Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

LANDSCAPE MODIFICATION MEETS SURROGATE OPTIMIZATION: TOWARDS DEVELOPING AN IMPROVED STOCHASTIC RESPONSE SURFACE METHOD

Michael C.H. Choi

Yale-NUS College and
Department of Statistics and Data Science
National University of Singapore
21 Lower Kent Ridge Rd
Singapore 119077, SINGAPORE

Venkatkrishna Karumanchi

Yale-NUS College
National University of Singapore
21 Lower Kent Ridge Rd
Singapore 119077, SINGAPORE

ABSTRACT

In global optimization, surrogate optimization algorithms such as the Stochastic Response Surface (SRS) method are often employed when the objective function is expensive to evaluate or when the gradient information is unavailable. The aim of this paper is to propose and analyze an improved SRS method that instead targets a transformed objective function. The core idea of the transformation rests on introducing a threshold parameter in which the landscape is modified when the algorithm is above this threshold, making the algorithm easier to climb out of a local minimum basin while preserving the set of the stationary points. We prove the asymptotic convergence of the proposed improved SRS method, and provide positive numerical results on some common global optimization benchmark functions which demonstrate the improved convergence of the proposed method. We stress that the proposed method can be implemented with minimal additional computational costs.

1 INTRODUCTION

Mathematical optimization is a discipline that studies the theory in minimizing a given objective function f and develops fast algorithms to perform such task. Among various acceleration techniques in optimization, landscape modification is a promising method recently introduced by the first author in (Choi 2020a; Choi 2020b). The core idea of landscape modification rests on introducing a threshold parameter c so that the landscape of the objective function f is modified once the algorithm is above c with a reduced critical height in the context of simulated annealing. It has been proved to yield faster convergence rate in the setting of kinetic simulated annealing Choi (2020a), Curie-Weiss model in statistical physics Choi (2020b) as well as the travelling salesman problem Choi (2020b) in operations research.

When the objective function f is computationally-expensive to evaluate, for instance in the context of Bayesian optimization Regis and Shoemaker (2007), a suitable optimization method in this setting is known as surrogate optimization. It consists of building an inexpensive approximate model from a set of randomly generated points, known as the response surface, to carry out the optimization task. The algorithm then proceeds iteratively to identify the next promising point to evaluate.

This paper aims at connecting landscape modification with surrogate optimization for cross-fertilization of the two areas. In particular, we propose a new method of constructing the response surface by utilizing the landscape modified objective function instead of the original objective function. The rest of this paper is organized as follows. In Section 2.1, we first give a review on the Stochastic Response Surface (SRS) method, and in Section 2.2 we summarize the literature on landscape modification. In Section 3, we present the main result of this paper that guarantees the convergence of the SRS method applied to the landscape-modified objective function. We then proceed to give encouraging and promising numerical results of the proposed method on some standard global optimization benchmark functions in Section 4. Finally, we conclude the paper in Section 5.

2 PRELIMINARIES

2.1 Multistart Local Metric Stochastic Response Surface Method (SRS)

This paper builds on a previous work by Regis and Shoemaker (2007) who use a response surface model to find the global minimum of expensive 'black-box' functions. That is, assuming a continuous function f has a unique global minimizer x^* over the domain $\mathscr{D}$, the SRS method applied to f on $\mathscr{D}$ converges almost surely to x^*. In each iteration of the SRS method, exactly one point is selected deterministically from a set of randomly generated candidate points. This selected point is then used to update (and improve) the response surface model to generate more candidate points in later iterations. This paper aims to improve on the *Multistart Local Metric* SRS Method proposed by Regis and Shoemaker (2007). The *Metric* SRS (MSRS) is a special case of SRS where the function evaluation point in each iteration is chosen deterministically, using a weighted score between the estimated function value obtained from the response surface model, and minimum distance from previously evaluated points. The *Multistart Local* MSRS method makes use of a Radial Basis Function (RBF) as the response surface and a varying step size to generate candidate points.

The framework for the *Multistart Local* MSRS function is as follows. Here, n is the number of previously evaluated points, $\mathscr{A}_n$ is the set of previously evaluated points, and $s_n(x)$ is the response surface model after n function evaluations.

Inputs:

1. A continuous real-valued function f defined on a compact hyper cube $\mathscr{D} = [a,b]^m \subseteq \mathbb{R}^d$, $m \leq d$.
2. A particular response surface model, in this case, radial basis functions.
3. A set of initial evaluation points $\mathscr{I} = \{x_1, \ldots, x_{n_0}\}$. These points come from a space-filling experimental design and various such techniques are discussed in Regis et al. (2007).
4. The number of candidate points in each iteration, denoted by t. We require $t = O(d)$.
5. The maximum number of function evaluations allowed denoted by N_{max}.

