Volume 1, Number 4
October–December 2008

ISSN: 1995-4239

NUMERICAL ANALYSIS AND APPLICATIONS

English Translation of *Sibirskii Zhurnal Vychislitel'noi Matematiki*

http://www.maik.ru
http://www.springerlink.com

PLEIADES PUBLISHING
МАИК "НАУКА/INTERPERIODICA"

Distributed by Springer

ISSN 1995-4239, Numerical Analysis and Applications, 2008, Vol. 1, No. 4, pp. 376–389. © Pleiades Publishing, Ltd., 2008.
Original Russian Text © S.P. Shary, 2008, published in Sibirskii Zhurnal Vychislitel'noi Matematiki, 2008, Vol. 11, No. 4, pp. 457–474.

Randomized Algorithms in Interval Global Optimization

S. P. Shary*

Institute of Computational Technologies, Siberian Branch,
Russian Academy of Sciences, pr. Akad. Lavrent'eva 6, Novosibirsk, 630090 Russia

Received August 23, 2007; in final form, March 7, 2008

Abstract—This paper is a critical survey of the interval optimization methods aimed at computing global optima for multivariable functions. To overcome some drawbacks of traditional deterministic interval techniques, we outline some ways of constructing stochastic (randomized) algorithms in interval global optimization, in particular, those based on the ideas of random search and simulated annealing.

DOI: 10.1134/S1995423908040083

Key words: *global optimization, interval methods, randomization, stochastic methods, random search, simulated annealing.*

1. INTRODUCTION

The subject of this paper is a global optimization problem for a real-valued function $F : \boldsymbol{X} \to \mathbb{R}$ on a rectangular bar $\boldsymbol{X} \subset \mathbb{R}^n$ whose sides are parallel to the coordinate axes, which reads thus:

$$\text{find} \quad \min_{x \in \boldsymbol{X}} F(x). \tag{1}$$

To solve problem (1) with no a priori information on the character of global behavior of an objective function and on the structure of its local extrema, we must use methods providing, in one form or another, search and comparison of all points in domain. Such are, for instance, methods of nonuniform covering [4] or various versions of multistart (see [5]). Among these, also, are various interval techniques for solving problem (1) based on adaptive (in accordance with a branch and bound strategy) partitioning of the domain of the function being optimized, and interval estimation of its range over the subdomains that obtain (see, for instance, [16, 17, 20, 22]). In dealing with optimization problems (1) of small to moderate scale n, such interval methods of global optimization, being deterministic procedures, make it possible to reliably find guaranteed two-sided bounds for both the optimum and corresponding argument values.

For practical optimization of functions of multiple variables, it is reasonable to use probabilistic methods. In modern interval analysis, however, such have not been developed. In this paper, general principles are propounded for constructing stochastic interval global optimization algorithms combining an interval subdomain estimation technique with execution control randomization. In particular, we present interval global optimization algorithms based on the following: (1) random search, and (2) "simulated annealing," a popular method of stochastic optimization, which borrows its name from a physical process in [19]. Preliminarily, some of the present results were announced in [11].

It should be noted that in the most general case, solving the problem of global optimization is a difficult task. The existing numerical methods basically aim at finding and comparing values of all local optima of the objective function. This is not due to the fact that these methods are bad or that modern optimization theory is underdeveloped, etc. In [3], it was proved that even for a polynomial objective function $F(x)$, global optimization problem (1) is NP-hard. This, in fact, is equivalent to admitting that the required cost is not less than an exponential function of the scale n. A recent effective result in this direction is the Kearfott–Kreinovich theorem [18] which says that the solution to global optimization problem (1) beyond the class of convex objective functions is NP-hard.

*E-mail: shary@ict.nsc.ru

2. FUNDAMENTALS OF INTERVAL TECHNIQUE

To make our argument self-contained, we recall some basic definitions and facts on interval analysis which will be used below.

One-dimensional *intervals* are closed sections of the real line $\mathbb{R}$; multidimensional intervals (also called *bars*) are direct products of one-dimensional intervals. A quantity is said to have *interval uncertainty* if its exact value is undefined but nevertheless known to be within some interval. A mathematical discipline that investigates methods to solve problems with interval uncertainty in data is called interval analysis (see, for instance, [6]).

Basic tools for interval analysis are so-called *interval arithmetics*, algebraic systems that formalize operations among intervals. The most popular *classical interval arithmetic* is the algebraic system $\mathbb{IR}$ formed by intervals of the real line $\boldsymbol{x} = [\underline{x}, \overline{x}] \subset \mathbb{R}$ so that arithmetic operations are defined on its representatives, that is, in accordance with the following fundamental principle:

$$\boldsymbol{x} \star \boldsymbol{y} = \{x \star y \mid x \in \boldsymbol{x},\ y \in \boldsymbol{y}\} \text{ for } \star \in \{+, -, \cdot, /\}.$$

In other words, the resulting interval of any operation is a set formed by all possible results of this operation at elements of interval operands. Explicit formulas for interval addition, subtraction, multiplication, and division are the following [2, 7, 16, 17, 20–22]:

$$\begin{aligned}
\boldsymbol{x} + \boldsymbol{y} &= [\underline{x} + \underline{y},\ \overline{x} + \overline{y}], \\
\boldsymbol{x} - \boldsymbol{y} &= [\underline{x} - \overline{y},\ \overline{x} - \underline{y}], \\
\boldsymbol{x} \cdot \boldsymbol{y} &= [\min\{\underline{x}\,\underline{y}, \underline{x}\,\overline{y}, \overline{x}\,\underline{y}, \overline{x}\,\overline{y}\},\ \max\{\underline{x}\,\underline{y}, \underline{x}\,\overline{y}, \overline{x}\,\underline{y}, \overline{x}\,\overline{y}\}], \\
\boldsymbol{x}/\boldsymbol{y} &= \boldsymbol{x} \cdot [1/\overline{y},\ 1/\underline{y}] \text{ for } \boldsymbol{y} \not\ni 0.
\end{aligned}$$

The interval arithmetic operations have the property of being monotone with respect to inclusion, i.e.,

$$\boldsymbol{x} \subseteq \boldsymbol{x}', \boldsymbol{y} \subseteq \boldsymbol{y}' \Rightarrow \boldsymbol{x} \star \boldsymbol{y} \subseteq \boldsymbol{x}' \star \boldsymbol{y}' \text{ for } \star \in \{+, -, \cdot, /\}.$$

For an interval, we can naturally define the following:

midpoint $\operatorname{mid} \boldsymbol{x} = \frac{1}{2}(\overline{x} + \underline{x})$,

width $\operatorname{wid} \boldsymbol{x} = \overline{x} - \underline{x}$,

absolute value $|\boldsymbol{x}| = \max\{|\underline{x}|, |\overline{x}|\}$.

