J Glob Optim (2016) 64:469–482
DOI 10.1007/s10898-014-0256-1

The robust constant and its applications in random global search for unconstrained global optimization

Zheng Peng · Donghua Wu · Wenxing Zhu

Received: 29 July 2013 / Accepted: 18 November 2014 / Published online: 30 November 2014

Abstract Robust analysis is important for designing and analyzing algorithms for global optimization. In this paper, we introduce a new concept, robust constant, to quantitatively characterize the robustness of measurable sets and functions. The new concept is consistent to the theoretical robustness presented in literatures. This paper shows that, from the respects of convergence theory and numerical computational cost, robust constant is valuable significantly for analyzing random global search methods for unconstrained global optimization.

Keywords Unconstrained global optimization · Robust constant · Random global search · Pure adaptive search · Algorithm analysis

1 Introduction

In this paper, we consider the unconstrained global minimization problem of the form:

$$c^* = \min_{x \in R^n} f(x) \tag{1.1}$$

This work was supported by the Natural Science Foundation of China (61170308), Major Science Foundation of Fujian Provincial Department of Education (JA14037) and talent foundation of Fuzhou University (XRC-1043).

Z. Peng (✉)
College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China
e-mail: pzheng@fzu.edu.cn

D. Wu
Department of Mathematics, College of Science, Shanghai University, Shanghai 200444, China
e-mail: dhwu@shu.edu.cn

W. Zhu
Center for Discrete Mathematics and Theoretical Computer Science, Fuzhou University, Fuzhou 350108, China
e-mail: wxzhu@fzu.edu.cn

where $f : R^n \to R$ is a summable continuous function but not necessarily convex. Assume that the function f is lower-bounded and the level set $\{x \in R^n : f(x) < f(x_0)\}$ is bounded for any $x_0 \in R^n$.

Generally, we need a certain smoothness assumption on the objective function f such that some gradient-based optimization methods, e.g., the steepest descent method or the Newton-type methods, can be used to find a local minimizer. Unless problem (1.1) has some special structures, e.g., convex programming or fractional linear programming, finding a global minimizer is an NP-hard problem, see Vavasis [27]. However, finding a global minimizer of a general non-convex function is a common task in the real-world applications. Thus, various global minimization techniques have been developed, interested readers are referred to [8,12,24,26] for excellent survey papers.

Among the developed methods, global search is frequently used in engineering applications. At each iteration, global search seeks a candidate solution from the feasible set of the problem to be solved, while local search finds a next candidate in the neighborhood of the current solution. Pure random search (PRS) and pure adaptive search (PAS) are two classical global search methods with random sampling techniques.

The PRS method was originally proposed by Brooks [4]. Let $H_0 = \{x \in R^n : f(x) < f(x_0)\}$, where $x_0 \in \mathrm{dom} f$ is an initial solution. If $f(x_0) > c^*$, then H_0 is nonempty. At the k-th iteration, PRS produces a candidate x_k such that

$$f(x_k) = \min\{f(x_i), i = 1, 2, \ldots, N,\ x_i \text{ i.i.d. on } H_0\}. \tag{1.2}$$

The PRS method is extremely robust, easy to implement but the convergence is very slow. Several stochastic methods have been developed as variations of PRS. Appel, Labarre and Radulovic [1] designed a simple accelerated random search (ARS) algorithm. Kabirian [10] proposed an algorithm that merges Ranking and Selection procedures with a large class of random search methods for continuous simulation optimization problems. Radulovic [19] proposed a pure accelerated random search (PARS), and proved that it converges for all measurable functions with the essential supremum, and for a very large subclass of functions the convergence rate is exponential. Price, Reale and Robertson [18] presented a variation of ARS called one side cut ARS method, which is a direct search method for bounded constrained global optimization. The other interesting random search methods, combining with partition, clustering, multi-level and multi-start, etc., have been presented by many authors. The readers are referred to [16,17,20,21,25] and the references therein.

The PAS method, originally studied by Zabinsky et al. [13,30], constructs a sequence of interior points uniformly distributed within the corresponding sequence of nested improving regions of the feasible space. For a given x_k, let $H_k = \{x \in R^n : f(x) < f(x_k)\}$ be the level set of the k-th iteration. The PAS method randomly generates x_{k+1} uniformly distributed in H_k. By this mechanism it always has $f(x_{k+1}) < f(x_k)$, and consequently $H_{k+1} \subset H_k$. It has been shown in [30] that, if $f(x)$ is Lipschitz then the convergence rate of the PAS method is exponential. Radulovic [19] claimed that the PAS method is "theoretical optimum" in the setting of pure, non-gradient based, stochastic optimization schemes. However, the PAS method is impossible to implement in general, since identifying the current level set H_k is intrinsically harder than actually finding an optimal solution. Zabinsky et al. [31] defined strong and weak variations of PAS in the setting of finite global optimization, and proved their linear (in dimension) complexity. Combining with the ideas of simulated annealing, Bulger and Wood [5] presented a unified theory which yields both the finite and continuous results for PAS. This method is allowed to " hesitate" before improvement continues, hence it is termed hesitant adaptive search. Baritompa et al. [2] introduced an algorithm termed pure localisation search which attempts to reach the practical ideal. This method is a relaxation

of PAS, and is more likely to be efficiently implementable, yet still possesses the desirable complexity of PAS. Bulger, Baritompa and Wood [6] proposed an implementation which uses the Grover quantum computational search algorithm to generate the PAS iterates, called the Grover adaptive search, to realize PAS for functions satisfying certain conditions. The cross-entropy (CE) method proposed by Rubinstein et al. [3,11,22,23] and the model reference adaptive search method proposed by Hu et al. [9] can be also looked as variations of PAS. They are random global search methods using importance sampling.