Output: The best point encountered by the algorithm.

Algorithm:

1. *(Initial expensive function evaluation).* Evaluate the function f at each point in $\mathscr{I}$. Set $n = n_0$ and set $\mathscr{A}_n = \mathscr{I}$. Let the point with the best function value in the set $\mathscr{A}_n$ be x_n^*.
2. While $n < N_{max}$:
   (a) *(Fit/Update Response Surface Model).* Use the data points $\mathscr{B}_n = \{(x_i, f(x_i)) : i = 1, \ldots, n\}$ to fit and update $s_n(x)$, the Surface Response Model.
   (b) *(Randomly Generate Candidate Points).* Randomly generate a set of t candidate points $\Omega_n = \{y_{n,1}, \ldots, y_{n,t}\}$ in $\mathbb{R}^d$ from a multivariate normal distribution with mean x_n^* and covariance matrix σI_n. The value of σ is modified contingent on the progress made by the algorithm. The value of σ is increased if the algorithm makes a lot of progress (the surrogate is accurate) and decreased when the algorithm is unable to make much progress (the surrogate model is inaccurate). If the value of σ is shrunk below a certain pre-defined threshold, the algorithm will be re-started using a different set of initial points, hence giving it the name of Multistart Local MSRS model.
   (c) *(Select next function evaluation point).* The evaluation point x_{n+1} is now selected deterministically from the t candidate points in Ω_n, using information from the response surface model $s_n(x)$ and the data points $\mathscr{B}_n = \{(x_i, f(x_i)) : i = 1, \ldots, n\}$. This process is explained in detail below.

(d) *(Do expensive function evaluation).* Evaluate the function f at x_{n+1}.
(e) *(Update information).* $\mathscr{A}_{n+1} := \mathscr{A}_n \cup x_{n+1}$; $\mathscr{B}_{n+1} := \mathscr{B}_n \cup (x_{n+1}, f(x_{n+1}))$. Let x^*_{n+1} be the point in $\mathscr{A}_{n+1}$ with the best function value. Reset $n := n+1$.

3. *(Return the best solution found).* Return $x^*_{N_{max}}$.

To implement the aforementioned algorithm, we need to specify a distance metric D on $\mathbb{R}^d$ and a set of non-negative weights $(w^R_n, w^D_n) : n = n_0, n_0+1, \ldots$ such that $w^R_n + w^D_n = 1$ for all $n \geq n_0$ for the response surface and the distance criteria. We will now proceed to describe how step $2(c)$ of the algorithm is implemented.

1. *(Estimate the value of the function at the candidate points).* Compute $s_n(x)\ \forall x \in \Omega_n$. Also, compute $s^{min}_n = min\{s_n(x) : x \in \Omega_n\}$ and $s^{max}_n = max\{s_n(x) : x \in \Omega_n\}$.
2. *(Determine the minimum distance from previously evaluated points).* Compute $\Delta_n(x) = min\{D(x, x_i) : 1 \leq i \leq n\}\ \forall x \in \Omega_n$. Also compute $\Delta^{max}_n = max\{\Delta_n(x) : x \in \Omega_n\}$ and $\Delta^{min}_n = min\{\Delta_n(x) : x \in \Omega_n\}$.
3. *(Compute the score for the response surface criterion).* For all $x \in \Omega_n$, compute $V^R_n(x) = (s_n(x) - s^{min}_n)/(s^{max}_n - s^{min}_n)$ if $s^{min}_n \neq s^{max}_n$ and $V^R_n(x) = 1$ otherwise.
4. *(Compute the score for the distance criterion).* For all $x \in \Omega_n$, compute $V^D_n(x) = (\Delta^{max}_n - \Delta_n(x))/(\Delta^{max}_n - \Delta^{min}_n)$ if $\Delta^{max}_n \neq \Delta^{min}_n$ and $V^D_n(x) = 1$ otherwise.
5. *(Compute the Weighted Score).* For all $x \in \Omega_n$, compute $\mathscr{W}_n(x) = w^R_n V^R_n(x) + w^D_n V^D_n(x)$.
6. *(Select the next evaluation point).* Let x_{n+1} be the point in Ω_n that minimizes $\mathscr{W}_n$.