An *interval vector* is a vector, column, or row with interval components. Its geometric image is a rectangular parallelepiped in $\mathbb{R}^n$ whose sides are parallel to the coordinate axes, which (as mentioned) is often called a bar. A sum (difference) of two equidimensional interval vectors is a vector of elementwise sums (differences) of operands. A topology on the space $\mathbb{IR}^n$ of all n-dimensional interval vectors is given by the metric

$$\operatorname{dist}(\boldsymbol{x}, \boldsymbol{y}) = \max\{\|\underline{x} - \underline{y}\|, \|\overline{x} - \overline{y}\|\},$$

where $\underline{x}$ and $\underline{y}$ are vectors of lower ends, $\overline{x}$ and $\overline{y}$ are vectors of upper ends for interval vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{IR}^n$, respectively, and $\|\cdot\|$ is a norm in $\mathbb{R}^n$ (equal to the absolute value for $n = 1$). The midpoint, width, and absolute value operations are applied to interval vectors componentwise. The interval vectors and matrices are assumed to respect a similar (componentwise) inclusion order. Finally, for an arbitrary set $D \subset \mathbb{R}^n$, $\mathbb{I}D$ denotes the set of all interval vectors contained in D, that is,

$$\mathbb{I}D = \{\boldsymbol{x} \in \mathbb{IR}^n \mid \boldsymbol{x} \subseteq D\}.$$

The problem of defining a range of a function on one or another subset in its domain, which is equivalent to an optimization problem, is reduced in interval analysis to calculating an *interval extension of the function.*

Definition 1. Let D be a nonempty subset of a space $\mathbb{R}^n$. An interval function $\boldsymbol{f} : \mathbb{I}D \to \mathbb{IR}^m$ is called an *interval continuation* of a real function $f : D \to \mathbb{R}^m$ if $\boldsymbol{f}(x) = f(x)$ for all $x \in D$.

Definition 2 [7, 17, 20–22]. Let D be a nonempty subset of a space $\mathbb{R}^n$. An interval function $\boldsymbol{f} : \mathbb{I}D \to \mathbb{IR}^m$ is called an *interval extension* of a real function $f : D \to \mathbb{R}^m$ if the following hold:

(1) $\boldsymbol{f}(\boldsymbol{x})$ is an interval continuation of $f(x)$,

(2) $\boldsymbol{f}(\boldsymbol{x})$ is monotone with respect to inclusion, that is, $\boldsymbol{x}' \subseteq \boldsymbol{x}'' \Rightarrow \boldsymbol{f}(\boldsymbol{x}') \subseteq \boldsymbol{f}(\boldsymbol{x}'')$ on $\mathbb{I}D$.

Thus, if $\boldsymbol{f}(\boldsymbol{x})$ is an interval extension of a function $f(x)$, then, for the range of f on a bar $\boldsymbol{X} \subset D$, we obtain the following external estimate (with the help of an enveloping set):

$$\{f(x) \mid x \in \boldsymbol{X}\} = \bigcup_{x \in \boldsymbol{X}} f(x) = \bigcup_{x \in \boldsymbol{X}} \boldsymbol{f}(x) \subseteq \boldsymbol{f}(\boldsymbol{X}).$$

Correspondingly, if $\boldsymbol{F}$ is an interval extension of an objective function $F(x)$ such as in (1), for solving problem (1), we then use

$$\min_{x \in \boldsymbol{X}} F(x) \geq \underline{\boldsymbol{F}(\boldsymbol{X})}.$$

An effective construction of interval extensions of functions is an important problem in interval analysis. Various solutions to this problem are being sought for. Within the framework of our brief review, it is appropriate to present some important results along these lines. Below is a result, which is often called the *main theorem of interval arithmetic.*

Theorem 1 [2, 7, 16, 17, 20, 21]. *If, for a rational function $f(x) = f(x_1, x_2, \ldots, x_n)$ on a bar $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n) \in \mathbb{IR}^n$, the result $\boldsymbol{f}_{\mathrm{nat}}(\boldsymbol{x})$ of substituting variation intervals $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ for its arguments and executing all actions on them in accordance with the rules of interval arithmetic is determined, then*

$$\{f(x) \mid x \in \boldsymbol{x}\} \subseteq \boldsymbol{f}_{\mathrm{nat}}(\boldsymbol{x}),$$

that is, $\boldsymbol{f}_{\mathrm{nat}}(\boldsymbol{x})$ contains the set of values of the function $f(x)$ on $\boldsymbol{x}$.

It is easy to see that, for a rational function $f(x)$, the interval function $\boldsymbol{f}_{\mathrm{nat}}(\boldsymbol{x})$ in the theorem is an interval extension. Such is called a *natural interval extension* and its calculation is a simple matter.

A natural interval extension often gives rather rough estimates of ranges of functions. Therefore, more advanced methods (forms) of finding interval extensions have been developed. A most popular method is the *centered form*

$$\boldsymbol{f}_c(\boldsymbol{x}, \tilde{x}) = f(\tilde{x}) + \sum_{i=1}^{n} \boldsymbol{g}_i(\boldsymbol{x}, \tilde{x})(\boldsymbol{x}_i - \tilde{x}_i),$$

where $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$ is a fixed point, which we call the center, and $\boldsymbol{g}_i(\boldsymbol{x}, \tilde{x})$ are interval extensions of some functions $g_i(x, \tilde{x})$ which are constructed from f and are generally dependent both on $\tilde{x}$ and $\boldsymbol{x}$. In particular, $\boldsymbol{g}_i(\boldsymbol{x}, \tilde{x})$ can be external interval estimates of ranges of derivatives $\partial f(x)/\partial x_i$ on $\boldsymbol{x}$. For more detailed information on the problem of constructing interval extensions of functions, we ask the reader to consult [2, 7, 16, 17, 20, 21].

For our further reasoning, it is important to note that the accuracy of interval estimation with any form of interval extension depends critically on the width of an estimation bar. If $f(\boldsymbol{x})$ denotes an exact range of an objective function on $\boldsymbol{x}$, that is, $f(\boldsymbol{x}) = \{f(x) \mid x \in \boldsymbol{x}\}$, for a natural interval extension of Lipschitz functions, we then have

$$\operatorname{dist}\bigl(\boldsymbol{f}_{\mathrm{nat}}(\boldsymbol{x}), f(\boldsymbol{x})\bigr) \leq C\|\operatorname{wid} \boldsymbol{x}\| \tag{2}$$

with a constant C, and this fact is equivalent to saying that the natural interval extension is first-order accurate. For the centered form, we write

$$\operatorname{dist}\bigl(\boldsymbol{f}_c(\boldsymbol{x}, \tilde{x}), f(\boldsymbol{x})\bigr) \leq 2(\operatorname{wid} \boldsymbol{g}(\boldsymbol{x}, \tilde{x}))^{\mathrm{T}} |\boldsymbol{x} - \tilde{x}|, \tag{3}$$

where $\boldsymbol{g}(\boldsymbol{x}, \tilde{x}) = (\boldsymbol{g}_1(\boldsymbol{x}, \tilde{x}), \boldsymbol{g}_2(\boldsymbol{x}, \tilde{x}), \ldots, \boldsymbol{g}_n(\boldsymbol{x}, \tilde{x}))$. Whenever interval estimates for the functions $\boldsymbol{g}_i(\boldsymbol{x}, \tilde{x})$ are first-order accurate, the centered form is totally second-order accurate, in accordance with (3). Such estimates are derived, for instance, in [21].

