Noname manuscript No.
(will be inserted by the editor)

Stopping Rules in k-Adaptive Global Random Search Algorithms

Anatoly Zhigljavsky · Emily Hamilton

Received: date / Accepted: date

Abstract In this paper we develop a methodology for defining stopping rules in a general class of global random search algorithms that are based on the use of statistical procedures. To build these stopping rules we reach a compromise between the expected increase in precision of the statistical procedures and the expected waiting time for this increase in precision to occur.

Keywords Global random search · global optimization · extreme order statistics · waiting times · k-th order statistic

1 Introduction

We consider a general minimization problem $f(x) \to \min_{x\in A}$ with objective function $f(\cdot)$ and feasible region A. We assume that $A \subseteq \mathbb{R}^d$ and $0 < \mathrm{vol}(A) < \infty$. Let x^* be the global minimizer; that is, x^* is a point in A such that $f(x^*) = m$ where $m = \min_{x\in A} f(x)$.

There is an extensive literature on global optimization. Intelligent global optimization techniques try to extract as much information as possible from prior information about the objective function $f(\cdot)$ and the feasible region A and from previously computed values of the objective function. There are as many purely deterministic techniques as purely stochastic. There are also techniques that can be considered as mixtures of both. Among them, one of the first and most influential is the so-called information-statistical technique developed by R.G. Strongin, see [8]. It has been further developed by R.G. Strongin and his coauthors into a powerful methodology of modern global optimization. For a comprehensive description of this methodology and related topics we refer to [9].

In this paper, we restrict ourselves to stochastic methods and consider the following general class of global random search algorithms (see Algorithm 2.2 in [15]).

School of Mathematics, Cardiff University,
Senghennydd Rd, CF24 4AG, Cardiff UK
Tel.: +44(0)2920875076
E-mail: ZhigljavskyAA@cardiff.ac.uk
E-mail: HamiltonEA@cardiff.ac.uk

Algorithm 1.

1. Choose a probability distribution $\mathbb{P}_1$ on A; set the step number to $j = 1$.
2. Obtain N_j points $x_1^{(j)}, \ldots, x_{N_j}^{(j)}$ in A by independently sampling from the distribution $\mathbb{P}_j$.
3. Using the points $x_{l(i)}^i$ $(l(i) = 1, \ldots N_i;\ i = 1, \ldots j)$ and the objective function values at these points, construct a distribution $\mathbb{P}_{j+1}$ on A.
4. Substitute $j + 1$ for j and return to 2.

Algorithm 1 is a convenient general scheme for describing *population-based methods* including the celebrated genetic algorithms. These methods have proved to work well in complex global optimization problems of reasonably high dimension (like $d \cong 10$). As long as the objective function is not too nasty and simulation of random points in A and its subsets is not too difficult, all reasonable rules for choosing the probability distributions $\mathbb{P}_j$ lead to a satisfactory performance of the related algorithms. One of the attractive features of these algorithms is the fact that very often they are reasonably robust with respect to the difficulty of the problem (expressed through f and A) and to the choice of parameters of the algorithms; for more discussion on this topic see Sections 2.1, 2.2 and 3.5 in [15].

There are several *stopping rules* involved in Algorithm 1. There is a global stopping rule which defines how many steps j $(j = 1, 2, \ldots)$ should be run. There are also stopping rules at each step j; these are defined in Algorithm 1 as numbers N_j $(j = 1, 2, \ldots)$. In this paper, we shall be concerned with the problem of choosing the stopping rules N_j $(j = 1, 2, \ldots)$.

We assume that at step j we only use the results of the current step j; the results of previous $j - 1$ steps are only used to construct the distribution $\mathbb{P}_j$. We thus drop the index j and formulate the problem as follows.

Assume we have $x_1, x_2, \ldots$, a sequence of independent identically distributed points in A with distribution $\mathbb{P}$, and the corresponding sequence of values of the objective function at these points: $y_1 = f(x_1), y_2 = f(x_2), \ldots$ After computing n values $y_1, \ldots, y_n$ we construct a confidence interval (c.i.) for $m = \operatorname{ess\,inf} y$ (here y is the random variable with the same distribution as $y_1, y_2, \ldots$). We need to make a decision for choosing between the following two alternatives: (a) carry on computing values $y_{n+1}, y_{n+2}, \ldots$ until the next update of the c.i., and (b) stop the computations and either terminate the algorithm or move to the next step of Algorithm 1 (by updating the distribution $\mathbb{P} = \mathbb{P}_j$).

Random variables $y_i = f(x_i)$, $x_i \sim \mathbb{P}$, have the c.d.f.

$$F(t) = \int_{f(x) \le t} \mathbb{P}(dx)\,. \tag{1}$$

If the distribution $\mathbb{P}$ is such that

$$\mathbb{P}\left(B_\varepsilon(z^*)\right) > 0 \text{ for any } \varepsilon > 0\,, \tag{2}$$

where $B_\varepsilon(z^*) = \{z \in A : \|z - z^*\| \le \varepsilon\}$, then the lower end-point of the distribution with c.d.f. $F(t)$, $m = \operatorname{ess\,inf} y$, is at the same time $m = \min_{x \in A} f(x)$. Otherwise, if the condition (2) is not met, $m = \inf_B f(x)$ may be larger than $\min_{x \in A} f(x)$; here B is the support of the distribution $\mathbb{P}$.

