Enhancing Random Search with Surrogate Models for Lipschitz Continuous Optimization

Qi Zhang and Jiaqiao Hu, *Member, IEEE*

***Abstract*— We propose a random search algorithm for solving Lipschitz continuous optimization problems. The algorithm samples candidate from a parameterized probability distribution over the solution space and uses the previously sampled data to fit a surrogate model of the objective function. The surrogate model is then used to modify the parameterized distribution in a way that concentrates the search on the set of high quality solutions. We prove the global convergence of the algorithm and provide numerical examples to illustrate its performance.**

I. INTRODUCTION

We address optimization problems of the form

$$x^* \in \arg\max_{x \in \mathscr{X}} H(x), \tag{1}$$

where the feasible region $\mathscr{X}$ is a compact subset of $\mathscr{R}^d$, and the objective function $H : \mathscr{X} \to \mathscr{R}$ is Lipschitz continuous. We consider a black-box scenario where the explicit form of H is not available and no other functional properties, such as convexity or differentiability, are known a priori. However, for a given solution in the domain $\mathscr{X}$, we assume that the objective function can be evaluated exactly without error.

A popular strategy that has been proven effective for attacking such problems is to use random search, where the basic idea is to construct a sequence of random iterates (e.g., candidate solutions, promising subsets, probability distributions), and then use the sequence to successfully approximate the optimal solution. Some examples of random search algorithms include simulated annealing [11], genetic algorithms [6], the nested partitions [15], tabu search [5], and model-based methods [8], [14]. Another general approach, which is frequently adopted when function evaluations are computationally expensive, is to use a surrogate model to approximate the response curve of the unknown objective function. This has given rise to a variety of optimization techniques based on surrogate modeling or response surface methodologies (RSMs); see, e.g., [1], [2], [7], [10], and [12].

In this paper, we propose an algorithm called enhanced annealing random search (EARS) that integrates surrogate modeling techniques within the class of model-based random search methods for solving (1). The algorithm is an improved version of the model-based annealing random search (MARS) method developed in [9], which searches for an optimal solution by repeatedly sampling from a sequence of parameterized distributions that gradually shift their probability mass to the set of high quality solutions. However, a potential drawback of MARS is that the the distribution is updated at each step based only on the currently generated solutions, which results in the previously sampled data being either discarded or insufficiently utilized. EARS addresses this limitation by fully retaining the previous sampling data and using them to build a surrogate model of the objective function. Thus, a main difference compared to MARS is that in EARS, the parameterized distribution is constructed based on an approximation of the entire objective function as opposed to be based on the set of sampled solutions. Intuitively, a well-built surrogate model can be used to effectively predict the response values at un-sampled solutions, and thus allows the algorithm to improve future search by exploiting knowledge of the objective function. In addition, as will be discussed in Section II, another benefit of using surrogate modeling is that it provides a way to resolve a certain ratio bias issue in the original MARS algorithm, which allows us to show the global convergence of EARS when only a single candidate solution is sampled at each iteration. This is in contrast to the original MARS algorithm, which requires the number of sampled solutions to increase polynomially with the number of algorithm iterations.

The rest of this paper is organized as follows. In Sections II and III, we motivate our work and describe the EARS algorithm. The global convergence result of EARS under general conditions is presented in Section IV. Illustrative numerical examples and some preliminary comparison results are provided in Section V. Finally, we conclude the paper in Section VI.

II. BACKGROUND AND MOTIVATION

The general idea of the MARS algorithm for solving (1) is to construct a sequence of parameterized sampling distributions $\{f_{\theta_k}\}$ that approximates Boltzmann distributions of the form

$$g_k(x) := \frac{e^{H(x)/t_k}}{\int_{\mathscr{X}} e^{H(x)/t_k} dx}, \tag{2}$$

where $t_k > 0$ is a temperature parameter. It is well known that as t_k tends to zero, g_k will converge to a degenerate distribution that concentrates only on the optimal solutions. Thus, if f_{θ_k} is a good approximation to g_k, then candidate solutions sampled from f_{θ_k} will be close to x^* with high probabilities as k becomes large. The approximation is carried out by choosing the parameterized distribution f_θ from the so-called natural exponential family (NEF) (see

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

Q. Zhang and J. Hu are with the Department of Applied Mathematics and Statistics, Stony Brook University, NY 11794, USA (e-mail: zhangqi.math@gmail.com; jqhu@ams.sunysb.edu).