3. INTERVAL METHODS OF GLOBAL OPTIMIZATION

Estimates (2) and (3) show that asymptotically, as the size of a bar in domain tends to zero, interval extensions give ever more accurate estimates for ranges of functions. However, if the bar is not small in size, then the interval estimation error for the range can be rather large, for any form of the interval extension of a function. What should be done in this case?

A possible way to meet this difficulty is to use a forced partitioning of a bar in domain so as to decrease its size and, hence, the interval estimation error. In other words, if the initial bar $\boldsymbol{X}$ is divided into two subbars, $\boldsymbol{X}'$ and $\boldsymbol{X}''$, which compose, when combined, the entire $\boldsymbol{X}$, i.e., $\boldsymbol{X}' \cup \boldsymbol{X}'' = \boldsymbol{X}$, then

$$\{F(x) \mid x \in \boldsymbol{X}\} = \{F(x) \mid x \in \boldsymbol{X}'\} \cup \{F(x) \mid x \in \boldsymbol{X}''\}.$$

As a new estimate of the objective function minimum with respect to $\boldsymbol{X}$ we can now take

$$\min\{\underline{\boldsymbol{F}(\boldsymbol{X}')}, \underline{\boldsymbol{F}(\boldsymbol{X}'')}\},$$

which, generally speaking, is more accurate than the initial $\underline{\boldsymbol{F}(\boldsymbol{X})}$, since the bars $\boldsymbol{X}'$ and $\boldsymbol{X}''$ are smaller in size than the initial $\boldsymbol{X}$. In turn, we can divide the descendant bars $\boldsymbol{X}'$ and $\boldsymbol{X}''$ into finer parts, find interval extensions for these, correct the estimate for the minimum, and then again repeat the procedure, etc.

It is reasonable to bring some order into this successive partitioning-estimating process. A lower estimate for $\min\limits_{x \in \boldsymbol{X}} F(x)$ is usually corrected in accordance with a strategy borrowed from the branch and bound method, which enjoys wide application in combinatorial optimization. Specifically, we do the following:

- store all bars $\boldsymbol{Y}$ appearing in partitioning the initial bar $\boldsymbol{X}$ and their estimates $\underline{\boldsymbol{F}(\boldsymbol{Y})}$;
- at each step of the algorithm, divide only one bar $\boldsymbol{Y}$—that providing the lowest estimate $\underline{\boldsymbol{F}(\boldsymbol{Y})}$ for $\min\limits_{x \in \boldsymbol{X}} F(x)$.

Thus, in executing the algorithm, we keep a work list $\mathcal{L}$ consisting of record pairs $\left(\boldsymbol{Y}, \underline{\boldsymbol{F}(\boldsymbol{Y})}\right)$, where $\boldsymbol{Y}$ is an interval bar in $\mathbb{R}^n$ such that $\boldsymbol{Y} \subseteq \boldsymbol{X}$. To simplify processing of the list $\mathcal{L}$, it is convenient to arrange its records so that they are stored in order of increasing $\underline{\boldsymbol{F}(\boldsymbol{Y})}$. At this step of the algorithm, a first record on the work list, the corresponding bar $\boldsymbol{Y}$, and the estimate $\underline{\boldsymbol{F}(\boldsymbol{Y})}$ are referred to as leading.

An important question of organizing our algorithm is partitioning of bars. On the one hand, sizes of leading bars must be guaranteed to tend to zero during its execution. On the other hand, the cost must not be too excessive. A traditional partitioning method is that in a bar subject to division, we divide only one or few widest interval components into equal (or approximately equal) parts; see [16, 17, 22].

As a result, we are led to an algorithm described by the pseudocode in Table 1, where $\leftarrow$ denotes an assignment operator. At the initial stage of this algorithm, the leading bars are concentrated around points providing local minima for the objective function $F(x)$. Then, as these local minima are sufficiently enough corrected (that is, as the leading bars are refined), the algorithm gradually eliminates those minima that are not global. Specifically, every point of a nonglobal local minimum has a neighborhood such that starting with some step, the neighborhood accepts no leading bars. Sooner or later, all leading bars will be concentrated only around values of the argument providing global minima (there may be more than one such minimum). Thereafter, the algorithm finally corrects results, that is, values of these global minima. Such is an ideal scheme for the algorithm operation. In specific problems, however, some of its stages can be omitted. For instance, if the scale n is large and the number of steps in the algorithm (or its operation time) is bounded above, then it can terminate before some individual local minima are corrected. Figure 1 shows a typical configuration of the bar in the domain of an objective function of two variables resulting from the operation of our interval optimization algorithm.

The simple global optimization algorithm presented in Table 1 can hardly be successfully applied to solve serious practical problems. In fact, as the size of a domain that is seemed to have a global minimum decreases, the main step to be taken is halving (bisecting), whose effect becomes ever less pronounced with increasing n. Therefore, it is reasonable to consider the algorithm presented in Table 1 as a "skeleton," a basis for constructing practical optimization methods by making various modifications that would considerably accelerate the convergence. Below is a (not complete) list of such modifications (see, for instance, [4, 16, 17, 22]):

Table 1. Simple interval adaptive algorithm for global optimization of functions

Input

Bar $\boldsymbol{X} \subset \mathbb{R}^n$. Given accuracy $\epsilon > 0$.

Interval extension $\boldsymbol{F} : \mathbb{I}\boldsymbol{X} \to \mathbb{IR}$ of objective function F.

Output

Lower estimate of global minimum F^* of function F on $\boldsymbol{X}$.

Algorithm

$\boldsymbol{Y} \leftarrow \boldsymbol{X}$;

calculate $\boldsymbol{F}(\boldsymbol{Y})$ and initialize list $\mathcal{L}$ by record $\{(\boldsymbol{Y}, \underline{\boldsymbol{F}(\boldsymbol{Y})})\}$;

DO WHILE $(\text{wid}\,(\boldsymbol{F}(\boldsymbol{Y})) \geq \epsilon)$

select component l in which bar $\boldsymbol{Y}$ has largest length, that is, $\text{wid}\,\boldsymbol{Y}_l = \max\limits_i \text{wid}\,\boldsymbol{Y}_i$;

halve $\boldsymbol{Y}$ along lth coordinate into bars $\boldsymbol{Y}'$ and $\boldsymbol{Y}''$ such that:

$$\boldsymbol{Y}' = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{l-1}, [\underline{\boldsymbol{Y}}_l, \text{mid}\,\boldsymbol{Y}_l], \boldsymbol{Y}_{l+1}, \ldots, \boldsymbol{Y}_n),$$

$$\boldsymbol{Y}'' = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{l-1}, [\text{mid}\,\boldsymbol{Y}_l, \overline{\boldsymbol{Y}}_l], \boldsymbol{Y}_{l+1}, \ldots, \boldsymbol{Y}_n);$$

calculate interval estimates $\boldsymbol{F}(\boldsymbol{Y}')$ and $\boldsymbol{F}(\boldsymbol{Y}'')$;

remove record $(\boldsymbol{Y}, \underline{\boldsymbol{F}(\boldsymbol{Y})})$ from work list $\mathcal{L}$;

enter records $(\boldsymbol{Y}', \underline{\boldsymbol{F}(\boldsymbol{Y}')})$ and $(\boldsymbol{Y}'', \underline{\boldsymbol{F}(\boldsymbol{Y}'')})$ into $\mathcal{L}$ in order of increasing second field;

denote lead record in $\mathcal{L}$ by $(\boldsymbol{Y}, \underline{\boldsymbol{F}(\boldsymbol{Y})})$;

