 1)^2 (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)) \\ & (30 + (2x_1 - 3x_2)^2 (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)) + \phi_1 \end{aligned}$$

where $-3 \le x_i \le 3,\ i = 1,2$ and $\phi_1 \sim \mathcal{N}(0, 100)$. The function $H_1(x) = E_{\phi_1}[h_1(x,\phi_1)]$ has three local minima $(-0.6,-0.4), (1.8,0.2), (1.2,0.8)$ and a global minumum $x_1^* = (0,-1)$ with function value $H_1(x_1^*) = 3$.

(2) Schwefel function with additive noise (d=10)

$$h_2(x,\phi_2) = 201.8432n - \sum_{i=1}^{d} x_i sin(\sqrt{|x_i|}) + \phi_2$$

where $-200 \le x_i \le 250,\ i = 1,2,\cdots,d$ and $\phi_2 \sim \mathcal{N}(0, 100)$. The function $H_2(x) = E_{\phi_2}[h_2(x,\phi_2)]$ has global minimum at $x_2^* = (203.814, 203.814, \cdots, 203.814)$ with function value $H_2(x_2^*) = 0$.

(3) Rastrigin function with additive noise (d=10)

$$h_3(x,\phi_3) = 10d + \sum_{i=1}^{d} x_i^2 - 10cos(2\pi x_i) + \phi_3$$

where $-5.12 \le x_i \le 5.12,\ i = 1,2,\cdots,d$ and $\phi_3 \sim \mathcal{N}(0, 25)$. The function $H_3(x) = E_{\phi_3}[h_3(x,\phi_3)]$ has global minimum at $x_3^* = (0,0,\cdots,0)$ with function value $H_3(x_3^*) = 0$.

(4) Trigonometric function with additive noise (d=10)

$$h_4(x,\phi_4) = \sum_{i=1}^{d} \left[8\sin^2\left(7(x_i - 0.9)^2\right) + 6\sin^2\left(14(x_i - 0.9)^2\right) + (x_i - 0.9)^2 \right] + \phi_4$$

where $-2 \le x_i \le 3,\ i = 1,2,\cdots,d$ and $\phi_4 \sim \mathcal{N}(0, 25)$. The function $H_4(x) = E_{\phi_4}[h_4(x,\phi_4)]$ has global minimum at $x_4^* = (0.90009, 0.90009, \cdots, 0.90009)$ with function value $H_4(x_4^*) = 0$.

5.1 Numerical Results

We compare the performance of our algorithm with that of the the well-known simultaneous perturbation stochastic approximation (SPSA) (Spall 1992). To illustrate the effect of surrogate models on the algorithm

performance, we have also included in our comparison a version of the SOPS algorithm called SOP. SOP has the same structure as SOPS but without the surrogate model approximation at each iteration. The choices of the parameters for these algorithms are listed in Table 1 and Table 2, where k is the number of iterations, r_k is the shrinking ball radius, δ is used in constructing the promising area as defined in Step 3 of SOPS, and α_k, c_k are the respective step size and the simultaneous perturbation size used in SPSA, which are chosen by trial and error.

Table 1: The Choices of r_k and δ in SOPS and SOP.

	r_k	δ
$h_1(d=2)$	$\frac{0.2}{k^{0.5/2}}$	0.1
$h_2(d=10)$	$\frac{20}{k^{0.5/10}}$	0.1
$h_3(d=10)$	$\frac{0.4}{k^{0.5/10}}$	0.1
$h_4(d=10)$	$\frac{1}{k^{0.5/10}}$	0.1

Table 2: The Choices of α_k and c_k in SPSA.

	α_k	c_k
$h_1(d=2)$	$\frac{1}{(k+5000)^1}$	$\frac{1}{(k+50000)^{0.25}}$
$h_2(d=10)$	$\frac{10}{(k+1)^1}$	$\frac{1}{(k+50000)^{0.25}}$
$h_3(d=10)$	$\frac{1}{(k+1000)^1}$	$\frac{1}{(k+50000)^{0.25}}$
$h_4(d=10)$	$\frac{1}{(k+1000)^1}$	$\frac{1}{(k+50000)^{0.25}}$

Figure 1 shows the performance of the three comparison algorithms, averaged over 30 independent replication runs on each test case. The use of promising region search in SOPS and SOP allows the algorithms to quickly locate promising subsets of the solution space. We see that both algorithms show faster initial improvements than SPSA in all cases. Moreover, as compared with SOP, SOPS uses a surrogate model to approximate the response surface of the objective function. This may provide a better and accurate prediction of promising regions, leading to improved algorithm performance. We see that in almost all cases, SOPS finds better solutions than SOP does.