Note that the construction of confidence intervals for m is the key statistical element in the so-called 'branch and probability bound methods', see [13] and [15], pp.83-84.

Additionally, the main statistical tests for testing the hypotheses about $m = \text{ess inf} y$ are based on the c.i. for m defined by (3), see Sect. 7.1.5 in [14].

Of course if the sample size n increases then the length of the c.i. decreases. However, the decrease of this length slows down as n increases and, as we will see later, one needs to wait longer and longer for the next change of the length of the c.i. Reaching a compromise between these two contradictory criteria (keeping n reasonably small and reaching short c.i.) is the main point of discussions in this paper.

Let k be a fixed integer and $y_{1,n} \leq \ldots \leq y_{k,n}$ be order statistics corresponding to the sample $\{y_1, \ldots, y_n\}$. The statistical inference in global random search algorithms are often based only on the lowest k order statistics $y_{1,n}, \ldots, y_{k,n}$ rather than on the whole sample $\{y_1, \ldots, y_n\}$, see [14,15]. Two main reasons for this are: (a) larger order statistics carry very little information about the value of the minimum m, and (b) in this case one can use the asymptotic extreme value theory to develop statistical procedures and study the quality of these procedures.

Note that the choice of k can be an important practical issue. If k is too small (for example, $k = 2$ or 3) then the precision of related statistical procedures can be low while large values of k require much larger values of the sample size n (otherwise the statistical procedures will be heavily biased). For further discussion on the choice of k, see [4] and Sect. 2.4.3 in [15].

We shall always follow the rule of using the lowest k order statistics for choosing values N_j in Algorithm 1. We can thus refer to Algorithm 1 as a *k-adaptive global random search algorithm.* There is a relation of this class of algorithms to the so-called 'pure adaptive search', see [10–12] and especially to the 'pure adaptive search of order k', see Sect. 2.2.4 in [15]. It is assumed in the pure adaptive search algorithm or order k that at each iteration the values of the lowest k order statistics are updated by ignoring the procedure of obtaining these order statistics. In Algorithm 1 these order statistics are obtained by means of the straightforward independent sampling from the given distribution. In particular cases, other techniques reducing the cost of obtaining the values of the lowest k order statistics could perhaps be used.

2 Confidence Intervals for m and Results From the Theory of Order Statistics

2.1 Confidence intervals

We shall use the following c.i. for $m = \text{ess inf}\, y$ based on the order statistics from the sample $\{y_1, \ldots, y_n\}$:

$$I_{k,\delta} = [y_{1,n} - r_{k,\delta}(y_{k,n} - y_{1,n}), y_{1,n}]\,. \tag{3}$$

Here $r_{k,\delta}$ is some constant, $r_{k,\delta} = 1/\left(\left(1-\delta^{1/k}\right)^{1/\alpha} - 1\right)$, and k is fixed. The value of α is defined below (in most cases $\alpha = d/2$, where d is the dimension of A). It is shown in [1], Sect. 7.1.4 in [14] and Sect. 2.4.2 in [15], that the asymptotic confidence level of the c.i. (3) tends to $1-\delta$ as $n \to \infty$. The construction of confidence intervals according to (3) has been suggested in [1]; this construction is the most suitable in the problems of global random search. Indeed, the c.i. (3) is simple and semi-parametric in the sense discussed below; no semi-parametric confidence intervals are known in the literature that lead to shorter confidence intervals than (3). Additionally, as has been

noted already, the main statistical tests for testing the hypotheses about m are based on the c.i. (3).

The length of the c.i. (3) is proportional to $y_{k,n} - y_{1,n}$ and is decreasing as $n \to \infty$. Each time $y_{k,n}$ is updated this length changes. Let n_i index the values of n in which the k-th order statistic changes; that is, the value y_{n_i} is smaller than the previous k-th order statistic: $y_{n_i} < y_{k,n_i-1}$.