Manuscript 110 submitted to 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE). Received February 27, 2019.

Definition 2.1) and then minimizing its Kullback-Leibler (KL) divergence to g_k, i.e.,

$$\theta_{k+1} = \arg\min_{\theta\in\Theta} \left\{ \mathscr{D}(g_k, f_\theta) := E_{g_k}\left[\ln \frac{g_k(X)}{f_\theta(X)}\right]\right\}. \quad (3)$$

Definition 2.1: A parameterized family of density/mass functions $\{f_\theta : \theta \in \Theta \subseteq \mathscr{R}^k\}$ is called a natural exponential family if there exist mappings $\Gamma(\cdot) : \mathscr{R}^d \to \mathscr{R}^k$ and $K(\cdot) : \mathscr{R}^k \to \mathscr{R}$ such that $f_\theta(x) = \exp(\theta^T\Gamma(x) - K(\theta))$, where $K(\theta) := \ln \int_{\mathscr{X}} \exp(\theta^T\Gamma(x))dx$ is a normalization constant and the superscript T denotes vector transposition.

In MARS, a convex combination of f_{θ_k} and g_k, i.e., $\tilde{g}_k(x) := \alpha_k g_k(x) + (1-\alpha_k)f_{\theta_k}(x)$, $\alpha_k \in (0,1)$, is used in lieu of g_k when minimizing the KL divergence (3). This leads to the following recursion connecting the parameter vectors θ_k and θ_{k+1} obtained at successively iterations of the algorithm [9]:

$$\begin{aligned} m(\theta_{k+1}) &= m(\theta_k) - \alpha_k \nabla_\theta \mathscr{D}(g_k, f_\theta)|_{\theta=\theta_k} \\ &= m(\theta_k) + \alpha_k(E_{g_k}[\Gamma(X)] - m(\theta_k)), \end{aligned} \quad (4)$$

where $m(\theta) := E_{f_\theta}[\Gamma(X)]$ is called the mean parameter function, which is a one-to-one, invertible transformation of θ. Since the Boltzmann density g_k depends on the unknown objective function H, the term $E_{g_k}[\Gamma(X)]$ in (4) is estimated using a change of measure under which the density g_k is replaced with the sampling distribution f_{θ_k}, and then approximating the expectations by their corresponding sample averages, i.e.,

$$\begin{aligned} E_{g_k}[\Gamma(X)] &= \frac{\int_{\mathscr{X}} e^{H(x)/t_k}\Gamma(x)dx}{\int_{\mathscr{X}} e^{H(x)/t_k}dx} \\ &= \frac{E_{f_{\theta_k}}[e^{H(X)/t_k}\Gamma(X)f_{\theta_k}^{-1}(X)]}{E_{f_{\theta_k}}[e^{H(X)/t_k}f_{\theta_k}^{-1}(X)]} \\ &\approx \frac{\frac{1}{N_k}\sum_{i=1}^{N_k} e^{H(x_i)/t_k}\Gamma(x_i)f_{\theta_k}^{-1}(x_i)}{\frac{1}{N_k}\sum_{i=1}^{N_k} e^{H(x_i)/t_k}f_{\theta_k}^{-1}(x_i)}, \end{aligned} \quad (5)$$

where $\{x_i : i = 1, ..., N_k\}$ is the set of candidate solutions independently sampled from f_{θ_k}.

Although the estimator given by (5) is straightforward to implement, its construction only relies on solutions sampled in the current step and thus involves a ratio bias for any finite sample size N_k. Therefore, the analysis of MARS requires N_k to increase polynomially with k to obtain an asymptotically unbiased estimator. In practice, this may result in a computational burden that becomes prohibitive as search proceeds, especially when the cost of function evaluation is high.