2.2 Landscape Modification

In Choi (2020b), the idea of landscape modification is first proposed and applied in a finite state space setting in Markov chain Monte Carlo algorithms such as the Metropolis-Hastings and simulated annealing algorithms. Instead of optimizing with respect to the original objective function f, in landscape modification we instead target a transformed objective function $f^g_{\varepsilon,c}$, mathematically defined to be

$$f^g_{\varepsilon,c}(x) := \int_{f_{min}}^{f(x)} \frac{1}{g((u-c)_+) + \varepsilon}\, du,$$

where $f_{min} := \min_x f(x)$ and $a_+ = \max\{a, 0\}$ is the non-negative part of $a \in \mathbb{R}$. This transformation is based on the introduction of three parameters, namely

- a threshold parameter c with $c \geq f_{min}$ in which the landscape is modified once the algorithm is above this threshold parameter c. A common way to tune this threshold parameter c is to set it to be the running minimum generated by the algorithm on the fly.
- a non-negative and non-decreasing function g with $g(0) = 0$ that describes how the landscape is transformed. Common choices are $g(x) = x$ or $g(x) = x^2$.
- a temperature parameter $\varepsilon > 0$. A common choice is to take $\varepsilon = 1$. We shall also denote the inverse temperature to be $\beta = 1/\varepsilon$.

It was then proven that this technique brings about benefits and speedups in the analysis of the Curie-Weiss model in statistical physics and stochastic optimization using simulated annealing. The core idea of the transformation relies on introducing a threshold parameter c in which the landscape is modified once the algorithm is above this threshold c, making the algorithm easier to climb out of a local minimum basin while preserving the set of the stationary points. In the context of simulated annealing it is proved that this transformation can effectively reduce the critical height of the landscape and thus offer promising speedup. This paper will investigate the benefit of landscape modification to the *Multistart Local* MSRS method discussed in Section 2.1.

We now proceed to provide a few examples of applying landscape modification to a function. As it is sometimes impossible for us to know a piori the global minimum of the target function, we will instead consider the difference of the target function for $x, y \in X$.

$$f_{\varepsilon,c}^{g}(y) - f_{\varepsilon,c}^{g}(x) = \int_{f(x)}^{f(y)} \frac{1}{g((u-c)_+)+\varepsilon}\,du.$$

The difference of the target function can then be computed in a piecewise manner owing to the following integration:

$$f_{\varepsilon,c}^{g}(y) - f_{\varepsilon,c}^{g}(x) = \begin{cases} \beta(f(y)-f(x)), & \text{if } c \geq f(y) > f(x), \\ \beta(c-f(x)) + \int_{c}^{f(y)} \frac{1}{g(u-c)+\varepsilon}\,du, & \text{if } f(y) > c \geq f(x), \\ \int_{f(x)}^{f(y)} \frac{1}{g(u-c)+\varepsilon}\,du, & \text{if } f(y) > f(x) > c. \end{cases} \tag{1}$$

The remaining cases of $\{f(y) < f(x)\}$ is omitted owing to symmetry. We now specialize into two cases, namely linear landscape modification with $g(x) = x$ and quadratic landscape modification $g(x) = x^2$ on an objective function $f : \mathbb{R}^n \mapsto \mathbb{R}$:

1. Linear landscape modification and log-transformed landscape: We take $g(u) = u$. For $x, y \in \{f(y) > f(x) \geq c\}$, since

   $$\int_{f(x)}^{f(y)} \frac{1}{u-c+\varepsilon}\,du = \ln\left(\frac{f(y)-c+\varepsilon}{f(x)-c+\varepsilon}\right),$$

   putting the expression back into (1) gives

   $$f_{\varepsilon,c}^{g}(y) - f_{\varepsilon,c}^{g}(x) = \begin{cases} \beta(f(y)-f(x)), & \text{if } c \geq f(y) > f(x), \\ \beta(c-f(x)) + \ln\left(\frac{f(y)-c+\varepsilon}{\varepsilon}\right), & \text{if } f(y) > c \geq f(x), \\ \ln\left(\frac{f(y)-c+\varepsilon}{f(x)-c+\varepsilon}\right), & \text{if } f(y) > f(x) > c. \end{cases}$$

2. Quadratic landscape modification and inverse-tangent-transformed landscape: We specialize into $g(u) = u^2$. In this case, the effect of landscape modification gives an inverse-tangent-transformed objective function whenever $f(x) > c$.
   For $x, y \in \{f(y) > f(x) \geq c\}$, using the inverse-tangent difference formula we obtain

   $$\int_{f(x)}^{f(y)} \frac{1}{(u-c)^2+\varepsilon}\,du = \sqrt{\beta}\arctan\left(\frac{\sqrt{\beta}(f(y)-f(x))}{1+\beta(f(y)-c)(f(x)-c)}\right),$$

   and substituting back into (1) gives

   $$f_{\varepsilon,c}^{g}(y) - f_{\varepsilon,c}^{g}(x) = \begin{cases} \beta(f(y)-f(x)), & \text{if } c \geq f(y) > f(x), \\ \beta(c-f(x)) + \sqrt{\beta}\arctan\left(\sqrt{\beta}(f(y)-c)\right), & \text{if } f(y) > c \geq f(x), \\ \sqrt{\beta}\arctan\left(\frac{\sqrt{\beta}(f(y)-f(x))}{1+\beta(f(y)-c)(f(x)-c)}\right), & \text{if } f(y) > f(x) > c. \end{cases}$$