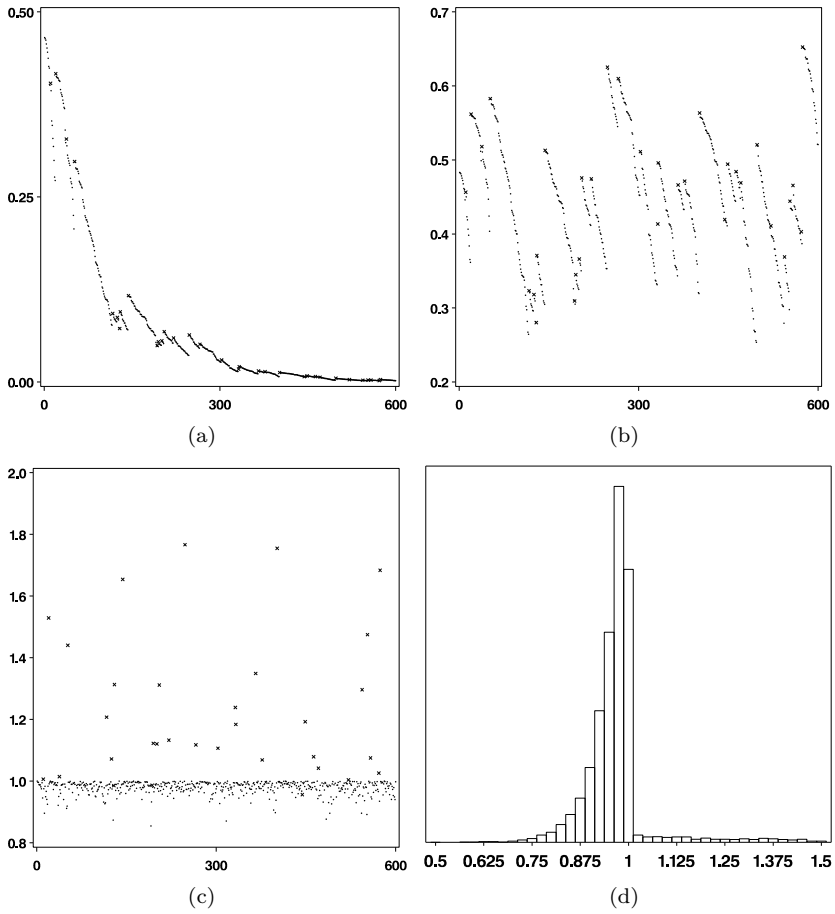


Fig. 1 (a) Typical trajectory of the sequence of $y_{k,n_i} - y_{1,n_i}$ $(i = 1, \ldots, 600)$; $\alpha = 5$, $k = 20$; (b) related sequence of normalized values $(y_{k,n_i} - y_{1,n_i})/(y_{k,n_i} - m)$; (c) related sequence of ratios $R_{k,i}$ defined in (4); (d) the histogram of the ratios plotted in (c).

Figure 1(a) shows a typical sequence of values of $y_{k,n_i} - y_{1,n_i}$ plotted only at the values $n = n_i$ $(i = 1, 2, \ldots)$. It is clearly seen at this figure that $y_{k,n_i} - y_{1,n_i}$ tends to zero as $n_i \to \infty$. To illustrate the rate of this convergence to zero, we normalized these values as $(y_{k,n_i} - y_{1,n_i})/(y_{k,n_i} - m)$ and plotted the related sequence in Figure 1(b). Another natural way of renormalizing the values of $y_{k,n_i} - y_{1,n_i}$ is to consider the

ratios

$$R_{k,i} = \left(y_{k,n_{i+1}} - y_{1,n_{i+1}}\right) / \left(y_{k,n_i} - y_{1,n_i}\right) . \tag{4}$$

The sequence of these ratios computed from the sequence depicted in Figure 1(a) is shown in Figure 1(c). The corresponding histogram is presented in Figure 1(d). This histogram gives an idea about the distribution of the ratios (4). In Section 3 we study this distribution.

The simulations used to produce the plots in Fig. 1 are made for the c.d.f. $F(t) = t^\alpha$, $0 \le t \le 1$. We have made similar simulations for the c.d.f.'s expressed through (1) and therefore arising from the test functions of global optimization, see [4]. The plots are very similar to the ones displayed in Fig. 1.

2.2 Some facts from the asymptotic theory of order statistics

Let $\{y_1, \ldots, y_n\}$ be an independent sample of values of a r.v. y with a c.d.f. $F(\cdot)$ such that $m = \operatorname{ess\,inf} \eta > -\infty$ and let $y_{1,n} \le \ldots \le y_{n,n}$ be the related order statistics. In global random search applications, the c.d.f. $F(\cdot)$ has the form (1).

Consider the asymptotic distribution of the sequence of the (normalized) lowest order statistics $y_{1,n}$, as $n \to \infty$. The following result is classical in the asymptotic theory of order statistics (for proofs, discussions and generalizations, see [2,3,5–7]).

Theorem 1. *Assume that* $\operatorname{ess\,inf} y = m > -\infty$, *where* y *has c.d.f.* $F(t)$, *and the function* $V(v) = F\left(m + \frac{1}{v}\right)$ *regularly varies at infinity with some exponent* $(-\alpha)$, $0 < \alpha < \infty$; *that is,*

$$\lim_{v\to\infty} \left[V(tv)/V(v)\right] = t^{-\alpha}, \qquad \text{for each } t > 0 . \tag{5}$$

Then $\lim_{n\to\infty} F_{1,n}(m + (\kappa_n - m)z) = \Psi_\alpha(z)$, *where* $F_{1,n}$ *is the c.d.f. of the minimum order statistics* $y_{1,n}$,

$$\Psi_\alpha(z) = \begin{cases} 0 & \text{for } z < 0, \\ 1 - \exp\left(-z^\alpha\right) & \text{for } z \ge 0 \end{cases} \tag{6}$$

defines the so-called Weibull distribution and κ_n *is the* $\left(\frac{1}{n}\right)$*-quantile of* $F(\cdot)$*:* $\kappa_n = \inf\{u | F(u) \ge 1/n\}$.