To address this limitation of MARS, we propose an enhanced version of the algorithm by resorting to the use of surrogate modeling techniques. The idea is to construct a surrogate model $S_k(x)$ of $H(x)$ based on all solutions sampled up to iteration k, and then estimate the expectation $E_{g_k}[\Gamma(X)]$ by replacing the true objective function $H(x)$ with its approximation $S_k(x)$, i.e.,

$$E_{g_k}[\Gamma(X)] \approx E_{\hat{g}_k}[\Gamma(X)],$$

where

$$\hat{g}_k(x) := \frac{e^{S_k(x)/t_k}}{\int_{\mathscr{X}} e^{S_k(x)/t_k}dx}. \quad (6)$$

Since the surrogate model often has an analytical expression, S_k is typically much cheaper to evaluate than the objective function H. Consequently, we assume that $E_{\hat{g}_k}[\Gamma(X)]$ can be computed arbitrarily accurately, e.g., via numerical integration or quasi-Monte Carlo methods.

III. ALGORITHM DESCRIPTION

We begin by specifying the following quantities:

- an exponential family of density functions $\{f_\theta(x), x \in \mathscr{X} : \theta \in \Theta\}$.
- an exploration factor λ, a step-size sequence $\{\alpha_k\}$, and an annealing schedule $\{t_k\}$.

The detailed algorithmic steps of EARS are then given below.

0: Choose an initial parameter θ_1. Initialize the set of sampled solutions $\Lambda_0 = \emptyset$ and set the iteration counter $k = 1$.

1: Sample a solution x_k from $(1-\lambda)f_{\theta_k} + \lambda U$, where U denotes the uniform distribution on $\mathscr{X}$. Evaluate the performance of x_k and obtain $y_k = H(x_k)$. Set $\Lambda_k = \Lambda_{k-1} \cup \{x_k\}$.

2: Construct a surrogate model $S_k(x)$ on $\mathscr{X}$ that interpolates the data $\{(x_i, y_i) : x_i \in \Lambda_k\}$.

3: Update the sampling distribution parameter

$$\begin{aligned} \eta_{k+1} &= \eta_k + \alpha_k(E_{\hat{g}_k}[\Gamma(X)] - \eta_k) \\ \theta_{k+1} &= m^{-1}(\eta_{k+1}). \end{aligned} \quad (7)$$

4: If a stopping criterion is met, then terminate the algorithm; Otherwise, set $k = k + 1$ and go to Step 1.

Intuitively, the use of the uniform distribution at Step 1 allows the algorithm to explore the entire solution space. This ensures that each subset of $\mathscr{X}$ with a positive volume will have a strictly positive probability of being visited by the algorithm. Equation (7) at Step 3 is essentially identical to the updating equation (4) with the only difference being that the idealized Boltzmann model g_k is now replaced by its approximation $\hat{g}_k$ (see (6)). Note that since Step 2 requires the use of an interpolation-based surrogate modeling technique, all previous sampling data are explicitly retained and directly contribute to the construction of $E_{\hat{g}_k}[\Gamma(X)]$. Thus, it is reasonable to expect that the estimator will provide an increasingly accurate approximation to $E_{g_k}[\Gamma(X)]$ even when a single solution is sampled at each iteration. Moreover, if the surrogate model S_k is able to accurately represent the true objective function H as data accumulates, then the distribution parameters calculated in (7) based upon the model (as opposed to be based on the randomly selected solutions as in (5)) could exhibit significantly reduced variability over time, leading to stable and robust algorithm performance.

Manuscript 110 submitted to 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE). Received February 27, 2019.

IV. GLOBAL CONVERGENCE

In this section, we analyze the asymptotic behavior of EARS and show its convergence to the global optimal solution of (1). Since the algorithm is randomized, it induces a probability distribution over the set of all sample paths generated. We denote by $P(\cdot)$ and $E[\cdot]$ the probability and expectation taken with respect to this distribution. Probability one convergence (w.p.1) of a sequence of random events is to be understood with respect to P. Throughout our analysis, a sequence $\{a_k\}$ is said to be $\mathcal{O}(b_k)$ if there exists a positive real number $c > 0$ and a positive integer $N > 0$ such that $a_k \leq cb_k$ for all $k \geq N$. We make the following assumptions regarding the problem and the algorithm parameters.

