Stopping Rules in Global Random Search Algorithms

A. Zhigljavsky

Cardiff University, U.K., ZhigljavskyAA@cf.ac.uk

In this talk we describe 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.

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 $m = \min_{x\in A} f(x)$. In this talk, we restrict ourselves with stochastic methods and consider the following general class of global random search algorithms (see Algorithm 2.2 in [1]).

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 independent sampling from distribution $\mathbb{P}_j$.
3. Using the points $x^i_{l(i)}$ $(l(i) = 1, \ldots N_i;\ i = 1, \ldots j)$ 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.

There are several stopping criteria involved in this algorithm. 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 talk, 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 for $m = \mathrm{vrai\ inf}\ y$ (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 confidence interval, and (b) stop the computations completely 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. $F(t) = \int_{f(x)\le t} \mathbb{P}(dx)$. If the distribution $\mathbb{P}$ is such that

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

where $B_\varepsilon(z^*) = \{z \in A : \|z - z^*\| \le \varepsilon\}$, then lower end-point of the distribution with c.d.f. $F(t)$, $m = \mathrm{vrai\ inf}\ y$, is at the same time $m = \min_{x\in A} f(x)$. Otherwise, if this condition 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 $\mathbb{P}$.

Let $y_{1,n} \le \ldots \le y_{k,n}$ 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 smallest k order statistics $y_{1,n}, \ldots, y_{k,n}$ rather than on the whole sample $\{y_1, \ldots, y_n\}$, see [1]. We shall always follow this rule of using the smallest k order statistics for choosing values N_j in Algorithm 1.

Of course if the sample size n increases then the length of the confidence interval for m 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 confidence interval. Reaching a compromise between these two contradictory criteria (keeping n reasonably small and reaching short confidence interval) is the main point of discussions in this talk.

References

[1] A. Zhiglajvsky and A. Zilinskas, *Stochastic Global Optimization.* Springer-Verlag, 2008.