3 LANDSCAPE MODIFICATION MEETS SURROGATE OPTIMIZATION

In this section, we first give an asymptotic convergence result on using landscape modification coupled with the SRS method. The following result guarantees that the sequence of points generated by the SRS method with landscape modification converges almost surely to the desired global minimum under appropriate regularity conditions:

Theorem 1 Let f be a function defined on the compact hypercube $\mathscr{D} \subseteq \mathbb{R}^d$ with a unique global minimum x^* such that $f(x^*) > -\infty$. For a fixed non-negative and non-decreasing function g, threshold parameter $c \geq f(x^*)$ and temperature parameter $\varepsilon > 0$, the SRS method applied to $f^g_{\varepsilon,c}$ generates a sequence $\{X^*_n\}_{n\geq 1}$ such that $X^*_n \longrightarrow x^*$ almost surely.

Proof. As it has been proved in Choi (2020b) that the global minimum of the landscape-modified objective function $f^g_{\varepsilon,c}$ is exactly the same as the global minimum of the original objective function f, proving that the SRS Algorithm extracts the global minimum of an arbitrary function passed as input would suffice. Then, as the SRS algorithm can approximate the global minimum of landscape modified objective function, and as this minimum would be the same as the minimum of the orginal objective function, we will have proved that the SRS Algorithm coupled with landscape modification provides us with global minimum we desire. □

3.1 The Classical Proof of the Convergence of the SRS Method Applied to f

In this subsection, we recall the classical proof of the convergence of SRS method applied to f as in Regis and Shoemaker (2007). Before we commence our proof, we will clarify certain notation.

In the reminder of this section, uppercase letters will denote random vectors to distinguish them from ordinary vectors in $\mathbb{R}^d$. Assume that $\mathscr{D} = [a,b]^m \subseteq \mathbb{R}^d$, $m \leq d$. Consider any random vector X which is realised in $\mathbb{R}^d$. The random vector $X_{\mathscr{D}}$, whose realisations are always in $\mathscr{D}$, will be defined as follows. Consider a sample point w:

$$X_{\mathscr{D}}(w) = \begin{cases} X(w), & \text{if } X(w) \in \mathscr{D} \\ min(max(a, X(w), b)), & \text{else.} \end{cases}$$

In the second case, note that the minimum and the maximum are taken componentwise. In this case, $X_{\mathscr{D}}(w)$ is the point in $\mathscr{D}$ which is nearest to $X(w)$. In the notation below, X_n is the random vector representing the nth function evaluation point x_n and $Y_{n,j}$ is the random vector representing the random candidate point $y_{n,j}$ before it was forced to be in $\mathscr{D}$. Note that for all $n \geq n_0$, the value of X_{n+1} is selected deterministically from the values of the random vectors $(Y_{n,1})_{\mathscr{D}}, (Y_{n,2})_{\mathscr{D}}, \ldots, (Y_{n,t})_{\mathscr{D}}$.

For each $n \geq n_0$, let $\mathscr{E}_n := \{X_1, \ldots, X_{n_0}, Y_{n_0,1}, \ldots, Y_{n_0,t}, \ldots, Y_{n,1}, \ldots, Y_{n,t}\}$. We define $\mathscr{E}_{n_0-1} := \{X_1, \ldots, X_{n_0}\}$. Let $n \geq n_0$. After the nth function evaluation, the entire history of the algorithm is completely determined by the random vectors in $\mathscr{E}_{n-1}$. $B(x,\delta)$ is the open ball of radius δ centred at x and $\sigma(\mathscr{E}_{n-1})$ is the $\sigma-$ field generated by the random vectors in $\mathscr{E}_{n-1}$.

We now proceed with proof of the convergence of the SRS method.