Theorem 1 yields that for a wide class of distributions the distribution of the sequence of random variables $(y_{1,n} - m)/(\kappa_n - m)$ converges (as $n \to \infty$) to the random variable which has the Weibull distribution with c.d.f. (6). This distribution has only one parameter, α, which is called the tail index.

In the case $\operatorname{ess\,inf} y = m > -\infty$, the c.d.f. $\Psi_\alpha(z)$, along with its limiting case $\Psi_\infty(z) = \lim_{\alpha\to\infty} \Psi_\alpha(1 + z/\alpha) = 1 - \exp(-\exp(z))$, $-\infty < z < \infty$, are the only non-degenerate limits of the c.d.f.'s of the sequences $(y_{1,n} - a_n)/b_n$, where $\{a_n\}$ and $\{b_n\}$ are arbitrary sequences of positive numbers.

If there exist numerical sequences $\{a_n\}$ and $\{b_n\}$ such that the c.d.f.'s of $(y_{1,n} - a_n)/b_n$ converge to Ψ_α, then we say that $F(\cdot)$ belongs to the domain of attraction of $\Psi_\alpha(\cdot)$ and express this as $F \in D(\Psi_\alpha)$. The conditions stated in Theorem 1 are necessary and sufficient for $F \in D(\Psi_\alpha)$. There are two conditions: $m = \operatorname{ess\,sup} \eta < \infty$ and the condition (5). The first one is easily met, in particular it is always valid in global random search applications. The condition (5) can be written as $F(t) = c_0(t - m)^\alpha + \mathrm{o}((t - m)^\alpha)$ as $t \downarrow m$, where c_0 is a function of $v = 1/(t - m)$, slowly

varying at infinity as $v \to \infty$. In particular, c_0 may be any positive constant, but the actual range of eligible functions is much wider.

Below we shall need one more important result from the asymptotic theory of order statistics which is actually a generalization of Theorem 1, see e.g. [15], p. 57.

Theorem 2. *If the conditions of Theorem 1 are met and $n \to \infty$, then the asymptotic distribution of*

$$\left(\frac{y_{1,n} - m}{\kappa_n - m}, \ldots, \frac{y_{k,n} - m}{\kappa_n - m} \right)$$

is the same as of the random vector

$$\left(\nu_1^{1/\alpha}, (\nu_1 + \nu_2)^{1/\alpha}, \ldots, (\nu_1 + \ldots + \nu_k)^{1/\alpha} \right) ,$$

where $\nu_1, \ldots, \nu_k$ are i.i.d. r.v. with density e^{-x}, $x \geq 0$.

2.3 The value of the tail index

In problems of global random search the value of the tail index can often be exactly identified, see [15], Sect. 2.5.3. Assume, for simplicity, that the global minimizer x^* is uniquely defined and belongs to the interior of A. Assume also that $f(\cdot)$ is piece-wise continuous in A and twice continuously differentiable in the vicinity of x^*, $\nabla f(x^*) = 0$ and the Hessian $\nabla^2 f(x^*)$ is non-degenerate. Additionally, assume the condition (2) concerning the distribution $\mathbb{P}$. Then the value of the tail index α is uniquely determined and equals $\alpha = d/2$, where d is the dimension of A. This fact and the results quoted in the previous section justify the use of the c.i. (3) in algorithms of global random search.

2.4 Waiting time until the next update

Here we derive the distribution and related moments for waiting times to the next update of the k-th order statistic assuming that we have an i.i.d. random sample $\{y_1, \ldots, y_n\}$ of size n. Note that this distribution only depends on n and k and does not depend on the record number and on the distribution of $\{y_1, \ldots, y_n\}$, as long as this distribution is continuous (which is an assumption below).

Let $y_{n+1}, y_{n+2}, \ldots$ be i.i.d.r.v. with the same distribution as $\{y_1, \ldots, y_n\}$ and let $n' = n + w$ be the first value of y_{n+i} such that $y_{n+i} < y_{k,n}$. Then the r.v. $w = W_{k,n}$ is the waiting time until the next update. Using the standard arguments used in the theory of k-th records (see, for example, [6], Lecture 19) we derive the distribution of $W_{k,n}$.

