Journal of Global Optimization **10:** 185–206, 1997.

Parallel Characteristical Algorithms for Solving Problems of Global Optimization

VLADIMIR A. GRISHAGIN[1], YAROSLAV D. SERGEYEV[1,2] and ROMAN G. STRONGIN[1]
[1]*Software Department, University of Nizhni Novgorod, Gagarin Ave. 23, Nizhni Novgorod, 603600, Russia (e-mail: strongin@nnucnit.unn.ac.ru);* [2]*ISI CNR, c/o DEIS University of Calabria, Rende (CS), Italy (e-mail: yaro@si.deis.unical.it)*

(Received: 24 March 1993; accepted: 20 June 1996)

Abstract. A class of parallel characteristical algorithms for global optimization of one-dimensional multiextremal functions is introduced. General convergence and efficiency conditions for the algorithms of the class introduced are established. A generalization for the multidimensional case is considered. Examples of parallel characteristical algorithms and numerical experiments are presented.

Key words: global optimization, parallel computations, characteristical algorithms.

1. Introduction

In this paper we introduce a class of parallel algorithms to solve the one-dimensional global optimization global optimization problem over an interval, i.e. to find the absolute minimum of a real-valued function $\phi(x), x \in [a, b]$. Many known methods were designed for sequential computers (see, e.g. [2,8, 10–14, 17–19, 21, 26, 30–31]) and therefore executed *trials* (evaluations of the objective function) sequentially. Parallel algorithms under consideration perform several trials simultaneously during each iteration – one trial at each of the processors of the multiprocessor system the algorithm is implemented on. Such a procedure allows us to accelerate solving the problems in which the performance of even one trial requires a lot of time.

Different approaches to designing the parallel computational methods are given in [3, 4, 6, 7, 16, 20, 27–29]. In particular, parallel computations can be used to accelerate executing the optimization algorithm decision rules and to decrease the time of conducting each trial (i.e. computing the function value in one point). Anyhow, such parallelization is specific for each concrete algorithm and concrete problem, whereas the subject under consideration in this paper is the construction of some general principles of parallel choice of trial points for the class of parallel characteristical algorithms.

The paper is structured as follows. Section two introduces the class of parallel characteristical algorithms and gives examples of algorithms belonging to

this class. Conditions providing various kinds of convergence for parallel characteristical algorithms are established in Section three. Section four is devoted to a generalization of the multidimensional case. Efficiency of parallelization is theoretically estimated in Section 5, while results of numerical experiments are described in Section 6. Section 7 concludes the paper.

2. Parallel Characteristical Algorithms

A global optimization algorithm minimizing a function $\phi(x), x \in [a, b]$ is called a *parallel characteristical algorithm* if trial points are chosen according to the rules found below.

Trials of the first $n \geq 1$ iterations are performed in arbitrary $K = k(n) = p(1) + p(2) + \cdots + p(n)$ points of the interval $[a, b]$, where $p(i), i \geq 1$, denotes the number of trials of the i-th iteration. Trial points corresponding to any subsequent Q-th iteration, $Q > n$, are chosen according to the rule:

(1) points of the set

$$X_k = \{x^1, \ldots, x^k\} \cup \{a\} \cup \{b\}, \tag{2.1}$$

including the boundaries of the interval $[a, b]$ and the coordinates $x^j, 1 \leq j \leq k$, of preceding trials, where

$$k = k(Q-1) = p(1) + \cdots + p(Q-1) \tag{2.2}$$

are renumbered (by subscripts) in the order of increasing the coordinates, namely

$$a = x_0 < x_1 < \cdots < x_\tau = b \tag{2.3}$$

where $\tau + 1 = \tau(Q) + 1$ is the quantity of (different) elements of the set X_k from (2.1);

(2) a real number $R(i)$ is assigned to each subinterval $(x_{i-1}, x_i), 1 \leq i \leq \tau$, where $R(i)$ is called the *characteristic* of this subinterval;