A1: The optimal solution x^* is unique. The objective function H is Lipschitz continuous with a Lipschitz constant L_1.

A2: The sufficient statistic $\Gamma(x)$ is bounded, i.e., $\exists B > 0$ such that $||\Gamma(x)|| \leq B$ for all $x \in \mathscr{X}$.

A3: The annealing schedule $\{t_k\}$ satisfies $t_k \to 0$ as $k \to \infty$ and $(\frac{\ln k}{k})^{\frac{1}{d}}/t_k \to 0$ as $k \to \infty$.

A4: The surrogate model satisfies $S_k(x) = H(x)$ for all $x \in \Lambda_k$. Moreover, all S_k's are Lipschitz continuous with their Lipschitz constants uniformly bounded by L_2 w.p.1.

A5: The step-size sequence $\{\alpha_k\}$ satisfies $\alpha_k \to 0$ as $k \to \infty$ and $\sum_{k=1}^{\infty} \alpha_k = \infty$.

Note that A1 and A2 are the respective assumptions on the objective function and the parameterized sampling distribution family. A3 states that the temperature t_k should gradually decrease to zero but at a rate that is sufficiently slow. A4 requires the use of interpolation-based surrogate modeling techniques with the fitted models themselves being Lipschitz continuous. A5 is a well-known condition used in studying stochastic approximation type of algorithms.

The following result states that for the class of optimization problems characterized by condition A1, the sequence of Boltzmann distributions $\{g_k\}$ given by (2) will converge to a limiting distribution that concentrates only on the optimal solution x^*. The proof can be found in Lemma 3.1 of [9].

Proposition 1: If Assumptions A1, A2 and A3 hold, then we have

$$E_{g_k}[\Gamma(X)] \to \Gamma(x^*) \ as \ k \to \infty, \tag{8}$$

where the limit is taken component-wise.

For a given point $x \in \mathscr{X}$ and a constant $r > 0$, let $B(x, r)$ denote a ball centered at x with radius r. Let $u_1, \ldots, u_k$ be k independent random points uniformly generated from a compact set $\mathscr{X}$. The following result, adapted from [4], is well-known in multidimensional spacing and provides a strong bound on the minimum radius r required for the collection of balls $\{B(u_i, r),\ i = 1, \ldots, k\}$ to cover $\mathscr{X}$.

Lemma 1: Let $u_1, u_2, ..., u_k$ be k i.i.d. random points uniformly sampled from $\mathscr{X}$ and define $r_k = \min\{r > 0 : \mathscr{X} \subseteq \cup_{i=1}^{k} B(u_i, r)\}$. Then $r_k = \mathcal{O}\big((\frac{\ln k}{k})^{\frac{1}{d}}\big)$ as $k \to \infty$ w.p.1.

Lemma 1, together with the Lipschitz continuity of H and S_k, implies the uniform convergence of the sequence of surrogate models $\{S_k\}$ to the objective function H.

Lemma 2: If Assumptions A1, A3 and A4 hold, then

$$\max_{x \in \mathscr{X}} |S_k(x) - H(x)|/t_k \to 0 \ w.p.1 \ as \ k \to \infty. \tag{9}$$

Proof: Let U_k denote the set of sample solutions generated from the uniform distribution up to iteration k. For each $x \in \mathscr{X}$, we have

$$\begin{aligned} |S_k(x) - H(x)| \leq & |S_k(x) - S_k(z_k)| + |S_k(z_k) - H(z_k)| \\ & + |H(z_k) - H(x)| \\ \leq & (L_1 + L_2)||x - z_k|| \\ \leq & (L_1 + L_2)||x - u_k||, \end{aligned}$$