6 CONCLUSION

We have proposed a random search-based optimization algorithm for solving SO problems with continuous decision variables. The algorithm combines ideas of promising area search and surrogate model approximation, and estimates the objective function values at sampled points using the shrinking ball method. We have discussed the local convergence property of the algorithm and carried out preliminary simulation experiments to illustrate its performance.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under Grant CMMI-1634627.

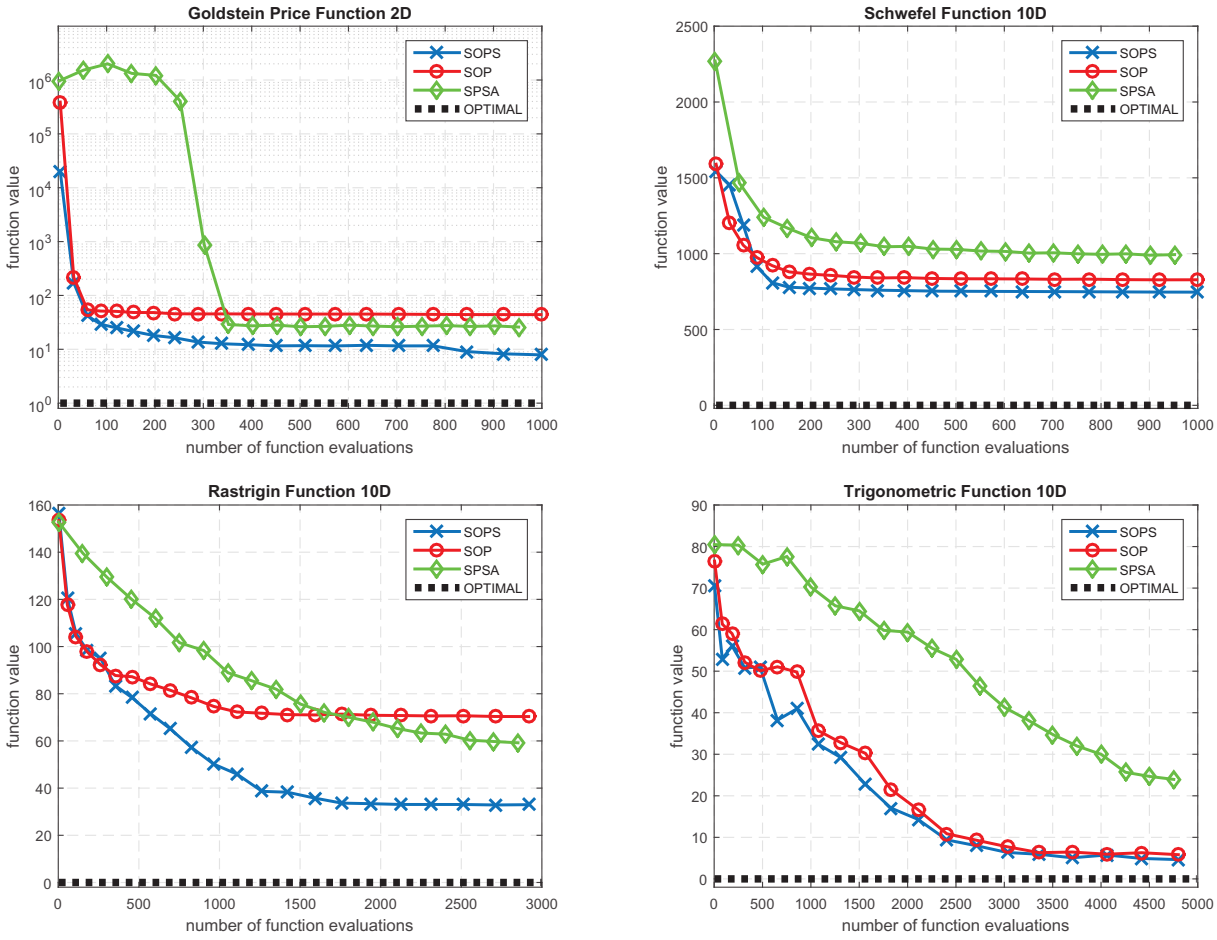

Figure 1: Averaged Performance of the Test Functions.

REFERENCES

Alrefaei, M. H., and S. Andradóttir. 1999. "A Simulated Annealing Algorithm with Constant Temperature for Discrete Stochastic Optimization". *Management Science* 45:748–764.