END DO

$F^* \leftarrow \underline{\boldsymbol{F}(\boldsymbol{Y})}$;

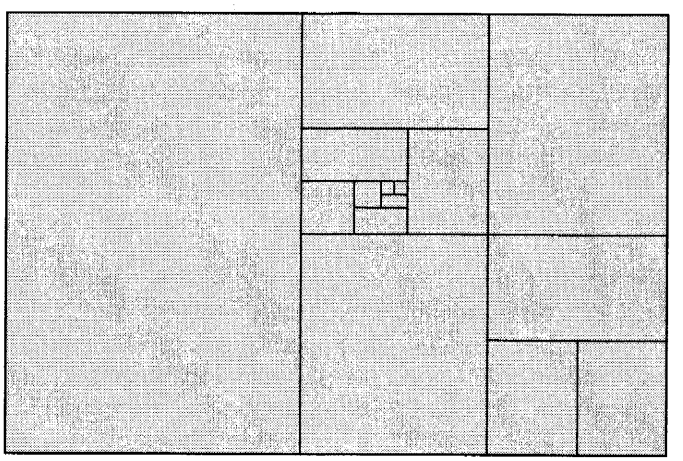

Fig. 1. Typical configuration of the bar in the domain of a function of two variables resulting from its partitioning by the algorithm in Table 1.

- monotonicity of the objective function $F(x)$ with respect to some variables on bars $\boldsymbol{Y}$ in the list $\mathcal{L}$ is used to decrease the dimensions of the bars;

- construction of a more qualitative interval extension for $F(x)$;
- minimization procedures that are more effective than bisection (for instance, gradient descent methods in those bars $\boldsymbol{Y}$ in which $F(x)$ is smooth and convex) are used in relevant bars based on specific local properties of the objective function;
- along with estimation of the objective function F in entire bars, its values at some points of the bars are calculated so as to produce an upper bound for the sought-for global minimum, which makes it possible to eliminate those records in $\mathcal{L}$ that a priori cannot be leading.

Interval global optimization methods, similar to the algorithm described above, evolved considerably in the last decades of the past century. Using these, a lot of difficult practical problems were solved (see, for instance, [13]) and a corresponding theory was entered into all encyclopedias and reference books on optimization (see, in particular, [15]). Summing up, we can say that the interval global optimization methods described above are efficient for problems in small to moderate dimensions (when n does not exceed several tens) and for good enough objective functions.

At the same time, the exploitation of interval global optimization methods has also revealed some problems. If the dimension of the problem under study is high, and/or the objective function has many local extrema, then the leading bars on the work list cannot decrease considerably in size within a reasonable time. The sizes remain comparable with that of the initial bar in domain. This causes a large total error in the interval algorithm.

The above-mentioned difficulty is often aggravated by the phenomenon of "stagnation" in interval estimation (see Fig. 2). The point is that, despite the presence of upper estimates (2) and (3), the error of interval estimation of the objective function can behave in a very complicated way. In particular, for large-sized bars, it may not decrease monotonically in the same way as does the bar size, and if such a bar becomes leading, then detecting this event may lead to large time losses.

Characteristic features of the interval optimization methods considered above are the following:

(1) provability (or guarantability) of results: the estimate obtained for values of a global minimum is guaranteed to estimate the global minimum from below (it is also not difficult to estimate an optimum from below and above simultaneously);

(2) purely deterministic character of the algorithm: each step is unambiguously determined by results of previous steps and by properties of an objective function.

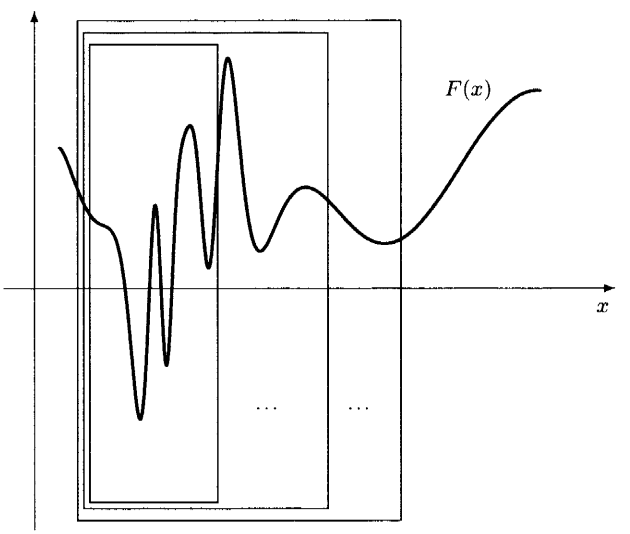

Fig. 2. The phenomenon of stagnation in interval estimation: as the bar width decreases considerably (along the x-axis), the quality of interval estimation of the range of the objective function (bar height) improves only slightly.

The provability of results is ensured by properties of interval extensions of functions, by the estimates obtained via these, and by the computational scheme of the algorithms based on the pseudocode in Table 1. Thus, both of the features indicated above are closely related to each other.

The provability (guarantability) of results of modern interval methods for global optimization is of course a very important property, which is rare in noninterval methods. On the other hand, provability is achieved at a significant sacrifice, say, in computational efficiency, which is much worse for interval methods than for their classical noninterval counterparts.

There arises the question as to whether a loss in provability and/or deterministic character of interval methods will be a sacrifice to pay for gaining in their computational efficiency or some other useful properties. It is this question that stimulated the present work. Below we will try to cautiously justify a positive answer.

We can outline several principal ways of implementing the idea that we have dubbed.

First, to estimate ranges of the objective function $F(x)$ by subsets of the domain $\boldsymbol{X}$, instead of the real interval extensions proper, we can well do with unguaranteed interval estimates having the property of being asymptotically accurate: an error of the estimate derived via them tends to zero as the corresponding bar decreases in size. Of course, the interval estimates that obtain will loose in provability, but this can be compensated for by a decrease in cost in calculating interval estimates $\boldsymbol{F}(\boldsymbol{Y})$. We can show, for instance, that when the interval extension $\boldsymbol{F}$ is replaced by an asymptotically accurate interval continuation in the algorithm depicted in Table 1, the resulting estimate F^* will no longer approximate the global minimum from below, but the consistency of this estimate, i.e., the property of tending to the sought-for global minimum, will be preserved.

Second, in constructing interval optimization algorithms, we can do away with a purely deterministic computational scheme and randomize the algorithm by letting some stochastic transitions in. This approach will be considered in detail below.

4. STOCHASTIC OPTIMIZATION METHODS

Stochastic optimization methods for functions is a large and intensively evolving field of knowledge with specific approaches and values (see, for instance, [5, 10, 15] and references therein). An elementary stochastic optimization algorithm is a passive random search based on the idea of a point being repeatedly thrown so as to be randomly scattered over domains of an objective function. For purely didactic reasons, here, we consider this algorithm as a model example and prototype for an elementary stochastic interval optimization method.