where $z_k := \arg\min_{x_l \in \Lambda_k} ||x - x_l||$ and $u_k := \arg\min_{x_l \in U_k} ||x - x_l||$. Let $\Omega_1 = \{\omega \in \Omega : |U_k(\omega)| \geq \tilde{\lambda} k\}$ for some $0 < \tilde{\lambda} < \lambda$, where $|A|$ denotes the cardinality of the set A and Ω is the set of all sample path generated by the algorithm. Let $\Omega_2 = \{\omega \in \Omega : r_k(\omega) = \mathcal{O}\big((\frac{\ln |U_k(\omega)|}{|U_k(\omega)|})^{\frac{1}{d}}\big)\ i.o.\}$, where $r_k := \min\{r > 0 : \mathscr{X} \subseteq \cup_{x_l \in U_k} B(x_l, r)\}$ and $i.o.$ means infinitely often. Thus, on every $\omega \in \Omega_1 \cap \Omega_2$, there exists a constant $c > 0$ such that when k is sufficiently large, we have

$$\begin{aligned} \frac{|S_k(x) - H(x)|}{t_k} & \leq \frac{L_1 + L_2}{t_k} c\Big(\frac{\ln |U_k(\omega)|}{|U_k(\omega)|}\Big)^{\frac{1}{d}} \\ & \leq \frac{L_1 + L_2}{t_k} c\Big(\frac{\ln \tilde{\lambda} k}{\tilde{\lambda} k}\Big)^{\frac{1}{d}} \\ & = (L_1 + L_2) \frac{\mathcal{O}\Big((\frac{\ln k}{k})^{\frac{1}{d}}\Big)}{t_k}, \end{aligned}$$

which tends to zero as $k \to \infty$. Therefore, the result follows from the strong law of large number and Lemma 1, which imply $P(\Omega_1) = 1$ and $P(\Omega_2) = 1$, respectively. ■

The above lemma gives rise to the next result, showing that the error of the estimator $E_{\hat{g}_k}[\Gamma(X)]$ decreases to zero as the number of iterations increases.

Proposition 2: If Assumptions A1, A2, A3 and A4 hold, then we have

$$\big|\big|E_{\hat{g}_k}[\Gamma(X)] - E_{g_k}[\Gamma(X)]\big|\big| \to 0 \ w.p.1 \ as \ k \to \infty. \tag{10}$$

Proof: Note that

$$\begin{aligned} & ||E_{\hat{g}_k}[\Gamma(X)] - E_{g_k}[\Gamma(X)]|| \\ & \quad \leq \int_{\mathscr{X}} ||\Gamma(x)|| \big|\hat{g}_k(x) - g_k(x)\big| \ dx \\ & \quad \leq B \int_{\mathscr{X}} \hat{g}_k(x) \big|1 - \frac{g_k(x)}{\hat{g}_k(x)}\big| \ dx \\ & \quad \leq B\Big(\exp(2 \max_{x \in \mathscr{X}} \frac{|S_k(x) - H(x)|}{t_k}) - 1\Big), \end{aligned}$$

where the second inequality follows from A2 and the last inequality is due to the fact that

$$\begin{aligned} \frac{g_k(x)}{\hat{g}_k(x)} & = E_{g_k}[e^{\frac{S_k(X) - H(X)}{t_k}}] e^{\frac{H(x) - S_k(x)}{t_k}} \\ & \leq \exp\Big(2 \max_{x \in \mathscr{X}} \frac{|S_k(x) - H(x)|}{t_k}\Big). \end{aligned}$$

Manuscript 110 submitted to 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE). Received February 27, 2019.

Thus, Lemma 2 implies that $||E_{\hat{g}_k}[\Gamma(X)] - E_{g_k}[\Gamma(X)]|| \to 0$ w.p.1. ■

We now present the main convergence result, which states that the value of the mean parameter function η_k converges to $\Gamma(x^*)$ with probability one.

Theorem 1: If Assumptions A1 to A5 hold, and the surrogate sampling distributions are taken from the NEF, then $\{\eta_k\}$ generated by the EARS algorithm satisfies

$$\eta_k \to \Gamma(x^*) \ w.p.1 \ as \ k \to \infty, \tag{11}$$

where the limit is taken component-wise.