Theorem 2 Let f be a function defined on the compact hypercube $\mathscr{D} \subseteq \mathbb{R}^d$ with a unique global minimum x^* such that $f(x^*) > -\infty$. Define the sequence of random vectors $\{X^*_n\}_{n\geq 1}$ as follows: $X^*_1 = X_1$ and $X^*_n = X_n$, if $f(X_n) < f(X^*_{n-1})$ while $X^*_n = X^*_{n-1}$ otherwise. Suppose the SRS method generates random vectors $\{X_n\}_{n\geq 1}$ and $\{Y_{n,1}, \ldots, Y_{n,t}\}_{n\geq n_0}$ from a normal distribution centred at X^*_n with covariance matrix $\sigma^2_n I_n$, where $\inf_{n\geq n_0} \sigma_n > 0$. Then $X^*_n \longrightarrow x^*$ almost surely.

Proof. Fix $\varepsilon > 0$ and $n \geq n_0 + 1$. It is observed that $[X_n \in \mathscr{D} : f(X_n) < f(x^*) + \varepsilon] = [X_n \in \mathscr{D} : |f(X_n) - f(x^*)| < \varepsilon]$. Since f is continuous on x^*, there exists $\delta(\varepsilon) > 0$ such that $|f(x) - f(x^*)| < \varepsilon$ whenever $\|x - x^*\| < \delta(\varepsilon)$. Hence, $[X_n \in \mathscr{D} : |f(X_n) - f(x^*)| < \varepsilon] \supseteq [X_n \in \mathscr{D} : \|x - x^*\| < \delta(\varepsilon)]$, and so,

$$
\begin{aligned}
P[X_n \in \mathscr{D} : |f(X_n) - f(x^*)| < \varepsilon | \sigma(\mathscr{E}_{n-2})] &\geq P[X_n \in \mathscr{D} : \|X_n - x^*\| < \delta(\varepsilon) | \sigma(\mathscr{E}_{n-2})] \\
&= P[X_n \in B(x^*, \delta(\varepsilon)) \cap \mathscr{D} | \sigma(\mathscr{E}_{n-2})].
\end{aligned}
$$

Note that if $(Y_{n-1,j})_{\mathscr{D}} \in B(x^*, \delta(\varepsilon)) \cap \mathscr{D}$ for each $j = 1, \ldots, t$, then the evaluation point $X_n \in B(x^*, \delta(\varepsilon)) \cap \mathscr{D}$. So,

$$
\begin{aligned}
P[X_n \in B(x^*, \delta(\varepsilon)) \cap \mathscr{D} | \sigma(\mathscr{E}_{n-2})] &\geq P[(Y_{n-1,j})_{\mathscr{D}} \in B(x^*, \delta(\varepsilon)) \cap \mathscr{D}, j = 1, \ldots, t | \sigma(\mathscr{E}_{n-2})] \\
&\geq P[Y_{n-1,j} \in B(x^*, \delta(\varepsilon)) \cap \mathscr{D}, j = 1, \ldots, t | \sigma(\mathscr{E}_{n-2})] \\
&= \prod_{j=1}^{t} P[Y_{n-1,j} \in B(x^*, \delta(\varepsilon)) \cap \mathscr{D} | \sigma(\mathscr{E}_{n-2})].
\end{aligned}
$$

The equality in the product is owing to the conditional independence of the $Y_{n-1,j}$ vectors $\forall j \in \{1, \ldots, t\}$ given the random vectors in $\mathscr{E}_{n-2}$.

Now we are left to prove that $P[Y_{n-1,j} \in B(x^*, \delta(\varepsilon)) \cap \mathscr{D} | \sigma(\mathscr{E}_{n-2})] > 0$.

Define $\psi_{\mathscr{D}}(\delta) := \inf_{x \in \mathscr{D}} \mu(B(x, \delta) \cap \mathscr{D})$, where μ is the Lebesgue measure on $\mathbb{R}^d$. Observe that for the compact hypercube $\mathscr{D}$, we have $\psi_{\mathscr{D}}(\delta) > 0$ for any $\delta > 0$. Since $\inf_{n \geq n_0} \sigma_n > 0$, the conditional density of $Y_{n-1,j}$ for each $n-1 \geq n_0$ is given by

$$
g_{n-1,j}(y | \sigma(\mathscr{E}_{n-2})) = (2\pi\sigma_{n-1}^2)^{-d/2} \exp\left(\frac{-\|y - X_{n-1}^*\|^2}{2\sigma_{n-1}^2} \right), \; y \in \mathbb{R}^d.
$$

Note that

$$
g_{n-1,j}(y | \sigma(\mathscr{E}_{n-2})) \geq \left(2\pi \left(\sup_{n-1 \geq n_0} \sigma_{n-1} \right)^2 \right)^{-d/2} \exp\left(\frac{-\|y - X_{n-1}^*\|^2}{2\sigma_{n-1}^2} \right) =: C > 0,
$$

for all $y \in \mathscr{D}$.