Another popular probabilistic optimization algorithm is simulated annealing, which models the corresponding physical process and is propounded in updated form in [19]. Russian scientists call it *simulated annealing*, *tempering procedure* [1], or *simulated hardening* [5]. A pseudocode for this algorithm is presented in Table 2. Theoretical results on its convergence, as well as other references, can be found, for instance, in [1, 5, 12].

The algorithm has a nonnegative real parameter T, called "temperature," which is similar to the physical temperature in real annealing. During the operation of the algorithm, the quantity T gradually decreases from an initial value T_0 to a final value T_{fin}. For definiteness, we consider the sequence of T-values as a geometrical progression with a denominator α, $0 < \alpha < 1$, although in the general case it is, strictly speaking, not very easy to select a strategy for decreasing T so as to provide convergence to a global optimum [1].

The rule $\mathcal{S}(y)$ for selecting a new point z is typically random throwing with a probability density symmetric with respect to y. The probability $\mathsf{P}_T(y, z)$ of taking the new approximation z figuring in the algorithm is assumed to be

$$\mathsf{P}_T(y,z) = \begin{cases} 1 & \text{if } \Delta F \le 0, \\ \exp\left(-\dfrac{\Delta F}{kT}\right) & \text{if } \Delta F \ge 0, \end{cases}$$

where ΔF denotes the function increment to the old approximation, that is, $\Delta F = F(z) - F(y)$. The function $\mathsf{P}_T(y, z)$ versus ΔF for various temperatures T is shown in Fig. 3. Here the constant k in the

Table 2. Pseudocode for the simulated annealing algorithm

Input
Bar $\boldsymbol{X} \subset \mathbb{R}^n$. Objective function $F : \boldsymbol{X} \to \mathbb{R}$.
Initial T_0 and final T_{fin} values of temperature.
Output
Estimate F^* of global minimum of function F on $\boldsymbol{X}$.
Algorithm
Select initial approximation $y = x_0 \in \boldsymbol{X}$ and set $T \leftarrow T_0$; set number N_T of tests per temperature level; DO WHILE ($T > T_{\text{fin}}$) DO FOR $j = 1$ TO N_T randomly choose new point $z \in \boldsymbol{X}$ by rule $\mathcal{S}(y)$; take z with probability $\mathsf{P}_T(y, z)$ setting $y \leftarrow z$; END DO decrease temperature $T \leftarrow \alpha T$ END DO $F^* \leftarrow F(y)$;

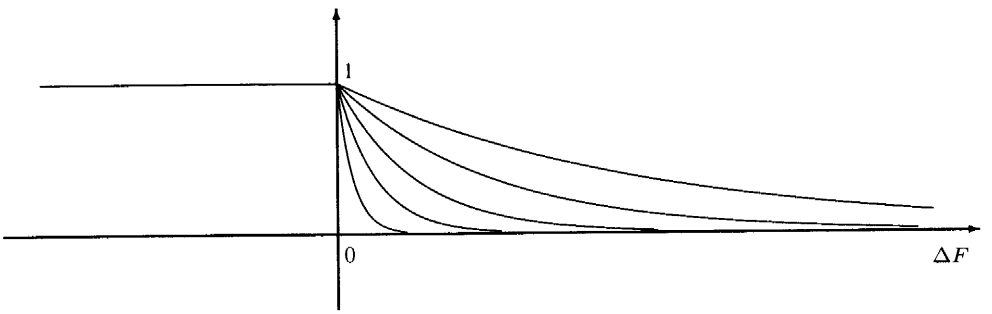

Fig. 3. Functions $\mathsf{P}_T(y, z)$ for various temperatures T.

denominator of the exponent argument is used for scaling and plays the role of a Boltzmann constant in physical formulas describing the real annealing process (see [1, 19]).

In the simulated annealing algorithm, a new approximation is accepted unconditionally if it provides a better value for the objective function. If, however, the new approximation is not better than the old, there still exists a nonzero (temperature-dependent) probability for its being accepted. As a result, simulated annealing exercises random "walks" about the domain to get out of valleys around local minima and thereby provide a global search for a solution to problem (1). As the temperature T decreases, the amplitude of random walks in simulated annealing also decreases and the algorithm in Table 2 turns into a random local search.

It should be noted that a similar problem—how to get out of wide "plateau" regions—is encountered in interval global optimization methods of the type presented in Table 1 when, due to stagnation of interval estimation within many steps of the algorithm, the leading estimate is achieved at some bars that do not contain the sought-for optimum. Can the main idea of simulated annealing be used to randomize interval global optimization algorithms and solve the problem of stagnation for interval estimation?

5. STOCHASTIC INTERVAL METHODS

When constructing stochastic interval optimization procedures, an interval extension of the objective function is assumed to be known: information about the function being minimized will be taken from this interval extension, not from values at individual points. A statistical test will be a random choice of a subbar in domain, followed by estimation, on the subbar, of the range of an objective function by an interval extension.

To provide a global search for an optimum, the subbars chosen randomly in the domains $\boldsymbol{X}$ must cover the entire initial bar $\boldsymbol{X}$. On the other hand, the subbars of nonzero intersection to estimate the objective function are undesirable in a cost effective algorithm. Finally, since the interval extension of the function gives a more accurate estimate of the range with decreasing estimation bar sizes, the sizes of the randomly chosen subbars in the domain of the objective function should be made as small as possible.

As a compromise between the requirements formulated above, it is reasonable to use some data handling and data organizing methods of the traditional interval global optimization algorithms described above, in Sec. 2. Specifically,

- all subbars being estimated will be obtained by successively partitioning the initial bar of the domain $\boldsymbol{X}$, and
- any subbar being estimated which is obtained by partitioning must be stored during the operation of the algorithm unless it has been found unproductive by some special methods.

In other words, we will in fact work with a discrete configuration on the initial bar of the domain $\boldsymbol{X}$ (shown in Fig. 1). Thus, the ideas of discrete stochastic optimization will be of special value for further developing our conception, that of optimizing a function by a suitable division of domain simultaneously with interval estimation.

An algorithm whose pseudocode is presented in Table 3 is a possible analog of the simplest optimization method of random search. It will be called *random interval partitioning*, because each trial of the method consists in halving a bar randomly chosen from those that cover the domain $\boldsymbol{X}$. All bars that obtain during the operation of the algorithm are stored in the work list $\mathcal{L}$ (there is no point in ordering the list now); the bar $\boldsymbol{Y}$ that provides the record estimate of an objective function minimum is also stored.

In the absence of a priori information about the objective function $F(x)$, it would be natural to randomly choose a bar $\boldsymbol{Z}$ in $\mathcal{L}$ that will be partitioned next in accordance with an equiprobable rule, that is, assuming that all bars have equal probabilities for being chosen. Such a choice, however, is not good, because the algorithm thus obtained turns out to be passive, in the sense that every step does not use any information produced at previous steps. In addition, as such an algorithm runs and the work list $\mathcal{L}$ increases in length, the probability that bars containing a global minimum are chosen will but decrease. To eliminate this shortcoming, it makes sense to introduce into the algorithm some procedures to remove unproductive bars from $\mathcal{L}$. Moreover, N.V. Panov proposed to dynamically assign higher probabilities for being chosen in $\mathcal{L}$ to bars providing lower estimates for the objective function. This modification of the algorithm in Table 3 is called random interval partitioning with priority.