Proof: Subtract $\Gamma(x^*)$ on both sides of equation (7), and let $Y_k = \eta_k - \Gamma(x^*)$, we have

$$Y_{k+1} = Y_k + \alpha_k f(Y_k) + \alpha_k b_k,$$

where $f(Y) := -Y$ and $b_k := E_{\hat{g}_k}[\Gamma(X)] - \Gamma(x^*)$. Note that since $||b_k|| \leq ||E_{g_k}[\Gamma(X)] - \Gamma(x^*)|| + ||E_{\hat{g}_k}[\Gamma(X)] - E_{g_k}[\Gamma(X)]||$, propositions 1 and 2 imply that $||b_k|| \to 0$ as $k \to \infty$ w.p.1. Thus, a direct application of Example 2 in Section 2.1 of [13] yields $Y_k \to 0$ w.p.1. ■

V. NUMERICAL EXPERIMENTS

In this section, we illustrate the performance of the EARS algorithm by considering some preliminary computational experiments on four benchmark problems and comparing its performance with that of the MARS algorithm and the simultaneous perturbation stochastic approximation (SPSA) algorithm proposed in [16]. The four test functions are 10 dimensional ($d = 10$) and their solution spaces are all in the form of box constraints $\mathscr{X} = \{x \in \mathscr{R}^d : -10 \leq x_i \leq 10, \ i = 1, ..., d\}$.

1) Sum squares function

$$H_1(x) = -\sum_{i=1}^{d} i(x_i)^2,$$

where $x^* = (0, ..., 0)^T$, $H_1(x^*) = 0$.

2) Griewank function

$$H_2(x) = -\frac{1}{40}\sum_{i=1}^{d}(x_i)^2 + \prod_{i=1}^{d} cos\Big(\frac{x_i}{\sqrt{i}}\Big) - 1,$$

where $x^* = (0, ..., 0)^T$, $H_2(x^*) = 0$.

3) Ackley function

$$\begin{aligned} H_3(x) =& 20\exp\left(-\frac{1}{5}\sqrt{\frac{1}{d}\sum_{i=1}^{d}x_i^2}\right) \\ &+ \exp\left(\frac{1}{d}\sum_{i=1}^{d}\cos(2\pi x_i)\right) - 20 - \exp(1), \end{aligned}$$

where $x^* = (0, ..., 0)^T$, $H_3(x^*) = 0$.

4) Trigonometric function

$$\begin{aligned} H_4(x) = -\sum_{i=1}^{d}[&8\sin^2(7(x_i - 0.9)^2) \\ &+ 6\sin^2(14(x_i - 0.9)^2) + (x_i - 0.9)^2], \end{aligned}$$

where $x^* = (0.9, ..., 0.9)^T$, $H_4(x^*) = 0$.

In our implementation, we have used radial basis function approximation [3] to construct the surrogate models. The specific approximator considered here is a linear combination of cubic basis functions, i.e., $S_k(x) := \sum_{i=1}^{k} w_i \phi(||x - x_i||)$, where $\phi(r) := r^3$, x_i's are sampled solutions, and w_i's are weights that can be obtained by solving a system of linear equations based on sampled information. In order to achieve a robust performance of the algorithm, we begin by uniformly sampling 50 solutions from $\mathscr{X}$ and using their function values to construct an initial surrogate model. The parameterized sampling distributions are taken to be normal densities with independent components. The initial mean vector is uniformly generated from $\mathscr{X}$ and the initial covariance matrix is set to a $d \times d$ diagonal matrix with all diagonal entries equal to 100. In all test cases, we set the annealing schedule, the step-size sequence, and the exploration factor to $t_k = \frac{1}{\log(k+1)}$, $\alpha_k = \frac{1}{(k+20)^{0.51}}$ and $\lambda = 0.1$, respectively. At each iteration, the value of $E_{\hat{g}_k}[\Gamma(X)]$ is calculated by quasi-Monte Carlo integration using 3^{10} Sobol points generated from $\mathscr{X}$.