In order for the next k-th record to occur at time $n + w$ (where $n \geq k$ and $w > 0$), no k-th records may occur at times $n + 1, \ldots, n + w - 1$, and a k-th record must occur at time $n + w$. For any $t \geq k$, $\mathbb{P}(I_{k,t} = 1) = k/t$ and $\mathbb{P}(I_{k,n} = 0) = (t - k)/t$, where $I_{k,t}$ is the indicator of the event that k-th record occurs at time t. Additionally, the random variables I_{k,t_1}, $I_{k,t_2}, \ldots$ are mutually independent for $k \leq t_1 < t_2 < \ldots$. This implies

$$\begin{aligned} \mathbb{P}(W_{k,n} = w) &= \mathbb{P}(I_{k,n+1} = 0, \ldots I_{k,n+w-1} = 0, I_{k,n+w} = 1) \\ &= \frac{k(n + w - k - 1)!n!}{(n + w)!(n - k)!}, \;\; n \geq k, \;\; w > 0 \, . \end{aligned}$$

Using this we can calculate the moments of $W_{k,n}$. In particular,

$$\mathbb{E}(W_{k,n}) = n/(k-1)\,. \tag{7}$$

The expected waiting time for the next update when $k = 1$ is infinite for all $n \geq 1$. Note that the distribution of $W_{k,n}$ (and hence the expectation $\mathbb{E}(W_{k,n})$) does not depend on the fact whether there was an update of the k-th record at time n.

An alternative way of establishing the result (7) would be to use the asymptotic theory of records (see e.g. [7]) according to which the moments of occurrence of k-th records asymptotically (as $n \to \infty$) follow the Poisson process with intensity $k \log n$ (this would imply the result (7) only asymptotically, as $n \to \infty$).

3 A Study of the Distribution of the Ratios (4)

3.1 Asymptotic length of the confidence interval (3)

Below we study the distribution of the ratios (4). Random values from this distribution give the multipliers for the length of the c.i. (3). Indeed, for a sample of size $n \geq k$, this length is proportional to $(y_{k,n} - y_{1,n})$, where the coefficient of proportionality is $r_{k,\delta}$. After N updates of the k-th record $y_{k,n}$, the length of the c.i. (3) is

$$L_N = r_{k,\delta}(y_{k,n_N} - y_{1,n_N}) = r_{k,\delta}\left[\prod_{i=1}^{N} \frac{(y_{k,n_i} - y_{1,n_i})}{(y_{k,n_{i-1}} - y_{1,n_{i-1}})}\right](y_{k,n_0} - y_{1,n_0}) = c\prod_{i=1}^{N} R_{k,i-1},$$

where $n_0 = k$ and $c = r_{k,\delta}(y_{k,n_0} - y_{1,n_0}) = r_{k,\delta}(y_{k,k} - y_{1,k})$. The length L_N tends to 0 as $N \to \infty$, see Figure 1(a), and it is natural to normalize this length by considering the reduction of the length per update, that is, to use $L_N^{1/N}$. Then we can consider the limit

$$\lim_{N\to\infty} L_N^{1/N} = \lim_{N\to\infty}\left[c\prod_{i=1}^{N} R_{k,i}\right]^{1/N} = \lim_{N\to\infty}\left[\prod_{i=1}^{N} R_{k,i}\right]^{1/N}.$$

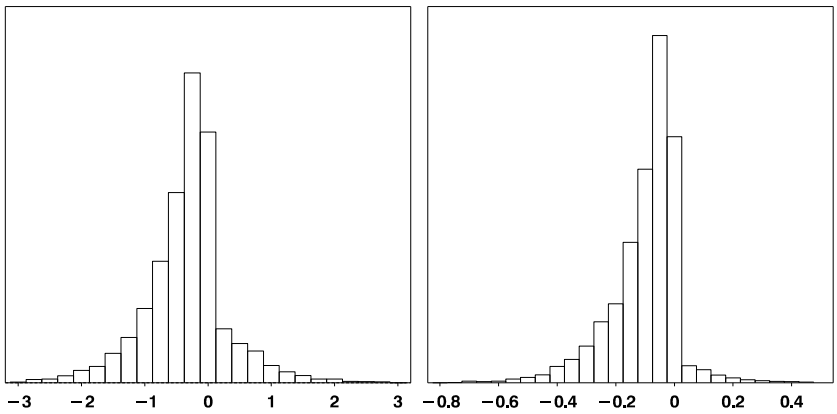
Equivalently, we can consider the asymptotic behaviour (as $N \to \infty$) of

$$\log L_N^{1/N} = \frac{1}{N}\left[\log c + \sum_{i=1}^{N} \log R_{k,i}\right] \longrightarrow \mathbb{E}\log R_k \quad (\text{as } N \to \infty), \tag{8}$$