The computational efficiency of random interval partitioning is clearly not high (this was confirmed by computer experiments performed by Panov). However, the method can be used in some practical situations (for justification of efficiency of the traditional random search, see, for instance, [5, 10]). Theoretical analysis of this algorithm presents no difficulties. Finally, on the basis of random interval partitioning, we can construct more advanced algorithms implementing some schemes of adaptation to the objective function.

A computational scheme for an interval analog of simulated annealing is shown in Table 4. Here the probability of partitioning a chosen bar $\boldsymbol{Z}$, which is similar to that of taking the next approximation in the classical simulated annealing method, is given by the formula

$$\mathsf{P}_T(\boldsymbol{Y},\boldsymbol{Z}) = \begin{cases} 1 & \text{if } \Delta\boldsymbol{F} \le 0, \\ \exp\left(-\dfrac{\Delta\boldsymbol{F}}{kT}\right) & \text{if } \Delta\boldsymbol{F} \ge 0, \end{cases}$$

Table 3. Pseudocode for the random interval partitioning method

Input
Bar $\boldsymbol{X} \subset \mathbb{R}^n$. Number N of trials.
Interval extension of objective function $\boldsymbol{F} : \mathbb{I}\boldsymbol{X} \to \mathbb{IR}$.
Output
Estimate F^* of global minimum of function F on $\boldsymbol{X}$.
Algorithm

$\boldsymbol{Y} \leftarrow \boldsymbol{X}$;
calculate $\boldsymbol{F}(\boldsymbol{Y})$ and set $F^* \leftarrow \underline{\boldsymbol{F}(\boldsymbol{Y})}$;
initialize work list $\mathcal{L}$ by record $\{\boldsymbol{Y}\}$;
DO FOR $j = 1$ TO N
    randomly choose bar $\boldsymbol{Z}$ in list $\mathcal{L}$;
    halve bar $\boldsymbol{Z}$ in longest component into descendant bars $\boldsymbol{Z}'$ and $\boldsymbol{Z}''$;
    remove bar $\boldsymbol{Z}$ from $\mathcal{L}$, enter bars $\boldsymbol{Z}'$ and $\boldsymbol{Z}''$ into $\mathcal{L}$;
    calculate estimates $\phi' \leftarrow \underline{\boldsymbol{F}(\boldsymbol{Z}')}$ and $\phi'' \leftarrow \underline{\boldsymbol{F}(\boldsymbol{Z}'')}$;
    IF ($\boldsymbol{Z} = \boldsymbol{Y}$) THEN
        set $F^* \leftarrow \min\{\phi', \phi''\}$;
    ELSE
        set $F^* \leftarrow \min\{ F^*, \phi', \phi''\}$
    END IF
    if F^* has changed, then denote by $\boldsymbol{Y}$ that bar $\boldsymbol{Z}'$ or $\boldsymbol{Z}''$ which provides new estimate F^*;
END DO

where $\Delta\boldsymbol{F} = \underline{\boldsymbol{F}(\boldsymbol{Z})} - \underline{\boldsymbol{F}(\boldsymbol{Y})}$ is the optimum estimate increment provided by the new approximation bar. A family of graphs of this function for various temperatures T is shown in Fig. 4.

The rule $\mathcal{S}(\boldsymbol{Y})$ in the algorithm given in Table 4 is randomly choosing a bar $\boldsymbol{Z}$ in the work list. The rule depends on the leading bar $\boldsymbol{Y}$ (and on the leading estimate) and can be organized in different ways depending on the available a priori information about the objective function.

It should be noted that, along with the procedure in Table 4, interval simulated annealing can be presented by a guaranteed version, in which the property that the estimate being calculated does not to exceed a global optimum is preserved during the execution of the algorithm. Why is it necessary to introduce stochastic elements into this computational scheme? The answer is that in so doing the scheme is randomized by providing, on average, a more uniform distribution of computational efforts at early stages of the algorithm operation. This partially mitigates the effect of stagnation in the interval estimation.

In Fig. 4 showing functions $\mathsf{P}_T(\boldsymbol{Y}, \boldsymbol{Z})$, the domain of negative $\Delta\boldsymbol{F}$ is shaded to show that it is inaccessible for the guaranteed version of interval simulated annealing. Actually, if $\boldsymbol{Y}$ is a bar that is leading at a given step of the algorithm, then $\underline{\boldsymbol{F}(\boldsymbol{Y})}$ is not greater than the estimate provided by any other bar on the list, and $\Delta\boldsymbol{F}$ is always nonnegative.

At more advanced stages of execution of the guaranteed version of interval simulated annealing, when the temperature T substantially decreases, the algorithm in Table 5 turns into a conventional interval global optimization algorithm, such as in Table 1.

First computational experiments with interval versions of the simulated annealing method proposed for small-scale problems in [8] showed that these versions are much worse in performance than

Table 4. Simple interval algorithm of simulated annealing

Input

Bar $\boldsymbol{X} \subset \mathbb{R}^n$. Initial T_0 and final T_{fin} temperature values.

Interval extension $\boldsymbol{F} : \mathbb{I}\boldsymbol{X} \to \mathbb{IR}$ of objective function F.

Output

Estimate F^* of global minimum of function F on $\boldsymbol{X}$.

Algorithm

set $\boldsymbol{Y} \leftarrow \boldsymbol{X}$ and $T \leftarrow T_0$;
set integer quantity N_T—number of trials per temperature level;
calculate $\boldsymbol{F}(\boldsymbol{Y})$ and initialize list $\mathcal{L}$ by record $\{(\boldsymbol{Y}, \underline{\boldsymbol{F}(\boldsymbol{Y})})\}$;
DO WHILE $(T > T_{\text{fin}})$
  DO FOR $j = 1$ TO N_T
    randomly choose record $(\boldsymbol{Z}, \underline{\boldsymbol{F}(\boldsymbol{Z})})$ in $\mathcal{L}$ in accordance with rule $\delta(\boldsymbol{Y})$;
    DO (with probability $\mathsf{P}_T(\boldsymbol{Y}, \boldsymbol{Z})$)
      halve bar $\boldsymbol{Z}$ in longest component into descendant bars $\boldsymbol{Z}'$ and $\boldsymbol{Z}''$;
      calculate $\boldsymbol{F}(\boldsymbol{Z}')$ and $\boldsymbol{F}(\boldsymbol{Z}'')$;
      remove record $(\boldsymbol{Z}, \underline{\boldsymbol{F}(\boldsymbol{Z})})$ from $\mathcal{L}$;
      enter records $(\boldsymbol{Z}', \underline{\boldsymbol{F}(\boldsymbol{Z}')})$ and $(\boldsymbol{Z}'', \underline{\boldsymbol{F}(\boldsymbol{Z}'')})$ into $\mathcal{L}$;
      denote by $(\boldsymbol{Y}, \underline{\boldsymbol{F}(\boldsymbol{Y})})$ that record $(\boldsymbol{Z}', \underline{\boldsymbol{F}(\boldsymbol{Z}')})$ or $(\boldsymbol{Z}'', \underline{\boldsymbol{F}(\boldsymbol{Z}'')})$ that has least second field;
    END DO
  END DO
  decrease temperature value $T \leftarrow \alpha T$
END DO
$F^* \leftarrow \underline{\boldsymbol{F}(\boldsymbol{Y})}$;