Thus,

$$
\begin{aligned}
P[Y_{n-1,j} \in B(x^*, \delta) \cap \mathscr{D} | \sigma(\mathscr{E}_{n-2})] &= \int_{B(x^*, \delta) \cap \mathscr{D}} g_{n-1,j}(y | \sigma(\mathscr{E}_{n-2})) dy \\
&\geq C\mu(B(x^*, \delta) \cap \mathscr{D}) \geq C\psi_{\mathscr{D}(\delta)} > 0.
\end{aligned}
$$

So,

$$
P[X_n \in B(x^*, \delta(\varepsilon)) \cap \mathscr{D} | \sigma(\mathscr{E}_{n-2})] \geq \prod_{j=1}^{t} v_j(x^*, \delta(\varepsilon)) =: L(\varepsilon) > 0.
$$

Thus, we have $P[X_n \in \mathscr{D} : f(X_n) < f(x^*) + \varepsilon | \sigma(\mathscr{E}_{n-2})] \geq L(\varepsilon)$. Thus, by following an argument similar to the result in page 40 of Spall (2005), we attain convergence almost surely. □

4 NUMERICAL EXPERIMENTS

In this section, we conduct numerical experiments for evaluating the performance of the SRS algorithm with and without landscape modification on some common global optimization benchmark functions. The experiments were run using the PySOT package in Python, a toolbox for the optimization of computationally expensive black-box objective functions. The source code for surrogate optimization in the PySOT package was modified to optimize with respect to the landscape-modified objective functions $f^g_{\varepsilon,c}$ and can be found here on the author's github repository. Experiments were performed by running the optimization algorithm (the number of evaluations vary according to the dimensionality of the benchmark function) 30 times for each benchmark objective function and the various landscape modified variants of the original objective function and then computing summary statistics. The random seed was the same for both SRS and the landscape-modified SRS method during each of the 30 runs for fair comparison. We also tune our threshold parameter value c by either setting it to be the running minimum found in each iteration of the surrogate optimization algorithm or to a certain fixed value which was found by comparing it with various other possible fixed values for c. We choose the temperature parameter $\varepsilon = 1$ in all our numerical experiments. Thus, our landscape-modified objective function is possibly adaptively changing throughout the course of the SRS algorithm, but we note that the set of stationary points is preserved throughout.

For the remainder of this section 'objective' will refer to the results of applying the SRS method to the original objective function f, 'linear a' will be the objective function in the linear modified landscape with an adaptive c, 'linear c' will be the objective function in the linear modified landscape with c fixed (for example, if c is fixed to 1, it will be denoted as linear 1), i.e. $f^{g(x)=x}_{1,c}$, 'quadratic a' will be the objective function in the quadratic modified landscape with an adaptive c, 'quadratic c' will be the objective function in the quadratic modified landscape with c fixed, i.e. $f^{g(x)=x^2}_{1,c}$.

We postpone the discussion of the numerical results and the plots in Section 4.6.

4.1 Ackley Function

$$f(x_1,\dots,x_d) = -20\exp\left(-0.2\sqrt{\frac{1}{d}\sum_{j=1}^{d} x_j^2}\right) - \exp\left(\frac{1}{d}\sum_{j=1}^{d}\cos(2\pi x_j)\right) + 20 - \exp(1)$$

where $-15 \le x_i \le 20$ and the global minimum is $f(0,\dots,0) = 0$, and $d = 10$.

Table 1: Summary statistics of the original SRS and the landscape-modified SRS algorithm on the Ackley function in 30 independent runs.

method	mean	median	standard deviation	best result	worst result
objective	0.9023	0.8279	0.4011	0.3110	2.2723
linear a	1.0348	0.7881	0.5366	0.3232	2.4860
linear 0	0.9567	0.8050	0.4752	0.4217	2.4860
quadratic a	1.1069	0.9119	0.5465	0.3709	2.9767
quadratic 0	1.0873	0.9557	0.4849	0.3709	2.4384

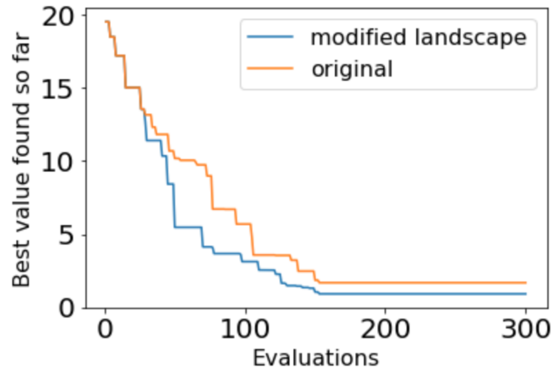

Figure 1: Sample run on the Ackley function in which 'linear a' found a minimum of 0.737749 and the original SRS found a minimum of 1.677980.