The parameters of MARS and SPSA are tuned separately for each test function to allow good performance of these algorithms. For MARS, the number of candidate solutions generated per iteration is taken to be $N_k = \lfloor K^{0.502} \rfloor$ and the step size is set to $\alpha_k = \frac{1}{(k+20)^{0.51}}$. An adaptive annealing schedule $t_k = \frac{0.1|h_k^*|}{\log(k+1)}$ is used in the algorithm, where h_k^* indicates the objective function value at the current best solution. For SPSA, the parameter values for each of the respective test cases are listed in Table I, where α_k is the gain sequence and c_k is the simultaneous perturbation size.

Each algorithm is run independently 50 times, and the averaged results are presented in Tables II, where H^* denotes the averaged function value at the final solution obtained by an algorithm and std_err is the standard error. Fig. 1 also shows the performance of the three comparison algorithms, where we plot the function values at the current best sampled solutions against the number of function evaluations used. We see that EARS shows superior performance over the two competing algorithms and finds solutions that are very close to the optimal ones in less than 200 function evaluations on all test functions. Note that due to the warm-up period used in EARS to construct the initial surrogate model, the algorithm is essentially conducting pure random search during the first 50 iterations. Therefore, EARS shows a slower initial improvement than MARS and SPSA. However, once an

TABLE I
CHOICES OF PARAMETERS IN SPSA.

	α_k	c_k
H_1	$\frac{1}{k+100}$	$\frac{1}{(k+1)^{0.25}}$
H_2	$\frac{50}{k+100}$	$\frac{5}{(k+1)^{0.25}}$
H_3	$\frac{10}{k+10}$	$\frac{2}{(k+1)^{0.25}}$
H_4	$\frac{1}{k+10}$	$\frac{5}{(k+1)^{0.25}}$

Manuscript 110 submitted to 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE). Received February 27, 2019.

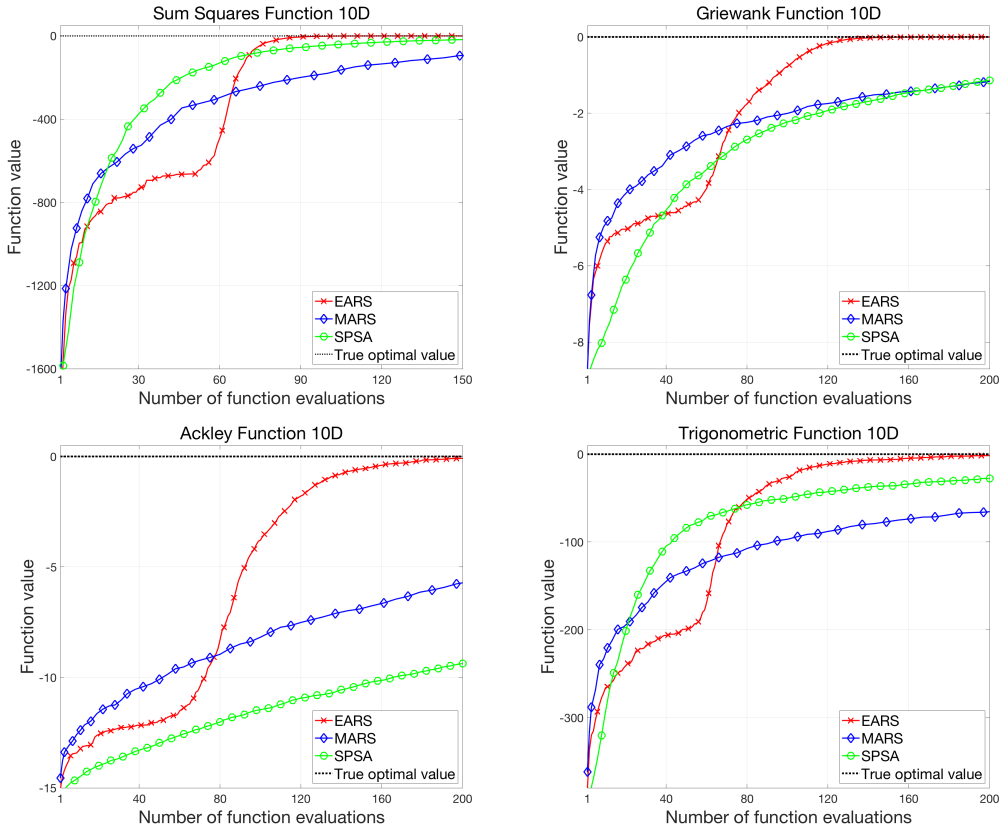

Fig. 1. Averaged performance of EARS, MARS and SPSA.