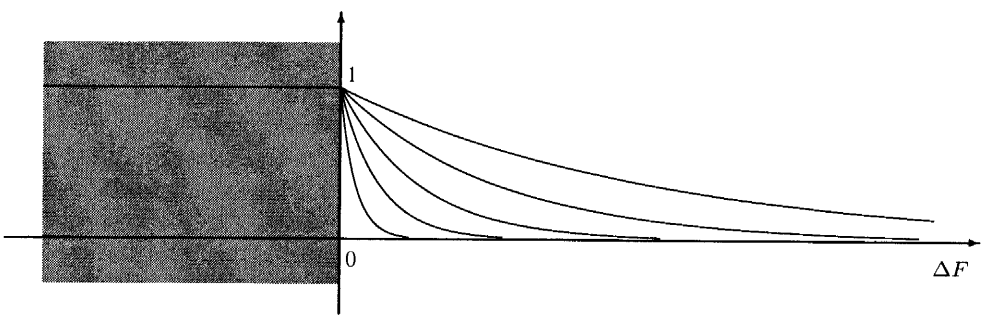

Fig. 4. Functions $\mathsf{P}_T(\boldsymbol{Y}, \boldsymbol{Z})$ for various temperatures T.

the deterministic algorithms of interval global optimization if the objective function $F(x)$ is not very complicated in structure. However, if $F(x)$ has many local extrema and is complex in shape, the advantages of deterministic optimization methods become small or negligible. In particular, in some test problems described in [9], the convergence rate of interval simulated annealing turned out to be higher than that in traditional interval global optimization methods based on adaptive partitioning of domain.

Table 5. Guaranteed version of interval simulated annealing algorithm

Input

Bar $\boldsymbol{X} \subset \mathbb{R}^n$. Given accuracy $\epsilon > 0$.

Interval extension $\boldsymbol{F} : \mathbb{I}\boldsymbol{X} \to \mathbb{IR}$ of objective function F.

Initial T_0 and final T_{fin} temperature values.

Output

Lower estimate F^* of global minimum of function F on $\boldsymbol{X}$.

Algorithm

set $\boldsymbol{Y} \leftarrow \boldsymbol{X}$ and $T \leftarrow T_0$;
set integer N_T—number of trials per temperature level;
calculate $\boldsymbol{F}(\boldsymbol{Y})$ and initialize list $\mathcal{L}$ by record $\{(\boldsymbol{Y}, \underline{\boldsymbol{F}(\boldsymbol{Y})})\}$;
DO WHILE $(T > T_{\text{fin}}$ and wid $(\boldsymbol{F}(\boldsymbol{Y})) \geq \epsilon)$
  DO FOR $j = 1$ TO N_T
    randomly choose record $(\boldsymbol{Z}, \underline{\boldsymbol{F}(\boldsymbol{Z})})$ in $\mathcal{L}$ in accordance with rule $\mathcal{S}(\boldsymbol{Y})$;
    DO (with probability $\mathsf{P}_T(\boldsymbol{Y}, \boldsymbol{Z})$)
      halve $\boldsymbol{Z}$ in longest component into bars $\boldsymbol{Z}'$ and $\boldsymbol{Z}''$;
      calculate $\boldsymbol{F}(\boldsymbol{Z}')$ and $\boldsymbol{F}(\boldsymbol{Z}'')$;
      remove record $(\boldsymbol{Z}, \underline{\boldsymbol{F}(\boldsymbol{Z})})$ from $\mathcal{L}$;
      enter records $(\boldsymbol{Z}', \underline{\boldsymbol{F}(\boldsymbol{Z}')})$ and $(\boldsymbol{Z}'', \underline{\boldsymbol{F}(\boldsymbol{Z}'')})$ into $\mathcal{L}$ in increasing order of second field;
    END DO
    denote lead record in $\mathcal{L}$ by $(\boldsymbol{Y}, \underline{\boldsymbol{F}(\boldsymbol{Y})})$
  END DO
  decrease temperature value $T \leftarrow \alpha T$
END DO
$F^* \leftarrow \underline{\boldsymbol{F}(\boldsymbol{Y})}$;

Our guess is that as the problem dimension n increases, the interval stochastic optimization methods should exhibit a more favorable behavior.

6. NUMERICAL EXAMPLES

The operation of the interval simulated annealing method is demonstrated by solving a problem of search in a $[-10, 10] \times [-10, 10]$ domain of a global minimum of the function

$$f(x, y) = 4x^2 - 2.1x^4 + \tfrac{1}{3}x^6 + xy - 4y^2 + 4y^4, \tag{4}$$

proposed in [14]. This function is also known as a six-hump camel function, since it has six local minima. Two of the minima are global, and their values are $f^* = -1.03163$ reached at points $(0.08984, -0.71266)$ and $(-0.08984, 0.71266)$, which are symmetric about the origin.

The function in question is commonly treated on a smaller domain: $[-5, 5] \times [-5, 5]$ or even $[-3, 3] \times [-2, 2]$. Outside a small region around the origin where all local minima are located, the six-hump camel function increases very rapidly, so that the width of the range of the function is greater than $3 \cdot 10^5$ in our case. By and large this function is a popular test for moderate complexity in global optimization methods.

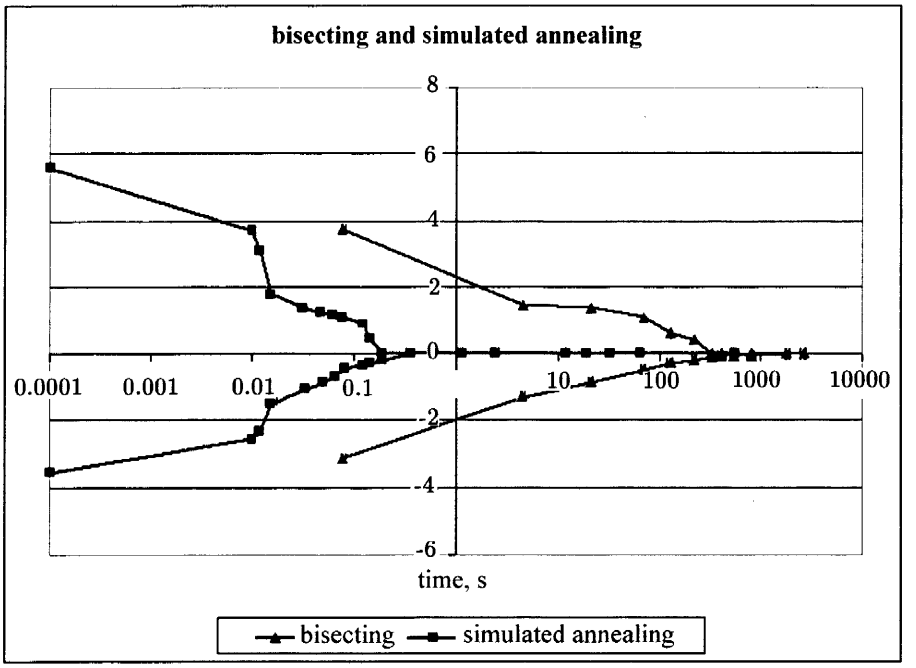

Fig. 5. Comparison of adaptive interval partitioning (bisecting) and interval simulated annealing for six-hump camel function (4).