4.2 Rosenbrock Function

$$f(x_1,\ldots,x_d) = \sum_{i=1}^{d} -1[100\left(x_{i+1}-x_i^2\right)^2 + (x_i-1)^2]$$

where $-2.048 \leq x_i \leq 2.048$ and the global minimum is $f(1,\ldots,1) = 0$, and $d = 10$.

Table 2: Summary statistics of the original SRS and the landscape-modified SRS algorithm on the Rosenbrock function in 30 independent runs.

method	mean	median	standard deviation	best result	worst result
objective	8.6185	8.8832	1.7402	4.4037	12.4250
linear a	7.3535	7.5624	1.3576	2.7582	9.6538
linear 1	7.2728	7.3866	0.7831	4.6646	8.5082
quadratic a	7.6232	7.8529	1.4907	2.7750	9.6461
quadratic 0	6.6163	7.2987	2.1640	1.0440	9.1723

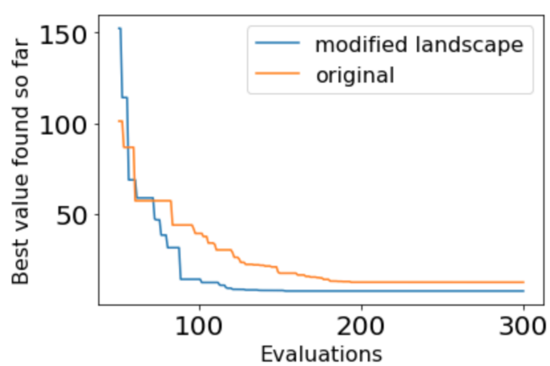

Figure 2: Sample run on the Rosenbrock function in which 'linear 1' found a minimum of 7.614494 and the original SRS method gives a minimum of 12.425069.

4.3 Dixon-Price Function

$$f(x_1,\ldots,x_d) = (x_1-1)^2 + \sum_{i=2}^{d} i(2x_i^2 - x_{i-1})^2)$$

where $-10 \leq x_i \leq 10$ and the global minimum is $f(x_1,\ldots,x_d) = 0$, when $x_i = 2^{-\frac{2^i-2}{2^i}}$, and $d = 10$.

Table 3: Summary statistics of the original SRS and the landscape-modified SRS algorithm on the Dixon-Price in 30 independent runs.

method	mean	median	standard deviation	best result	worst result
objective	8.6137	5.7372	7.3143	1.7414	36.5023
linear a	1.7440	0.7492	3.1117	0.5484	15.0939
linear 0	0.6453	0.6055	0.1050	0.5393	0.9126
quadratic a	122.4847	119.6667	99.5981	0.6550	377.3572
quadratic 0	2.0505	0.8139	3.3107	0.5923	18.0407

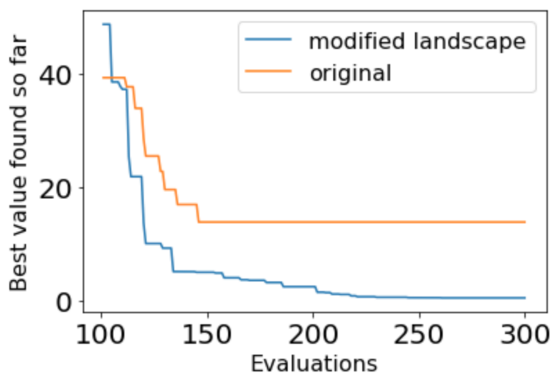

Figure 3: Sample run on the Dixon-Price function in which ‘quadratic 0’ found a minimum of 0.539344 and the original SRS method found a minimum of 13.865434.

4.4 Three-Hump Camel Function

$$f(x_1,x_2) = 2x_1^2 - 1.05x_1^4 + \frac{x_1^6}{6} + x_1x_2 + x_2^2$$

where $-5 \leq x_i \leq 5$ and the global minimum is $f(0,0) = 0$.