TABLE II
AVERAGE PERFORMANCE OF EARS, MARS, AND SPSA.

	EARS		MARS		SPSA	
	H^*	std_err	H^*	std_err	H^*	std_err
H_1	-3.69e-9	1.63e-11	-58.02	5.82e-1	-17.16	2.33
H_2	-1.52e-8	5.47e-10	-7.42e-1	6.61e-3	-1.11	5.42e-2
H_3	-8.96e-4	8.73e-5	-4.18	1.12e-2	-9.34	5.49e-1
H_4	-1.21e-2	1.01e-3	-70.34	2.76e-1	-27.62	1.23

initial model is obtained, it allows the algorithm to compute new distribution parameters using knowledge of the objective function predicted by the model. If the model can adequately capture the general trend of the response surface of the objective function, then the algorithm may quickly locate subsets of $\mathscr{X}$ containing high-quality solutions, leading to significantly improved performance.

VI. CONCLUSIONS

In this paper, we have proposed a random search algorithm called EARS for solving global optimization problems with Lipschitz continuous objective functions. EARS improves upon the MARS algorithm by incorporating a surrogate model to successively predict the response surface of an unknown objective function. The use of surrogate modeling provides a means to explicitly take into account the previous sampling information and thus allows the algorithm to conduct future search by exploiting knowledge of the objective function. Under appropriate conditions, we have shown that the algorithm converges globally to an optimal solution with probability one. Our preliminary computational results also indicate that the algorithm is promising and may find high-quality solutions very close to the optimal within a small number of function evaluations.

REFERENCES

[1] R. R. Barton, "Response Surface Methodology," *Encyclopedia of Operations Research and Management Science*, S. I. Gass and M. C. Fu, Eds. Springer, 2013, pp. 1307-1313.
[2] R. R. Barton and M. Meckesheimer, "Metamodel-Based Simulation Optimization," *Handbooks in Operations Research and Management Science: Simulation*, S. G. Henderson and B. L. Nelson, Eds. Elsevier, 2006, pp. 535-574.
[3] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1995.
[4] P. Deheuvels, "Strong Bounds for Multidimensional Spacings," *Z. Wahrscheinlichkeitstheorie und verw*, vol. 64, pp. 411-424, 1983.
[5] F. Glover, "Tabu Search: A Tutorial," *Interfaces*, vol. 20, pp. 74-94, 1990.

[6] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Boston, MA: Kluwer, 1989.
[7] H. M. Gutmann, "A Radial Basis Function Method for Global Optimization," *Journal of Global Optimization*, vol. 19, pp. 201-227, 2001.
[8] J. Hu, M. C. Fu, and S. I. Marcus, "A model reference adaptive search algorithm for global optimization," *Oper Res*, vol. 55, pp. 549-568, 2007.
[9] J. Hu and P. Hu, "Annealing Adaptive Search, Cross-Entropy, and Stochastic Approximation in Global Optimization," *Nav Res Log*, vol. 58, pp. 457-477, 2011.
[10] D. Jones, M. Schonlau, and W. Welch, "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, vol. 13, pp. 455-492, 1998.
[11] S. Kirkpatrick, C. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671-680, 1983.
[12] J. P. C. Kleijnen, "Response Surface Methodology for Constrained Simulation Optimization: An Overview," *Simul Model Pract Th*, Vol. 16, pp. 50-64, 2008.
[13] H. J. Kushner and D. S. Clark, *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. New York: Springer-Verlag, 1978.
[14] R. Y. Rubinstein and D. P. Kroese, *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation, and Machine Learning*. New York: Springer, 2004.
[15] L. Shi and S. Ólafsson, "Nested Partitions Method for Global Optimization," *Oper Res*, vol. 48, pp. 390-407, 2000.
[16] J. C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Trans Autom Control*, vol. 37, no. 3, pp. 332341, Mar. 1992.