The integral level-set method (ILSM), originally proposed by Zheng [7,33], is a deterministic global search method. It constructs two sequences in the ILSM: a sequence of the level value $\{c_k\}$ and a sequence of the corresponding level set $\{H_{c_k}\}$, which are

$$c_{k+1} = \frac{1}{\mu(H_{c_k})} \int_{H_{c_k}} f(x) d\mu, \tag{1.3}$$

$$H_{c_{k+1}} = \left\{x \in R^n : f(x) < c_{k+1}\right\}, \tag{1.4}$$

where μ is the Lebesgue measure on R^n. Under the assumptions that $f(x)$ is lower semi-continuous and robust,[1] and the level-set H_0 is measurable for all $c_0 = f(x_0)$, Zheng [32] proved that the level-value sequence $\{c_k\}$ converges to the optimal value, and respectively, the level-set sequence $\{H_{c_k}\}$ converges to the optimal solution set. The ILSM has the same difficulty as that of PAS, i.e., the level set H_{c_k} is hard to be determined. So, the implementable algorithm of ILSM calculates integration in (1.3) and determines level set in (1.4) by the Monte-Carlo method. To do so, it leads to a drawback that convergence of the implementable algorithm cannot be guaranteed. To improve the ILSM, Yao et al. [29] presented an optimality condition and an algorithm with deviation integral for global optimization. Wu et al. [28] presented a sufficient and necessary condition for computing the essential infimum. To overcome the drawback of identifying the current level set H_{c_k}, Peng et al. [14] proposed a level-value estimation method (LVEM) based on the idea of ILSM, and implemented the LVEM by the Monte-Carlo method using importance sampling, and proved convergence of the implementable algorithm. The LVEM bridged the gap between the conceptual algorithm and implementation of the ILSM. Peng, Shen and Wu [15] proposed a modified integral level-set method (MILSM) based on importance sampling. Let

$$F_k(x) = \begin{cases} c_k, & \text{if } f(x) \geq c_k, \\ f(x), & \text{otherwise}, \end{cases} \tag{1.5}$$

the MILSM updates the level value by

$$c_{k+1} = \frac{1}{N} \sum_{t=1}^{N} F_k(X_t) \tag{1.6}$$

where X_t is independently identically distributed (i.i.d. for short) from distribution with the density $g_k(x)$ on R^n. The efficiency of the MILSM depends on sample distribution characterized by the density $g_k(x)$. The cross-entropy method provides a novel idea for choosing and updating rule of sampling density function.

Given a sample $x \in R^n$, we say it is an effectual sample with respect to the set H if and only if $x \in H$. Numerous computational experiments showed that, the ratio of effectual sample depends on not only the density $g(x)$ but also the set H. To characterize the property of a set related to the ratio of effectual sample, Zheng [32] introduced robust analysis for global optimization. Specifically, he presented some related concepts of robust set and robust

[1] The robustness of a function defined by Zheng [32] will be introduced in the next section.

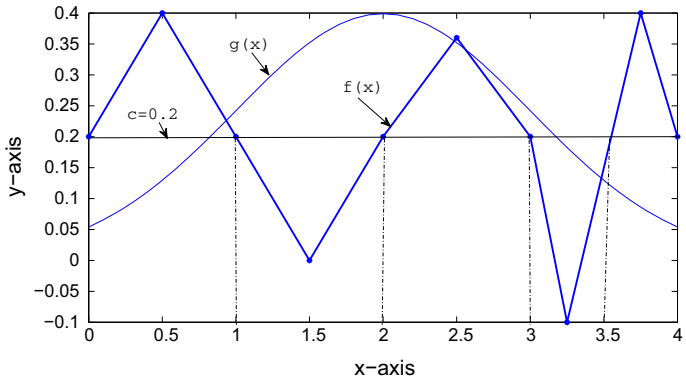

Fig. 1 The relaxation of level set and density function

function for qualitative description of the robustness. To quantitatively analyze the robustness of robust set and robust function, we introduce a new important concept, i.e., robust constant, in this paper. We will show that, robust constant is valuable significantly for analyzing convergence and efficiency of random global search techniques for global optimization.

The rest of this paper is organized as follows. Section 2 gives the concept of robust constant. Section 3 proposes a modified PAS method for global optimization, and analyzes its convergence by using robust constant. Utilizing the global robust constant, we give a simple and checkable stopping criterion for the proposed method. Section 4 presents a simple numerical experiment which shows that, the robust constant is an important parameter for global optimization when the modified PAS method (a representative of random global search technique) is used. Finally, Sect. 5 gives some concluding remarks.

2 Robust constant

A sample density function $g(x)$ is said to be "good" with respect to (w.r.t. for short) the set H if

$$\int_{\mathrm{clco}H} g(x)d\mu \geq p > 0, \tag{2.1}$$

where $\mathrm{clco}H$ is the closed convex hull of H.

Assume the density function $g_k(x)$ is "good" w.r.t. the level set H_{c_k} at the k-th iteration. Then, $\Delta c_k = c_k - c_{k+1}$, decrement of the level values generated by (1.5)–(1.6), depends obviously on the effectual sample set determined by $H_k = \{X_i : F(X_i) = f(X_i), i = 1, 2, \ldots, N\}$. For example, suppose the objective function $f(x)$ and the density function $g(x)$ are displayed in Fig. 1. The level set with $c = 0.2$ is

$$H_c = \{x : f(x) < c\} = (1, 2) \cup (3, 3.5).$$

It is obvious that $\mathrm{clco}H_c = [1, 3.5]$. Let $\rho = \int_1^{3.5} g(x)dx$, then the ratio of effectual sample is

$$\alpha = \rho \times \frac{(2-1)+(3.5-3)}{(3.5-1)} = \rho \times \frac{\mu((1,2)\cup(3,3.5))}{\mu([1,3.5])} = \rho \times \frac{\mu(H_c)}{\mu(\mathrm{clco}H_c)}. \tag{2.2}$$

It is worth to notice that, the ratio of effectual sample depends on $\frac{\mu(H_c)}{\mu(\text{clco}H_c)}$, which is a constant associated to the level set H_c.