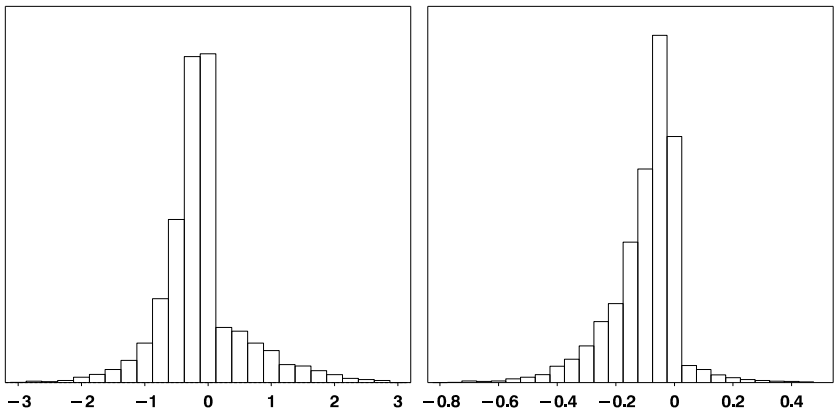
where R_k is a r.v. which has the asymptotic distribution of the ratios (4). The consideration of large N allows us to achieve the following: (i) to get rid of the constant c in (8), (ii) to use the law of large numbers in (8), and (iii) to use the asymptotic representation of Theorem 2 for studying the asymptotic distribution of the ratios (4) and their logarithms.

Taking the exponent of both sides in (8) leads us to the observation that asymptotically (for large N) the value $r_k = \exp\{\mathbb{E}\log R_k\}$ is an average multiplier in the expression for the length L_N. In the next section this value will be considered as the main characteristic of interest.

Note also that the distribution of the logarithms of the ratios (4) behaves better than the distribution of the ratios (4) themselves. This can be seen by comparing Figure 1(d) with histograms of Figure 2.



(a) $\alpha = 1$ $k = 3$ (b) $\alpha = 1$ $k = 10$



(c) $\alpha = 3$ $k = 3$ (d) $\alpha = 3$ $k = 10$

Fig. 2 Histograms of $\log R_{k,i}$ for different k and α; $i = 1, \ldots, 10000$.

3.2 Asymptotic distribution of the ratios (4)

Consider the ratio $R_{k,i} = (y_{k,n_{i+1}} - y_{1,n_{i+1}})/(y_{k,n_i} - y_{1,n_i})$ defined in (4) and assume that i is large (so that we can use the asymptotic representations). Note that $y_{k,n_i} - y_{1,n_i} = y_{k,n} - y_{1,n}$ for all n such that $n_i \le n < n_{i+1}$. We can therefore write the ratio $R_{k,i}$ in the form

$$R_{k,i} = R_k = \left(y'_k - y'_1\right) / \left(y_k - y_1\right), \tag{9}$$

where $y_1 = y_{1,n}$, $y_k = y_{k,n}$, $y'_1 = y_{1,n_{i+1}}$ and $y'_k = y_{k,n_{i+1}}$; here n_{i+1} is the smallest value of $n' > n$ such that the k-th order statistic is changed at time n': $n_{i+1} = \min\{n' > n : y_{n'} < y_{k,n}\}$.

We assume that $k \ge 3$, n is large and we use the representation of Theorem 2 which can be written in the form

$$\frac{y_k - y_1}{\kappa_n - m} \sim (\nu_1 + \eta + \nu_k)^{1/\alpha} - \nu_1^{1/\alpha} \tag{10}$$

where ν_1, ν_k and η are independent random variables, ν_1 and ν_2 have exponential density e^{-x}, $(x \geq 0)$ and η has the Gamma density

$$\varphi_{k-2}(x) = \frac{1}{\Gamma(k-2)} x^{k-3} e^{-x}, \;\; x \geq 0.$$

Our aim now is to write down the ratio (9) in a form similar to (10).

There are three possible cases for how the next update of the order statistics y_1 and y_k is made.

Case 1. The new observation is smaller than y_1; that is, the first order statistic is changed: $y_1' < y_1$; y_k' is therefore equal to $y_{k-1,n}$.

Case 2. The new observation is between $y_{1,n}$ and $y_{k-1,n}$. Then $y_1' = y_1$ and $y_k' = y_{k-1,n}$.

Case 3. The new observation is between $y_{k-1,n}$ and $y_{k,n}$. In this case, $y_1' = y_1$ and y_k' is larger than $y_{k-1,n}$ but smaller than $y_k = y_{k,n}$.

Conditionally on the values of ν_1, η and ν_k, the three cases occur with the following probabilities:

$$\begin{cases} \text{Case 1} \;\; \text{occurs with probability } \frac{\nu_1}{\nu_1+\eta+\nu_k}\,; \\ \text{Case 2} \;\; \text{occurs with probability } \frac{\eta}{\nu_1+\eta+\nu_k}\,; \\ \text{Case 3} \;\; \text{occurs with probability } \frac{\nu_k}{\nu_1+\eta+\nu_k}\,. \end{cases}$$

Consider the following representation for $y_k' - y_1'$ in each of these cases:

$$\frac{y_k' - y_1'}{\kappa_n - m} = \begin{cases} (\nu_1+\eta)^{1/\alpha} - (U\nu_1)^{1/\alpha} & \text{in Case 1}\,; \\ (\nu_1+\eta)^{1/\alpha} - \nu_1^{1/\alpha} & \text{in Case 2}\,; \\ (\nu_1+\eta+U\nu_k)^{1/\alpha} - \nu_1^{1/\alpha} & \text{in Case 3}\,. \end{cases}$$

Here U is a random variable with uniform distribution on $[0,1]$ and independent of ν_1, ν_2 and η.