(3) characteristics $R(i)$ of the subintervals $(x_{i-1}, x_i), 1 \leq i \leq \tau$, are ordered by decreasing

$$R(i_1) \geq R(i_2) \geq \cdots \geq R(i_\tau); \tag{2.4}$$

(4) the next p trials of the Q-th iteration are performed in the points of the set

$$T(Q) = \{x^{k+1}, \ldots, x^{k+p}\}, \tag{2.5}$$

where

$$x^{k+q} = S(i_q) \tag{2.6}$$

and $i_q, 1 \leq q \leq p$, are the first p indices from the series (2.4) and the function S is such that

$$x^{k+q} \in (x_{i_q-1}, x_{i_q}). \tag{2.7}$$

In this case it is supposed that

$$p = p(Q) \leq \tau, \quad Q > n. \tag{2.8}$$

The class of parallel characteristical algorithms introduced includes two known sets of purely sequential methods: the characteristically presented algorithms [10,11], and the GA methods [17]. These sets include algorithms whose computing scheme is easily derived from (2.1)–(2.8), if we assume in (2.8) that

$$p = p(Q) = 1, \quad Q \geq 1,$$

i.e. all trials are successively performed by one processor. The sequential scanning method, algorithms based on piecewise linear support functions (e.g. broken lines method [19] by Piyavskii), information-statistic global search algorithms [26], Kushner's method [14], Bayesian algorithms (e.g. Zilinskas' method [31]), and Pinter's methods [17] may all serve as examples of such algorithms.

Thus, the scheme (2.1)–(2.8) permits us both to construct parallel methods on the basis of known sequential algorithms and to create new parallel methods having no sequential prototypes. Below we present some parallel algorithms constructed on the basis of the purely sequential characteristical algorithms mentioned above. We use the following notations

$$\Delta_j = x_j - x_{j-1}, 1 \leq j \leq \tau; z_j = \phi(x_j), \quad 0 \leq j \leq \tau. \tag{2.9}$$

EXAMPLE 1. *The parallel scanning method.*

The characteristic of interval (x_{i-1}, x_i) for this method is $R(i) = x_i - x_{i-1}$, i.e., the length of the interval, and point $x^{k+q} \in T(Q)$ is formed as

$$x^{k+q} = \frac{1}{2}(x_{i_q-1} + x_{i_q}). \tag{2.10}$$

EXAMPLE 2. *Broken lines parallel method*.

This method places two first trials on the ends of interval $[a, b]$ and then follows the basic characteristical rule with characteristics

$$R(i) = \frac{m}{2}\Delta_i - \frac{1}{2}(z_i + z_{i-1}), \quad 1 \leq i \leq \tau = k - 1,$$

and trial points

$$x^{k+q} = \frac{1}{2}(x_{i_q-1} + x_{i_q}) - \frac{z_{i_q} - z_{i_q-1}}{2m}, 1 \leq q \leq p, \tag{2.11}$$

where $m > 0$ is a parameter of the method.

EXAMPLE 3. *The information algorithm with parallel iterations*

This algorithm [24] as a point of the first trial takes any inner point of the interval $[a, b]$ and then uses characteristics

$$R(1) = 2r\omega\Delta_1 - 4z_1, \tag{2.12}$$

$$R(\tau) = 2r\omega\Delta_\tau - 4z_{\tau-1}, \tag{2.13}$$

$$R(i) = r\omega\Delta_i + \frac{r(z_i - z_{i-1})^2}{\omega\Delta_i} - 2(z_i + z_{i-1}), \quad 1 < i < \tau, \tag{2.14}$$

where

$$\omega = \max\{|z_i - z_{i-1}|/\Delta_i, \quad 1 < i < \tau\}, \tag{2.15}$$

and $r > 1$ is a parameter of the method. If ω is impossible to determine according to (2.15) or $\omega = 0$, then $\omega = 1$. The trial points are calculated according to (2.10), if $i_q = 1$ or $i_q = \tau$ and, according to (2.11), with the replacement of m by $r\omega$, if $1 < i_q < \tau$.