This constant is related to an important property of H_c. To study this property, Zheng [32] gave some concepts of theoretical robustness. Let X be a topological space and D be a subset of X. By [32], we have

Definition 2.1 [32] A set D is robust if and only if

$$\text{cl}D = \text{cl}(\text{int } D), \tag{2.3}$$

where cl denotes the closure of a set. Obviously, by Definition 2.1 the empty-set $\emptyset$ is robust.

Definition 2.2 [32] A function $f : X \to R$ is robust if and only if its level set $H_c = \{x \in X : f(x) < c\}$ is robust for all $c \in (-\infty, +\infty)$, where X is a robust set.

Definition 2.3 [32] Let X be a normal topological space, Ω be a σ-field of subsets of X. A measure space (X, Ω, μ) is said to be a Q-measure space if the following conditions hold:

M1. each open set is in Ω;
M2. the measure of each nonempty open set is positive;
M3. the measure of each compact set is bounded.

Let (X, Ω, μ) be a Q-measure space. For a given measurable set $S \subset X$, let clco$S$ be the closed convex hull of S. We give the definition of robust constant as follows:

Definition 2.4 Let S be a robust set. The robust constant of S is given by

$$R(S) = \frac{\mu(S)}{\mu(\text{clco}S)}. \tag{2.4}$$

Particularly, set $R(\emptyset) = 1$ and $R(\{x\}) = 1$, where $\{x\} \subset X$ is a singleton.

Definition 2.5 Let $f : X \to R$ be a robust function where X is robust. The robust constant of f w.r.t. the level value c is given by

$$R(f, c) = \frac{\mu(H_c)}{\mu(\text{clco}H_c)}, \tag{2.5}$$

where $H_c = \{x \in X : f(x) < c\}$. By Definition 2.5, it is obvious that $R(f, c) = R(H_c)$.

In what follows, we present some examples for well-understanding of the new concepts.

Example 2.1 Let $X \in R^n$ be a closed convex set, then clco$X = X$, which implies that the robust constant of X is $R(X) = 1$.

Example 2.2 Let $f : R^n \to R$ be a strictly convex and lower-bounded function. Then, for any $c \in R$, $H_c = \{x \in R^n : f(x) < c\}$ is a convex set, which results in $\mu(H_c) = \mu(\text{clco}H_c)$. Thus, we have $R(f, c) = 1$.

Example 2.3 Let $X = [0, 1]$, and $f : X \to R$ be defined by

$$f(x) = \begin{cases} 1, & \text{if } x \text{ is an irrational number,} \\ 0, & \text{if } x \text{ is a rational number.} \end{cases}$$

Then we have

$$R(f, c) = \begin{cases} 1, & \text{if } c \geq 1, \\ 0, & \text{otherwise.} \end{cases}$$

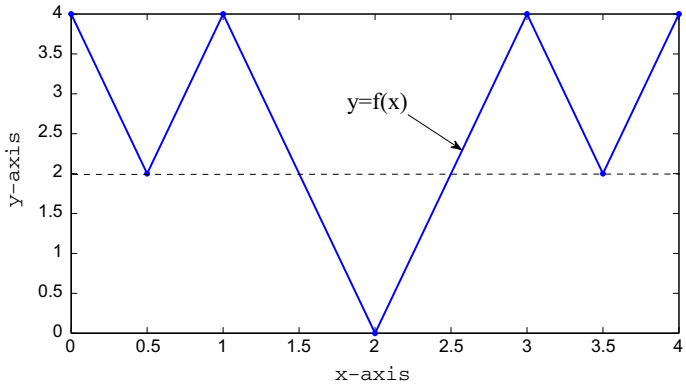

Fig. 2 The function $f : [0, 4] \to R$

Example 2.4 Let $X = [0, 4\pi]$, and $f : X \to R$ be defined by

$$f(x) = \begin{cases} \sin x, & x \in [0, 2\pi), \\ \dfrac{1}{\pi}x - 2, & x \in [2\pi, 4\pi]. \end{cases}$$

Then, by a simple computation we have: $R(f, c) = 1$ for all $c \geq 1$ or $c < 0$, and $R(f, \frac{1}{2}) = \frac{11}{15}$, and so on.

The robust constant of a function depends on the Lebesgue measure of its level set. The level set is difficult to identify, thus robust constant is hard to compute in general. However, it has significance in theory for global optimization, just like Lipschitz constant for convex analysis. Lipschitz constant is also hard to compute, but this does not obstruct the wide applications. This paper does not aim to overcome the drawback of computing robust constant. Alternatively, we focus on the new concept aiming to give a quantitative measure of robustness, and discuss the theoretical significance and applicability.

For global optimization, we are also interested in global robust constant.

Definition 2.6 A function $f : R^n \to R$ is global robust with global robust constant $r > 0$ if

$$R(f, c) \geq r, \quad \forall c \in (-\infty, +\infty). \tag{2.6}$$

Example 2.5 Suppose the function $f : [0, 4] \to R$ is displayed in Fig. 2. It is obvious that

$$R(f, c) \geq \lim_{\epsilon \to 0^+} R(f, 2 + \epsilon) = \frac{1}{3}, \quad \forall c \in (-\infty, \infty).$$

Thus, the global robust constant of the function f is $r = \frac{1}{3}$.

Indeed, the definition of robust constant is consistent to the concept of theoretical robustness presented in literatures. The consistency is shown in the following.

Theorem 2.1 *If $f : X \to R$ is a robust function where X is a robust set, then $R(f, c) > 0$ for all $c \in (-\infty, +\infty)$.*

Proof By Definition 2.2, the function $f : X \to R$ is robust which implies that $H_c = \{x \in X : f(x) < c\}$ is robust for all $c \in (-\infty, +\infty)$. Thus, by Definition 2.1, we have

$$\mathrm{cl}(H_c) = \mathrm{cl}(\mathrm{int}H_c). \tag{2.7}$$

If $H_c = \emptyset$, then $R(f, c) = R(H_c) = R(\emptyset) = 1 > 0$. Otherwise $H_c \neq \emptyset$, by (2.7) we have

$$\mathrm{cl}(\mathrm{int}H_c) = \mathrm{cl}(H_c) \neq \emptyset,$$

which means that $\mathrm{int}H_c \neq \emptyset$, and consequentially by Definition 2.3, $\mu(\mathrm{int}H_c) > 0$. Thus $\mu(H_c) \geq \mu(\mathrm{int}H_c) > 0$ which deduces $R(f, c) > 0$.