Therefore the ratio (9) can be expressed in terms of ν_1, ν_k, η and U as follows:

$$R_k = \frac{y_k' - y_1'}{y_k - y_1} = \begin{cases} \frac{(\nu_1+\eta)^{1/\alpha} - (U\nu_1)^{1/\alpha}}{(\nu_1+\eta+\nu_k)^{1/\alpha} - \nu_1^{1/\alpha}} & \text{with probability } \frac{\nu_1}{\nu_1+\eta+\nu_k}\,; \\ \frac{(\nu_1+\eta)^{1/\alpha} - \nu_1^{1/\alpha}}{(\nu_1+\eta+\nu_k)^{1/\alpha} - \nu_1^{1/\alpha}} & \text{with probability } \frac{\eta}{\nu_1+\eta+\nu_k}\,; \\ \frac{(\nu_1+\eta+U\nu_k)^{1/\alpha} - \nu_1^{1/\alpha}}{(\nu_1+\eta+\nu_k)^{1/\alpha} - \nu_1^{1/\alpha}} & \text{with probability } \frac{\nu_k}{\nu_1+\eta+\nu_k}\,. \end{cases} \tag{11}$$

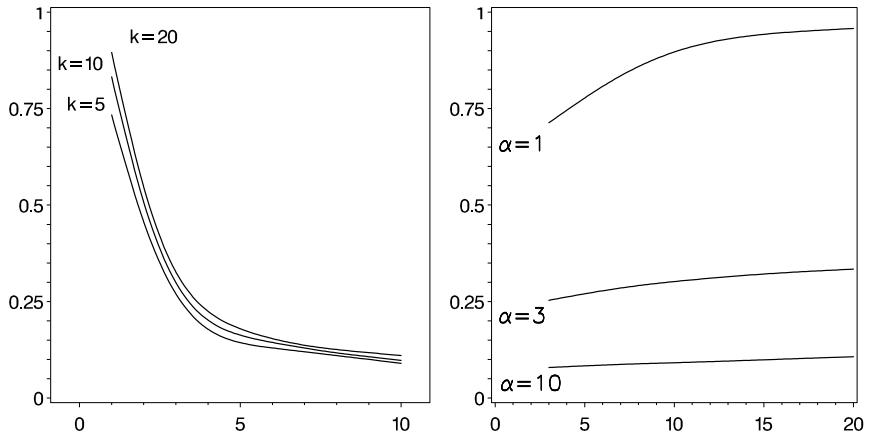
The update of the ratio R_k in Cases 2 and 3 is very similar and always leads to the values $R_k < 1$. Case 1, when y_1 is getting updated, is different. In this case, it is not untypical to obtain values $R_k > 1$. To distinguish the update of R_k in Case 1 and in Cases 2-3, in Figure 1 we marked the updates of R_k in Case 1 with crosses, while the updates of R_k in Cases 2-3 are simple dots.

Computation of moments of the random variable (11), including the most important characteristic $r_k = \exp\{\mathbb{E}\log R_k\}$, can be done either by simulation or by numerical integration: the four-dimensional intergal for computing the distribution of the r.v. (11) can easily be written down. In what follows, the value

$$u_k = -(k-1)\log r_k = -(k-1)\{\mathbb{E}\log R_k\}$$

is more important than r_k itself. Figure 3 displays the values of u_k as functions of α (for $k = 5, 10, 20$) and k (for $\alpha = 1, 3, 10$). The data in this figure have been prepared

on the base of extensive simulations. Additionally, the simulation results have been checked using the numerical integration of the related expressions for $\mathbb{E}\log R_k$ (which is a more difficult way to compute the four-dimensional integrals).



(a) $k = 5, 10, 20$ (b) $\alpha = 1, 3, 10$

Fig. 3 Values of $u_k = -(k-1)\{\mathbb{E}\log R_k\}$ as functions of α (for $k = 5, 10, 20$) and k (for $\alpha = 1, 3, 5$).

3.3 Stopping rules

Consider the k-adaptive global random search algorithms which are written in the form of Algorithm 1 and use the lowest k order statistics for constructing the c.i. (3) for the value of $m = \min_{x\in A} f(x)$. The key characteristics we suggest to consider for defining the stopping rules $N = N_j$ at each step j of Algorithm 1 are:

- n, the number of computations (evaluations of the objective function) already made at the step j;
- the expected waiting time to the next update $\mathbb{E}(W_{k,n}) = n/(k-1)$, see (7); and
- the values $r_k = \exp\{\mathbb{E}\log R_k\}$, see Figure 3 and the explanation of the importance of r_k in Section 2.1.