Andradóttir, S. 2014. "A Review of Random Search Methods". In *Handbook of Simulation Optimization*, edited by M. C. Fu, International Series in Operations Research and Management Science, 277–292. Springer.

Andradóttir, S., and A. A. Prudius. 2010. "Adaptive Random Search for Continuous Simulation Optimization". *Naval Research Logistics* 57:583–604.

Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*. New York, NY, USA: Oxford University Press, Inc.

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* 25:230–243.

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

Hong, L. J., and B. L. Nelson. 2006. "Discrete Optimization via Simulation Using COMPASS". *Operations Research* 54:115–129.

Hu, J. 2014. "Model-Based Stochastic Search Methods". In *Handbook of Simulation Optimization*, edited by M. C. Fu, International Series in Operations Research and Management Science, 319–340. Springer.

Hu, J., M. C. Fu, and S. I. Marcus. 2008. "A Model Reference Adaptive Search Method for Stochastic Global Optimization". *Communications in Information and Systems* 8:245–276.

Hu, J., P. Hu, and H. S. Chang. 2012. "A Stochastic Approximation Framework for a Class of Randomized Optimization Algorithms". *IEEE Transactions on Automatic Control* 57:165–178.

Huang, D., T. Allen, W. Notz, and N. Zeng. 2006. "Global Optimization of Stochastic Black-box Systems via Sequential Kriging Meta-models". *Journal of Global Optimization* 34:441–466.

Jones, D., M. Schonlau, and W. Welch. 1998. "Efficient Global Optimization of Expensive Black-box Functions". *Journal of Global Optimization* 13:455–492.

Kleijnen, J. P. 2014. "Response Surface Methodology". In *Handbook of Simulation Optimization*, edited by M. C. Fu, International Series in Operations Research and Management Science, 81–104. Springer.

Nakayama, H., M. Arakawa, and R. Sasaki. 2002. "Simulation-based Optimization Using Computational Intelligence". *Optimization and Engineering* 3:201–214.

Regis, R., and C. Shoemaker. 2007. "Improved Strategies For Radial Basis Function Methods For Global Optimization". *Journal of Global Optimization* 37:113–135.

Rubinstein, R. Y., and D. P. Kroese. 2004. *The Cross Entropy Method: A Unified Approach To Combinatorial Optimization, Monte-carlo Simulation (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Shi, L., and S. Ólafsson. 2000. "Nested Partitions Method for Stochastic Optimization". *Methodology And Computing In Applied Probability* 2:271–291.

Smith, R. L. 1984. "Efficient Monte Carlo Procedures for Generating Points Uniformly Distributed Over Bounded Regions". *Operations Research* 32:1296–1308.

Sóbester, A., S.J.Leary, and A. Keane. 2005. "On the Design of Optimization Strategies Based On Global Response Surface Approximation Model". *Journal of Global Optimization* 33:31–59.

Spall, J. C. 1992. "Multivariate Stochastic Approximation Using a Simultaneous Perturbation Gradient Approximation". *IEEE Transactions on Automatic Control* 37:332–341.

Xu, J., B. L. Nelson, and J. L. Hong. 2010. "Industrial Strength COMPASS: A Comprehensive Algorithm and Software for Optimization via Simulation". *ACM Transactions on Modeling and Computer Simulation* 20:3:1–3:29.

Yan, D., and H. Mukai. 1992. "Stochastic Discrete Optimization". *SIAM Journal on Control and Optimization* 30:594–612.

Zabinsky, Z. B. 2014. "Stochastic Adaptive Search Methods: Theory and Implementation". In *Handbook of Simulation Optimization*, edited by M. C. Fu, International Series in Operations Research and Management Science, 293–318. Springer.

AUTHOR BIOGRAPHIES

QI FAN is a Ph.D. student in the Department of Applied Mathematics and Statistic at the State University of New York, Stony Brook. He received the B.S. degree in mathematics from Zhejiang University, China in 2011. His research interests include Markov decision processes, optimization and simulation. His email address is qfan@ams.sunysb.edu.

JIAQIAO HU is an Associate Professor in the Department of Applied Mathematics and Statistics at the State University of New York, Stony Brook. He received the B.S. degree in automation from Shanghai Jiao Tong University, the M.S. degree in applied mathematics from the University of Maryland, Baltimore County, and the Ph.D. degree in electrical engineering from the University of Maryland, College Park. His research interests include Markov decision processes, applied probability, and simulation optimization. His email address is jqhu@ams.sunysb.edu.