By Theorem 2.1, the global robust constant $r > 0$ is a necessary condition for that the corresponding function is robust. The Dirichlet function is not robust thus its global robust constant is zero, see Example 2.3.

3 Applications in analyzing random global search

Robust constant is a useful concept for analyzing random global search technique for global optimization. In this section, we first propose a modified pure adaptive search method which is a representative of random global search methods, then we analyze convergence of the proposed method by using robust constant.

For problem (1.1), it is well-known that: if $H_c = \emptyset$, then $c < c^* = \min f(x)$. Suppose $f(x)$ is robust on R^n, and the scheme (1.5)–(1.6) is used to find the minimal value of f, where X_t $(t = 1, 2, \ldots, N)$ in (1.6) are i.i.d. random samples with density function $g(x)$ (determined by mean u and variance σ^2). Then, the ratio of effectual sample with respect to H_c is given by

$$\rho = R(f, c) \times \int_{\mathrm{clco}H_c} g(x)d\mu. \tag{3.1}$$

The expected number of effectual samples is $N^{\mathrm{elite}} = \rho N$.

Basing on this observation, we propose a modified pure adaptive search method for the problem (1.1), and analyze the proposed method by using robust constant. Via this analysis, one may realize the significance of robust constant.

Algorithm 3.1 A modified pure adaptive search method, MPAS

s0. Let X_0 be a random sample in R^n, $c_0 = f(X_0)$, and let $g_0(x)$ be the probability density function with mean vector u_0 and variance vector σ_0^2. Let $\varepsilon > 0$. Set $k = 0$.

s1. Generate a sample set $S_k = \{X_t, t = 1, 2, \ldots, N_k\}$ i.i.d. from the distribution with density $g_k(x)$. Let

$$c_{k+1} = \frac{1}{N_k} \sum_{t=1}^{N_k} F_k(X_t), \tag{3.2}$$

where $F_k(x)$ is defined by (1.5). Let $\hat{H}_k = \{X_t : F(X_t) = f(X_t) < c_k, t = 1, 2, \ldots, N_k\}$ be the effectual sample set, and $n_k = |\hat{H}_k|$ be the number of elements in the set.

s2. If $|c_{k+1} - c_k| < \varepsilon$, then stop. Otherwise, go to s3.

s3. Let

$$u_{k+1} = \frac{1}{n_k} \sum_{X_t \in \hat{H}_k} X_t, \quad \text{and} \quad \sigma^2_{k+1,i} = \frac{1}{n_k} \sum_{X_t \in \hat{H}_k} \left(X_{t,i} - u_{k+1,i}\right)^2, i = 1, 2, \ldots, n. \tag{3.3}$$

Smooth these parameters by the following scheme

$$u_{k+1} = \alpha u_{k+1} + (1-\alpha)u_k, \quad \sigma_{k+1,i} = \beta_k \sigma_{k+1,i} + (1-\beta_k)\sigma_{k,i}, \quad i = 1, 2, \ldots, n, \tag{3.4}$$

where $0.5 < \alpha < 0.9$, $0.8 < \beta < 0.99$ and $\beta_k = \beta - \beta(1 - \frac{1}{k})^q$, q is an integer (typically between 5 and 10), see Rubinstein et al. [22] and [11].
Construct the new density function $g_{k+1}(x)$ with mean vector u_{k+1} and variance σ^2_{k+1}, and let $k := k + 1$, go to s1.

Remark 3.1 The step s1 accepts the new level value c_{k+1} if it satisfies

$$c_{k+1} - c_k \le \lambda_k (c_k - c_{k-1}), \tag{3.5}$$

where $\lambda_k \in (0, 1)$ is the step length of the k-th iteration.

Remark 3.2 In engineering applications, one can update the sample-size N_k by the following style: at the k-th iteration, generate N_k samples from the distribution with density $g_k(x)$. Denote the sample set by $S_k = \{X_t, t = 1, \ldots, N_k\}$, and compute c_{k+1} by (3.2). If the criterion (3.5) is satisfied, then accept c_{k+1} as the new level value. Otherwise, generate $\lceil \frac{N_k}{10} \rceil$ samples $S'_k = \{X_j, j = N_k + 1, \ldots, N_{k+\lceil \frac{N_k}{10} \rceil}\}$ and let $S_k := S_k \cup S'_k$. Let $N_k = N_k + \lceil \frac{N_k}{10} \rceil$, re-compute c_{k+1} by (3.2) until the condition (3.5) is satisfied.

At the next iteration, we first reserve the effectual samples w.r.t. c_{k+1} in S_k, and denote it by $S^{\text{res}}_k = \{X_t \in H_{c_{k+1}} : X_t \in S_k\}$. Then generate N_{k+1} samples from the distribution with density $g_{k+1}(x)$, denote the sample set by $\widetilde{S}_{k+1}$. Let $S_{k+1} = \widetilde{S}_{k+1} \cup S^{\text{res}}_k$ and $N_{k+1} := N_{k+1} + |S^{\text{res}}_k|$. The rest is to compute c_{k+2} and check whether it satisfies the criterion (3.5) or not, and so on.

To prove convergence of the MPAS method, we have the following theorems.