Figure 5 shows the results obtained with PC Intel Pentium M at a clock frequency of 1.5 GHz using the interval simulated annealing method and conventional adaptive interval partitioning. The operation time of these algorithms in seconds is shown on a logarithmic scale along the x-axis; along the y-axis are the logarithms of absolute values for lower and upper estimates of a minimum which are taken with the minus sign for the negative values of the estimates with the aim to illuminate large-scale convergence. Here, for instance, -4 on the y-axis means -10^4. Interval estimation for ranges of the objective function on subbars were calculated by a simple natural interval extension (see Sec. 2).

It is rather difficult to simultaneously elucidate global and local convergences (near global extrema) in one plot, for these processes differ considerably in scale. Nevertheless, it is easy to see that an optimal value was found three orders of magnitude faster by interval simulated annealing than by the deterministic algorithm in Table 1.

We give yet another example. For Rastrigin's objective function of two variables

$$g(x, y) = x^2 + y^2 - \cos(18x) - \cos(18y)$$

on a $[-10, 10] \times [-10, 10]$ bar, interval simulated annealing with the same settings as in the previous case demonstrated that it converges more than ten orders of magnitude slower than the traditional adaptive interval partitioning. A possible reason is the specific structure of the objective function: globally, it is a simple paraboloid of revolution, but locally, fast oscillations are superimposed on its graph which are caused by the presence of terms with cosines.

7. CONCLUSIONS

Provability (guarantability) of results is a valuable property of modern interval methods of global optimization, but at the same time it is a reason for their poor computational efficiency compared to classical noninterval approaches. Refutation of provability and/or purely deterministic character of interval optimization methods may, we think, lead up to creation of numerical algorithms with essentially new properties, in particular, with better computational efficiency. Elements of stochastic control (randomization) can be introduced into interval methods in a number of different ways. Simplest stochastic interval algorithms are random interval partitioning and interval simulated annealing.

Practical prospects for the methods of stochastic interval optimization, specifically, random interval partitioning and interval simulated annealing, call for further investigation. In our opinion, the stochastic interval methods are doomed to figure prominently in stores of computational optimization.

ACKNOWLEDGMENTS

This work was supported by the Program for Leading Scientific Schools of Russia, grant NSh-9886.2006.9.

REFERENCES

1. Azencott, R., *Protsedura "otpuska"* (Simulated "Annealing"), *Proc. Sem. Bourbaki, 1988*, Moscow: Mir, 1990, pp. 235–251.
2. Alefeld, G. and Herzberger, J., *Vvedenie v intervalnye vychisleniya* (Introduction to Interval Computations), Moscow: Mir, 1987.
3. Gaganov, A.A., On Complexity of Calculating the Range of a Multivariable Polynomial, *Kibern.*, 1985, no. 4, pp. 6–8.
4. Evtushenko, Yu.G. and Rat'kin, V.A., Method of Bisection for Global Optimization of Multivariable Function, *Izv. Akad. Nauk SSSR, Tekhn. Kibern.*, 1987, no. 1, pp. 119–128.
5. Zhiglyavskii, A.A. and Zhilinskas, A.G., *Metody poiska global'nogo ekstremuma* (Methods of Search for a Global Extremum), Moscow: Nauka, 1991.
6. *Intervalnyi analiz i ego prilozheniya* (Interval Analysis and Its Applications); http://www.nsc.ru/interval/.
7. Kalmykov, S.A., Shokin, Yu.I., and Yuldashev, Z.Kh., *Metody interval'nogo analiza* (Methods of Interval Analysis), Novosibirsk: Nauka, 1986.
8. Panov, N.V. and Koldakov, V.V., Programmnyi Kompleks dlya Graficheskogo Predstavleniya Protsessa i Rezul'tatov Raboty Interval'nykh Algoritmov, *Trudy 5-oi mezhdunarodnoi konferentsii "Perspectivy sistem informatiki" pamyati akad. A.P. Ershova* (A. Ershov Fifth International Conference on Perspectives for System Informatics), Novosibirsk, 2003, pp. 38–45.
9. Panov, N.V. and Shary, S.P., Stochastic Approaches to Interval Methods of Global Optimization, *Vserossiiskoe soveshchanie po interval'nomu analizu i ego prilozheniyam INTERVAL-06* (All-Russian Meeting on Interval Analysis and Its Applications INTERVAL-06), St.-Petersburg, 2006, pp. 101–105.
10. Rastrigin, L.A., Statisticheskie metody poiska (Statistical Methods of Search), Moscow: Nauka, 1968.
11. Shary, S.P., Stochastic Approaches to Interval Global Optimization, *Trudy 13-oi Baikal'skoi mezhdunar. shkoly-seminara "Metody optimizatsii i ikh prilozheniya"* (Proc. 13th Baikal Intern. Workshop on Optimization Methods and Their Applications), Irkutsk, 2005, vol. 4., pp. 85–105.
12. Aarts, E. and Korst, J., *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*, Chichester: J. Wiley & Sons, 1989.
13. Corliss, G.F. and Kearfott, R.B., Rigorous Global Search: Industrial Applications, in *Developments in Reliable Computing*, Csendes, T., Ed., Dordrecht: Kluwer, 1999, pp. 1–16; http://interval.louisiana.edu/preprints/scan98.pdf/.
14. Dixon, L.C. and Szegö, G.P., The Global Optimization Problem: An Introduction, in *Towards Global Optimization II*, Dixon, L.C. and Szegö, G.P., Eds., Amsterdam: North Holland, 1978, pp. 1–15.
15. *Encyclopedia of Optimization*, Floudas, C.A. and Pardalos, P.M., Eds., Dordrecht: Kluwer, 2001, vols. I–VI.
16. Hansen, E. and Walster, G.W., *Global Optimization Using Interval Analysis*, New York: Marcel Dekker, 2004.
17. Kearfott, R.B., *Rigorous Global Search: Continuous Problems*, Dordrecht: Kluwer, 1996.
18. Kreinovich, V. and Kearfott, R.B., Beyond Convex? Global Optimization Is Feasible Only for Convex Objective Functions: A Theorem, *J. Glob. Optim.*, 2005, vol. 33, no. 4, pp. 617–624.
19. Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P., Optimization by Simulated Annealing, *Science*, 1983, vol. 220, pp. 671–680.
20. Moore, R.E., *Methods and Applications of Interval Analysis*, Philadelphia: SIAM, 1979.
21. Neumaier, A., *Interval Methods for Systems of Equations*, Cambridge: Cambridge Univ. Press, 1990.
22. Ratschek, H. and Rokne, J., *New Computer Methods for Global Optimization*, New York: Halsted Press, 1988.