Table 4: Summary statistics of the original SRS and the landscape-modified SRS algorithm on the Three-Hump Camel function in 30 independent runs.

method	mean	median	standard deviation	best result	worst result
objective	0.0206	0.0006	0.0405	7.0e-05	0.1557
linear a	3.3e-05	2.7e-05	2.9e-05	1.2e-07	0.0001
linear 1	3.9e-05	2.6e-05	3.8e-05	8.6e-07	0.0001
quadratic a	3.8e-05	3.6e-05	3.3e-05	8.4e-07	0.0001
quadratic 1	4.1e-05	3.6e-05	3.4e-05	6.1e-07	0.0001

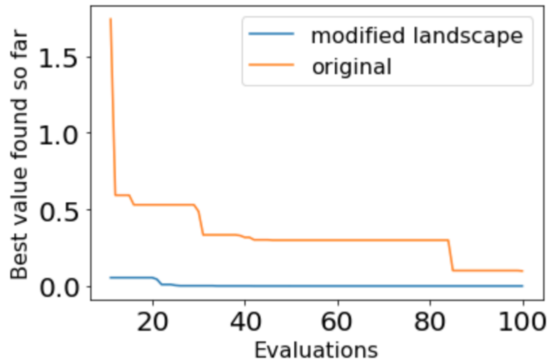

Figure 4: Sample run on the three-hump camel function in which 'quadratic 1' found a minimum of 4.482e-05 and the original SRS method found a minimum of 0.098218.

4.5 Six-Hump Camel Function

$$f(x_1,x_2) = \left(4 - 2.1x_1^2 + \frac{x_1^4}{3}\right) x_1^2 + x_1x_2 + \left(-4 + 4x_2^2\right) x_2^2$$

where $-3 \leq x_1 \leq 3$, $-2 \leq x_2 \leq 2$ and the global minimum is $f(0.0898, -0.7126) = -1.0316$.

Table 5: Summary statistics of the original SRS and the landscape-modified SRS algorithm on the Six-Hump Camel function in 30 independent runs.

method	mean	median	standard deviation	best result	worst result
objective	-1.0315	-1.0315	0.0001	-1.0316	-1.0309
linear a	-1.0315	-1.0315	2.7e-05	-1.0316	-1.0315
linear 0	-1.0315	-1.0315	5.74e-05	-1.0316	-1.0313
quadratic a	-1.0315	-1.0315	7.6e-05	-1.0316	-1.0312
quadratic 0	-1.0315	-1.0315	0.0001	-1.0316	-1.0311

4.6 Discussion

It is evident from the numerical results that the landscape modification of the objective function is a useful technique to improve the convergence to the global minimum of certain expensive black-box functions. In the aforementioned experiments it is observed that the landscape-modified surrogate optimization provided an advantage over the standard surrogate optimization algorithm for all benchmark functions (which are all valley-shaped), except the Ackley function. For Ackley function, we see from Figure 1 that there are instances in which the landscape-modified SRS algorithm outperforms the original SRS algorithm.

5 CONCLUSION

In this paper we propose a new SRS method that optimize with respect to the landscape-modified function $f^g_{\varepsilon,c}$ instead of the original objective function f. The proposed method provably converges to the global minimum of f as demonstrated in Theorem 1, and it offers promising numerical results that outperform the original SRS method on some global optimization benchmark functions.

ACKNOWLEDGMENTS

Both authors are grateful to the constructive feedback of the reviewers. Both authors acknowledge the support from the Yale-NUS College special pocket research grant and the startup funding from the Yale-NUS College.

REFERENCES

Choi, M. C. 2020a. "On the Convergence of an Improved and Adaptive Kinetic Simulated Annealing". *arXiv preprint arXiv:2009.00195*. https://arxiv.org/abs/2009.00195, accessed October 12, 2022.

Choi, M. C. 2020b. "On the Convergence of an Improved Discrete Simulated Annealing via Landscape Modification". *arXiv preprint arXiv:2011.09680*. https://arxiv.org/abs/2011.09680, accessed October 12, 2022.

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

Spall, J. C. 2005. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*, Volume 65. Upper Saddle River, New Jersey: John Wiley & Sons.

AUTHOR BIOGRAPHIES

MICHAEL C.H. CHOI is an Assistant Professor at the Yale-NUS College and the Department of Statistics and Data Science at National University of Singapore. He received his Ph.D. in Operations Research at the School of Operations Research and Information Engineering (ORIE) at Cornell University. His research interests center around the theory and applications of Markov chains and Markov processes, with a particular focus on stochastic algorithms driven by Markov chains in the area of Markov chain Monte Carlo (MCMC). His e-mail address is michael.choi@yale-nus.edu.sg. Here is his website.

VENKATKRISHNA KARUMANCHI is a rising senior student majoring at Mathematical, Computational and Statistical Sciences with a special focus on Mathematics at the Yale-NUS College, National University of Singapore. His academic interests include Mathematical Analysis and Probability Theory. His email address is krishnakarumanchi@u.yale-nus.edu.sg.