Theorem 3.1 *Suppose $f : R^n \to R$ is robust with global robust constant $r > 0$, and the sample density functions $g_k(x)$ is "good" w.r.t. H_{c_k}. Then, there exists a positive constant λ_B such that the step length $\lambda_k \ge \lambda_B$ for all k of the MPAS method.*

Proof At the k-th iteration, we have the sample set S_k. We partition S_k into two parts: one is $S^1_k = \left\{X_t \in S_k : F_k(X_t) > c_k - \eta(c_{k-1} - c_k)\right\}$, and the other is $S^2_k = \left\{X_t \in S_k : F_k(X_t) \le c_k - \eta(c_{k-1} - c_k)\right\}$, where $\eta \in (0, 1)$ is a constant. Let $N^1_k = |S^1_k|$ and $N^2_k = |S^2_k|$, then $N_k = N^1_k + N^2_k$. By $F_k(x) \le c_k$, we have

$$\begin{aligned} c_{k+1} &= \frac{1}{N_k} \sum_{X_t \in S_k} F_k(X_t) \\ &= \frac{N^1_k}{N_k} \left(\frac{1}{N^1_k} \sum_{X_t \in S^1_k} F_k(X_t) \right) + \frac{N^2_k}{N_k} \left(\frac{1}{N^2_k} \sum_{X_t \in S^2_k} F_k(X_t) \right) \end{aligned}$$

$$\leq \frac{N_k^1}{N_k}c_k + \frac{N_k^2}{N_k}\Big(c_k - \eta(c_{k-1} - c_k)\Big)$$
$$= c_k - \eta\frac{N_k^2}{N_k}(c_{k-1} - c_k).$$

Let $\lambda_k = \eta\dfrac{N_k^2}{N_k}$, we have

$$c_{k+1} \leq c_k - \lambda_k(c_{k-1} - c_k). \tag{3.6}$$

On the other hand, let $\hat{c}_k = c_k - \eta(c_{k-1} - c_k)$, then by (3.1) we have $N_k^2 = \rho_k N_k$, where

$$\rho_k = R(f, \hat{c}_k) \times \int_{\mathrm{clco}H_{\hat{c}_k}} g_k(x)d\mu.$$

Thus, we obtain

$$\lambda_k = \eta\rho_k = \eta \times R(f, \hat{c}_k) \times \int_{\mathrm{clco}H_{\hat{c}_k}} g_k(x)d\mu. \tag{3.7}$$

Note $f : R^n \to R$ is robust with global robust constant $r > 0$, i.e., $R(f, \hat{c}_k) \geq r$ for all k. The sample density function $g_k(x)$ is "good" w.r.t. H_{c_k}, and $c^* < \hat{c}_k < c_k$ which implies $\mathrm{clco}H_{c_k} \supset \mathrm{clco}H_{\hat{c}_k} \supset H_{\hat{c}_k} \neq \emptyset$. Thus, there is $0 < \hat{p} < p$ such that

$$\int_{\mathrm{clco}H_{\hat{c}_k}} g_k(x)d\mu \geq \hat{p}, \tag{3.8}$$

which deduces

$$\lambda_k \geq \eta r\hat{p} := \lambda_B > 0. \tag{3.9}$$

Theorem 3.2 *Suppose $f : R^n \to R$ and $g_k(x)$ satisfy the assumptions of Theorem 3.1. Then, for the sequence $\{c_k\}$ generated by the MPAS method, we have*

$$\lim_{k\to\infty}\big(c_{k-1} - c_k\big) = 0. \tag{3.10}$$

Proof By (1.5) and (3.2), it is easy to show that $c_{k-1} - c_k \geq 0$ for all k. Adding (3.6) from $k = 1$ to ∞, and let $c_\infty := \lim\limits_{k\to\infty} c_k$, we get

$$c_\infty \leq c_1 - \sum_{k=1}^{\infty}\lambda_k(c_{k-1} - c_k). \tag{3.11}$$

The objective function f is lower bounded implies that there exists a real b such that $f(x) \geq b$ for all $x \in R^n$, thus we have $c_\infty \geq b$. Combining (3.9) with (3.11), there exists $\lambda_B > 0$ such that

$$\lambda_B\sum_{k=1}^{\infty}(c_{k-1} - c_k) \leq \sum_{k=1}^{\infty}\lambda_k(c_{k-1} - c_k) \leq c_1 - c_\infty \leq c_1 - b < +\infty, \tag{3.12}$$

which deduces (3.10) directly.

Intuitively, by S3 of the MPAS method, we have $u_{k+1} \in \mathrm{clco}H_{c_k}$ and $\|\sigma_{k+1}\| \leq \mathrm{diam}(\mathrm{clco}H_{c_k})$, where $\mathrm{diam}(\mathrm{clco}H_{c_k}) = \max_{x,y\in\mathrm{clco}H_{c_k}}\|x - y\|$ is the diameter of $\mathrm{clco}H_{c_k}$. Thus, the density function $g_k(x)$ used in all iterations of the MPAS method is "good" w.r.t. H_{c_k}. However, to prove this assertion is very hard and beyond the interest of this paper, so it is left as an open problem.

Theorem 3.3 *Suppose the objective function $f(x)$ is robust with global robust constant $r > 0$, and the sample density function $g_k(x)$ is "good" w.r.t. H_{c_k} for all $k > 0$. Then, the sequence $\{c_k\}$ generated by the MPAS method converges with probability 1 to global minimal value of problem* (1.1).

Proof The objective function $f(x)$ is lower-bounded on R^n. By (3.2) at S1 of the MPAS method, the sequence $\{c_k\}$ monotonously decreases and $c_k \geq c^*(= \min_{x\in R^n} f(x))$ for all $k > 0$. Hence it converges.

The rest is to prove $\lim\limits_{k\to\infty} c_k = c^*$. By contradiction, assume without loss of generality that

$$\lim_{k\to\infty} c_k = \bar{c}, \quad \text{and} \quad \bar{c} > c^*.$$

Then, for arbitrarily given $\epsilon > 0$, there is a positive integer K, such that $|c_k - \bar{c}| < \epsilon$ holds for all $k > K$, which implies that

$$\bar{c} + \epsilon > c_k > \bar{c} - \epsilon.$$

Let $\epsilon = \frac{1}{2}(\bar{c} - c^*) > 0$, we get $c_k > \frac{1}{2}(\bar{c} + c^*) > c^*$.