The values of r_k do not depend on n (assuming n is large enough so that the theory above is applicable) while the expected wait to the next update is proportional to n. This implies that the usefulness of the values $f(x^{(j)}_{n+l})$ $(l = 1, 2, \ldots, o(n))$ is inversely proportional to n. It is worthwhile to remind at this point, see Section 2.4 for details, that the distribution of $W_{k,n}$ is the same whether or not there was an update of the k-th record at time n.

Assume, for example, that we want to achieve $100\beta\%$ of the current length of the c.i. (3) by making further evaluations of the objective function (here we assume $\beta < 1$ and β is not much smaller than 1). Each new k-th record gives us on average $100r_k\%$ decrease of the c.i. Thus on average we need $\tau = \log\beta/\log r_k$ decreases to get

$100\beta\%$ decrease of the current c.i. To get one decrease of the c.i., we need on average $\tilde{n} = n/(k-1)$ further iterations and hence on average we need about

$$\tilde{n}\tau = \frac{n}{k-1}\frac{\log\beta}{\log r_k} = \frac{-n\log\beta}{u_k}$$

iterations to reduce the length of the c.i. (3) to $100\beta\%$ of its current length. Let us now assign cost 1 to each evaluation of the objective function and gain C_β to a decrease of the length of the c.i. (3) from its current value to $100\beta\%$. Then the above considerations lead us to the following definition of the stopping rule:

$$N = \min\left\{n: \;\; n\,\frac{-\log\beta}{u_k} \geq C_\beta\right\} = \left\lceil \frac{C_\beta\, u_k}{-\log\beta}\right\rceil . \tag{12}$$

The stopping rule (12) is defined by the following constants: β, C_β, k and r_k. Let us briefly discuss the values of these constants and the freedom one can have in choosing some of them.

- k: see Section 1 for a discussion on the choice of k. Note that the dependence of u_k (and therefore the stopping rule (12)) on k is very weak, see Fig. 3(b).
- r_k: this constant is uniquely defined through k and α, see Sections 2.1 and 3.2.
- β: there are two constraints: $\beta < 1$ and $\beta \simeq 1$. A natural choice of β is $\beta = r_k$. In this case, the stopping rule (12) is simply $N = \lceil C_\beta(k-1)\rceil$.
- C_β: the interpretation of this constant is obvious. Basically, this is the only constant which is free to choose. Clearly, we have to have $C_\beta >> 1$ if β is not very close to 1.

The dependence of the stopping rule on the original optimization problem is expressed only through α, the tail index of the underlying extreme value distribution. The value of α is often uniquely determined in the problems of global optimization, see Section 2.2. Alternatively, it can be estimated.

Acknowledgements The authors are grateful to the reviewers for helpful suggestions.

References

1. P. Cooke, *Statistical inference for bounds of random variables*, Biometrika **66** (1979), no. 2, 367–374.
2. H. A. David and H. N. Nagaraja, *Order statistics*, third ed., Wiley Series in Probability and Statistics, Wiley, Hoboken, NJ, 2003.
3. J. Galambos, *The asymptotic theory of extreme order statistics*, second ed., Robert E. Krieger Publishing Co. Inc., Melbourne, FL, 1987.
4. E. Hamilton, Savani V., and Zhigljavsky A., *Estimating the minimal value of a function in global random search: comparison of estimation procedures*, Models and algorithms for global optimization, (Eds. A. Törn and J. Žilinskas), Springer, New York, 2007, pp. 193–214.
5. S. Kotz and S. Nadarajah, *Extreme value distributions*, Imperial College Press, London, 2000.
6. V. B. Nevzorov, *Records: mathematical theory*, Translations of Mathematical Monographs, vol. 194, American Mathematical Society, Providence, RI, 2001.
7. S.I. Resnick, *Extreme values, regular variation and point processes*, Springer, New York, 2008.
8. R. G. Strongin, *Numerical methods in multiextremal problems (information-statistical algorithms)*, "Nauka", Moscow, 1978, (in Russian).
9. R.G. Strongin and Y.D. Sergeyev, *Global optimization with non-convex constraints*, Kluwer Academic Publishers, Dordrecht, 2000.

10. G.R. Wood and Z.B. Zabinsky, *Stochastic adaptive search*, Handbook of global optimization, Vol. 2, Nonconvex Optim. Appl., vol. 62, Kluwer Acad. Publ., Dordrecht, 2002, pp. 231–249.
11. Z.B. Zabinsky, *Stochastic adaptive search for global optimization*, Nonconvex Optimization and its Applications, vol. 72, Kluwer Academic Publishers, Boston, MA, 2003.
12. Z.B. Zabinsky and R.L. Smith, *Pure adaptive search in global optimization*, Math. Programming **53** (1992), no. 3, Ser. A, 323–338.
13. A. Zhigljavsky, *Branch and probability bound methods for global optimization*, Informatica **1** (1990), no. 1, 125–140, 190, 201.
14. ———, *Theory of global random search*, Kluwer Academic Publishers, Dordrecht, 1991.
15. A. Zhigljavsky and A. Žilinskas, *Stochastic global optimization*, Springer, New York, 2008.