On the other hand, since $c_k > c^*$, we have $\mathrm{int}H_{c_k} \neq \emptyset$, and $\mu(H_{c_k}) = \mu(\mathrm{int}H_{c_k}) > 0$ by robustness of $f(x)$. Let $\tilde{c} = \frac{1}{4}(\bar{c} + 3c^*)$, then $c^* < \tilde{c} < c_k$ implies $H_{\tilde{c}} \subset H_{c_k}$ and $\mathrm{int}H_{\tilde{c}} \neq \emptyset$. Thus

$$\mathrm{clco}H_{\tilde{c}} \subset \mathrm{clco}H_{c_k}, \quad \mu(\mathrm{clco}H_{\tilde{c}}) \geq \mu(H_{\tilde{c}}) = \mu(\mathrm{int}H_{\tilde{c}}) > 0. \tag{3.13}$$

At the k-th iteration, a "good" density function $g_k(x)$ (w.r.t. H_{c_k}) is used in sampling. By (3.13) and the same discussion in Theorem 3.1, we have

$$\int_{\mathrm{clco}H_{\tilde{c}}} g_k(x)d\mu \geq \tilde{p} > 0.$$

The ratio of effectual sample with respect to $\tilde{c}$ is

$$\tilde{\rho} = R(f, \tilde{c}) \times \int_{\mathrm{clco}H_{\tilde{c}}} g_k(x)d\mu \geq r\tilde{p},$$

where $r > 0$ is the global robust constant of f by the assumption, i.e., $R(f, \tilde{c}) \geq r$.

We also have, at the k-th iteration, a sample set $S_k = \{X_t : t = 1, 2, \ldots, N_k\}$. Partition S_k to two parts: $S_k^1 = \{X_t \in S_k : F(X_t) \geq \tilde{c}\}$ and $S_k^2 = \{X_t \in S_k : F(X_t) < \tilde{c}\}$. It is obvious that $S_k^2 \subset H_{\tilde{c}}$. Among the N_k samples, with probability 1 there exist $\tilde{\rho}N_k$ samples in S_k^2. By (3.2), we get

$$\begin{aligned} c_{k+1} &= \frac{1}{N_k}\sum_{X_t\in S_k} F(X_t) = \frac{1}{N_k}\left(\sum_{X_t\in H_k^1} F(X_t) + \sum_{X_t\in H_k^2} F(X_t)\right) \\ &= \frac{(1-\tilde{\rho})N_k}{N_k}\left(\frac{1}{(1-\tilde{\rho})N_k}\sum_{X_t\in H_k^1} F(X_t)\right) + \frac{\tilde{\rho}N_k}{N_k}\left(\frac{1}{\tilde{\rho}N_k}\sum_{X_t\in H_k^2} F(X_t)\right) \\ &\leq (1-\tilde{\rho})c_k + \tilde{\rho}\tilde{c}. \end{aligned}$$

By the setting, $c_k > \frac{1}{2}(\bar{c} + c^*)$ and $\tilde{c} = \frac{1}{4}(\bar{c} + 3c^*)$, we have

$$c_k - c_{k+1} \geq (\tilde{\rho} - 1)c_k - \tilde{\rho}\tilde{c} + c_k = \tilde{\rho}(c_k - \tilde{c}) \geq \tilde{\rho}\left[\frac{1}{2}(\bar{c} + c^*) - \frac{1}{4}(\bar{c} + 3c^*)\right]$$
$$= \frac{1}{4}\tilde{\rho}(\bar{c} - c^*) \geq \frac{1}{4}r\tilde{\rho}(\bar{c} - c^*),$$

which yields a contradiction with the stopping criterion $|c_k - c_{k+1}| < \varepsilon$ for arbitrarily $\varepsilon > 0$. In summary, we have $\lim_{k\to\infty} c_k = c^*$ and complete the proof.

Theorem 3.2 gives a simple and utilizable stopping criterion for the MPAS method. Indeed, under the assumption that the optimized function is robust with global robust constant $r > 0$, $|c_{k+1} - c_k| = 0$ provides a necessary condition for global optimality.

4 Numerical results

The computational cost of a random global search method depends heavily on the robust constant of a function to be optimized. To illustrate this, we perform the MPAS method to minimize a set of simple test functions with given global robust constants. We observe the computational effort, measured by the number of function evaluations, while the MPAS method is used to minimize the objective function with different global robust constants. The test functions are given by the following function series: Let $x \in [-3, 3]$, $f_0(x) = |x|$,

$$\begin{cases} f_1(x) = \min\{0.8 \times f_0(x+2) + 0.2,\ f_0(x),\ 0.8 \times f_0(x-2) + 0.2\}, \\ f_2(x) = \min\{0.8 \times f_1(3(x+2)) + 0.2,\ f_1(3x),\ 0.8 \times f_1(3(x-2)) + 0.2\}, \\ \cdots \quad \cdots \\ f_{k+1} = \min\{0.8 \times f_k(3(x+2)) + 0.2,\ f_k(3x),\ 0.8 \times f_k(3(x-2)) + 0.2\}. \end{cases} \tag{4.1}$$

All the functions achieve global minimal value $c^* = 0$ at the unique global minimizer $x^* = 0$.

Denote the global robust constant of $f_k(x)$ by r_k where $k = 0, 1, 2, \cdots$. It is easy to show that $r_0 = 1$, and

$$r_1 = \lim_{\epsilon\to 0^+} R(f_1, 0.2 + \epsilon) = \frac{1}{10}, \ldots, r_{k+1} = \lim_{\epsilon\to 0^+} R(f_{k+1}, 0.2 + \epsilon) = \frac{1}{3}r_k.$$

Thus, $r_{k+1} = \frac{1}{3^k} \cdot \frac{1}{10}$, and $\lim_{k\to\infty} r_k = 0$. We list pictures of some selected test functions in Fig. 3.

In the numerical experiments, we focus on the relationship between computational cost (the number of function evaluations) and the global robust constant of the function to be minimized. In all tests, the MPAS method runs with the same settings: The sample density function is set to $\mathcal{N}(u_k, \sigma_k^2)$, the density of normal distribution, in which mean value u_k and variance σ_k^2 are adaptively updated by S3 of the MPAS method. The initial sample size is fixed to $N_k = 100$ for all instances. The other parameters are list in the following: $\varepsilon = 1.0 \times 10^{-7}$, $\alpha = 0.85, \beta = 0.90, q = 6$. We start the MPAS method from uniformly distributed sampling on $[-3, 3]$, then we get the initial mean value u_0 and variance σ_0^2 from these samples. The MPAS method runs 20 times on each instance. We list the averaged minimal solution (denoted by x^*) and the averaged number of function evaluations (denoted by N_f) in Table 1.

More clearly, the relationship between computational cost and global robust constant is displayed in Fig. 4.

By a simple linear regression, we have

$$\log N_f = -0.4508 \log r + 3.2218.$$

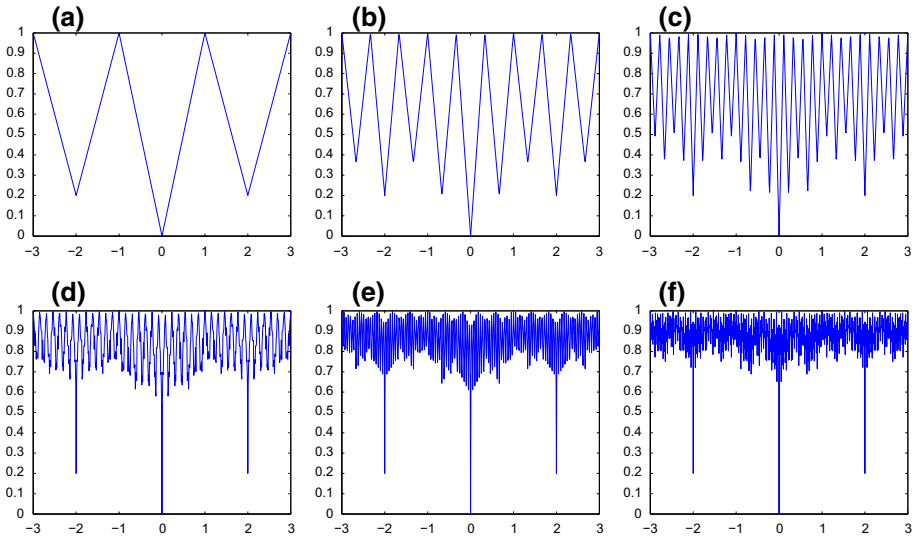

Fig. 3 Pictures of selected test functions, (**a**) Picture of f1(x), (**b**) Picture of f2(x), (**c**) Picture of f3(x), (**d**) Picture of f8(x), (**e**) Picture of f9(x), (**f**) Picture of f10(x)

Table 1 Computational cost and global robust constant

Prob. id.	f0	f1	f2	f3	f4	f5
r_k	1	$\frac{1}{10}$	$\frac{1}{3} \times \frac{1}{10}$	$\frac{1}{3^2} \times \frac{1}{10}$	$\frac{1}{3^3} \times \frac{1}{10}$	$\frac{1}{3^4} \times \frac{1}{10}$
x^*	3.8254e−7	−6.2651e−8	2.1930e−8	−4.8603e−9	−5.0626e−8	−4.5715e−8
N_f	2228	3678	10065	13120	16530	20870
Prob. id.	f6	f7	f8	f9	f10	
r_k	$\frac{1}{3^5} \times \frac{1}{10}$	$\frac{1}{3^6} \times \frac{1}{10}$	$\frac{1}{3^7} \times \frac{1}{10}$	$\frac{1}{3^8} \times \frac{1}{10}$	$\frac{1}{3^9} \times \frac{1}{10}$	
x^*	1.7409e−9	1.8006e−9	2.4229e−9	−2.3364e−9	−7.9819e−9	
N_f	58190	100100	148275	267240	490250	

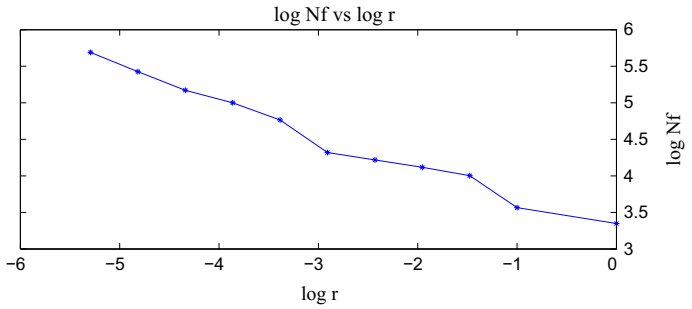

Fig. 4 The relationship between the computational cost and the global robust constant

5 Conclusions

In engineering optimization, random global search techniques are the most useful method for finding a global optimizer. However, many random global search methods have not a suitable stopping criterion. In this paper, we have proposed a new important concept, i.e., robust constant, to quantitatively describe the robustness of measurable sets and functions. We have shown that, global robust constant $r > 0$ is a necessary condition for that the corresponding function (defined on a robust set) is robust. The robust constant is a very useful concept for analyzing random global search methods for unconstrained global optimization.

To show applicability of the robust constant, we have proposed a modified pure adaptive (MPAS) method for unconstrained global optimization, and have analyzed the proposed method (by using global robust constant) from two sides: convergence theory and numerical experiment.

On convergence theory, under the assumption that the objective function is robust with global robust constant $r > 0$, we have proved the decrement of the level value, $|c_k - c_{k+1}|$, converges to zero. This provides a necessary condition for global optimality if the MPAS method is used. Basing on the assertion, we have proved convergence of the MPAS method. Meanwhile, we have given a suitable and checkable criterion to stop the MPAS method.

On numerical experiment, we have constructed a series of test functions with given global robust constants, and minimized these functions by the MPAS method. Numerical result shows that, the computational cost depends heavily on the global robust constant of the minimized function. This observation implies that, when the MPAS method (as a representative of the random global search technique) is used, robust constant is an important parameter of the objective function in unconstrained global optimization.

The MPAS method proposed in this paper can be improved and enhanced via various ways. However, the enhancement and improvement is beyond the main goal of this paper, and we leave it as a task for future research.

References

1. Appel, M.J., Labarre, R., Radulovic, D.: On accelerated random search. SIAM J. Optim. **14**, 708–731 (2003)
2. Baritompa, W., Zhang, B.P., Wood, G.R., Zabinsky, Z.B.: Towards pure adaptive search-a general framework and a one-dimensional realisation. J. Glob. Optim. **7**, 93–110 (1995)
3. De Boer, P.T., Kroese, D.P., Mannor, S., Rubinstein, R.Y.: A tutorial on the cross-entropy method. Ann. Oper. Res. **134**, 19–67 (2005)
4. Brooks, S.H.: A discussion of random methods for seeking maxima. Oper. Res. **6**, 244–251 (1958)
5. Bulger, D.W., Wood, G.R.: Hesitant adaptive search for global optimisation. Math. Progr. **81**, 89–102 (1998)
6. Bulger, D., Baritompa, W., Wood, G.R.: Implementing pure adaptive search with Grover's quantum algorithm. J. Optim. Theory Appl. **116**(3), 517–529 (2003)
7. Chew, S., Zheng, Q.: Integral Global Optimization (Theory, Implementation and Applications). Springer, Berlin (1988)
8. Floudas, C.A., Gounaris, C.E.: A review of recent advances in global optimization. J. Glob. Optim. **45**(1), 3–38 (2009)
9. Hu, J., Fu, M.C., Marcus, S.I.: A model reference adaptive search method for global optimization. Oper. Res. **55**(3), 549–568 (2007)
10. Kabirian, A.: Hybrid probabilistic search methods for simulation optimization. J. Ind. Syst. Eng. **2**(4), 259–270 (2009)
11. Kroese, D.P., Porotsky, S., Rubinstein, R.Y.: The cross-entropy method for continuous multi-extremal optimization. Methodol. Comput. Appl. Probab. **8**, 383–407 (2006)

12. Pardalos, P.M., Romeijn, H.E., Tuy, H.: Recent developments and trends in global optimization. J. Comput. Appl. Math. **124**, 209–228 (2000)
13. Patel, N.R., Smith, R.L., Zabinsky, Z.B.: Pure adaptive search in Monte-Carlo optimization. Math. Progr. **43**, 317–328 (1988)
14. Peng, Z., Wu, D., Zheng, Q.: A level-value estimation method and stochastic implementation for global optimization. J. Optim. Theory Appl. **156**(2), 493–523 (2013)
15. Peng, Z., Shen, Y., Wu, D.H.: A modified integral global optimization method and its asymptotic convergence. Acta Math. Appl. Sin (Engl. Ser.) **25**(2), 283–290 (2009)
16. Pinter, J.: Convergence qualification of adaptive partition algorithms in global optimization. Math. Progr. **56**, 343–360 (1992)
17. Price, C.J., Reale, M., Robertson, B.L.: A cover partitioning method for bound constrained global optimization. Optim. Methods Softw. **27**, 1059–1072 (2012)
18. Price, C.J., Reale, M., Robertson, B.L.: One side cut accelerated random search—A direct search method for bound constrained global optimization. Optim. Lett. **8**, 1137–1148 (2014)
19. Radulovia, D.: Pure random search with exponential rate of convergency. Optimization **59**(2), 289–303 (2010)
20. Rinnooy Kan, A.H.G., Timmer, G.T.: Stochastic global optimization methods part I: clustering methods. Math. Progr. **39**, 27–56 (1987)
21. Rinnooy Kan, A.H.G., Timmer, G.T.: Stochastic global optimizationmethods part II: multi-level methods. Math. Progr. **39**, 57–78 (1987)
22. Rubinstein, R.Y.: The cross-entropy method for combinatorial and continuous optimization. Methodol. Comput. Appl. Probab. **2**, 127–190 (1999)
23. Rubinstein, R.Y., Kroese, D.P.: The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation, and Machine Learning. Springer, New York (2004)
24. Schoen, F.: Stochastic techniques for global optimization: a survey of recent advances. J. Global Optim. **1**(3), 207–228 (1991)
25. Tang, Z.B.: Adaptive partitioned random search to global optimization. IEEE Trans. Autom. Control **11**, 2235–2244 (1994)
26. Thomas, W.: Global optimization algorithms – theory and application. http://www.it-weise.de/
27. Vavasis, S.A.: Complexity issues in global optimization: a survey. In: Horst, R., Pardalos, P.M. (eds.) Handbook of Global Optimization, pp. 27–41. Kluwer, The Netherlands (1995)
28. Wu, D.H., Yu, W.Y., Zheng, Q.: A sufficient and necessary condition for global optimization. Appl. Math. Lett. **23**(1), 17–21 (2010)
29. Yao, Y.R., Chen, L., Zheng, Q.: Optimality condition and algorithm with deviation integral for global optimization. J. Math. Anal. Appl. **357**(2), 371–384 (2009)
30. Zabinsky, Z.B., Smith, R.L.: Pure adaptive search in global optimization. Math. Progr. **53**, 323–338 (1992)
31. Zabinsky, Z.B., Wood, G.R., Steel, M.A., Baritompa, W.P.: Pure adaptive search for finite global optimization. Math. Progr. **69**, 443–448 (1995)
32. Zheng, Q.: Robust analsys and global optimization. Ann. Oper. Res. **24**, 273–286 (1990)
33. Zheng, Q., Zhuang, D.M.: Integral global minimization: algorithms, implementations and numerical tests. J. Glob. Optim. **7**, 421–454 (1995)