3. Convergence of Parallel Characteristical Algorithms

The definition of the characteristical algorithms introduced lacks the rule of stopping computations, i.e. it is supposed that the algorithm generates an infinite sequence of trials $\{x^s\}$ and a corresponding sequence $\{z^s\}$ of the values $z^j = \phi(x^j), j \geq 1$. The properties of these sequences reflect those of an algorithm itself. Therefore an investigation of optimization methods will be carried out studying the sequences $\{x^s\}$ and $\{z^s\}$ generated by them. This is why the section is devoted to the problem of the convergence of sequences $\{x^s\}$.

Let us first introduce the notations

$$\delta_i = \min\{z_{i-1}, z_i\}, \tag{3.1}$$

where z_{i-1}, z_i are from (2.9),

$$\hat{\delta}_i = \begin{cases} \delta_i, & \text{if } x_{i-1} \in \{x^s\} \text{ and } x_i \in \{x^s\}; \\ z_{i-1}, & \text{if } x_{i-1} \in \{x^s\} \text{ and } x_i \notin \{x^s\}; \\ z_i, & \text{if } x_{i-1} \notin \{x^s\} \text{ and } x_i \in \{x^s\}. \end{cases} \tag{3.2}$$

THEOREM 1. *Minimizing a function $\phi(x), x \in [a, b]$, by a parallel characteristical algorithm, let for characteristics $R(i)$, $1 \leq i \leq \tau$, and trial points $x^{k+q}, 1 \leq q \leq p$, of the algorithm hold the following relations:*

i. $$\lim_{Q\to\infty} \lim_{\substack{x_{i-1}\to\bar{x} \\ x_i\to\bar{x}}} R(i) = -\mu\phi(\bar{x}) + c; \tag{3.3}$$

ii. $$\lim_{Q\to\infty} R(i) > \alpha\Delta_i - \mu\hat{\delta}_i + c, \tag{3.4}$$

if $(x_{i-1}, x_i) \cap \{x^s\} = \emptyset$*;*

iii. $\max\{x^{k+q} - x_{i-1}, x_i - x^{k+q}\} \le \nu\Delta_{i_q}$, (3.5)

where μ, c, α, ν *are constants,* $\mu \ge 0$ *and*

$$0 < \nu < 1. \tag{3.6}$$

Then, with $\alpha \ge 0$ *for any inner point* $x^* \in [a, b]$ *being a limit point of the trial sequence generated by the algorithm, there exist two subsequences of trials, one of which converges to* x^* *from the left, the other one from the right.*

Proof. Let $x^* \notin \{x^s\}$ and $t = t(Q)$ be the number of the interval (x_{t-1}, x_t) containing the point x^* after the Q-th iteration. Since the point x^* is a limit one, then it follows from (2.7), (3.5), (3.6) that

$$\lim_{Q\to\infty} \Delta_t = 0. \tag{3.7}$$

Thus, as required subsequences we can use sequences $\{x_{t(Q)-1}\}$ and $\{x_{t(Q)}\}$ of the left and right ends of intervals (x_{t-1}, x_t) accordingly.

If $x^* \in \{x^s\}$, there will be such numbers $u \ge 1$ and $s \ge k(u)$ that $x^* = x^s \in T(u)$. Then, for any $Q > u$ there exists such a number $j = j(Q)$ that $x_j = x^*$. Let us assume that convergence to x^* from the left is absent (the case when no subsequence converging to x^* from the right is considered identically). Then, there will be such iteration number $d \ge u$ and a trial number $v > k(d)$ that $x^v \in T(d)$ and for all $Q \ge d$ trials will not get into interval $(x_{j-1}, x_j) = (x^v, x^s)$.

As a result of (3.4) we obtain

$$\lim_{Q\to\infty} R(j) > -\mu\phi(x^*) + c, \tag{3.8}$$

yet according to (3.3) for the adjacent interval $(x_{t-1}, x_t), t = j + 1$,

$$\lim_{Q\to\infty} R(t) = -\mu\phi(x^*) + c. \tag{3.9}$$

Therefore, the following inequality for a sufficiently large number Q will be true

$$R(j) > R(t). \tag{3.10}$$

Because of decision rules (2.1)–(2.8), (3.10) contradicts to the impossibility of performing trials within the interval (x_{j-1}, x_j). ■

NOTE. According to Theorem 1 for any interval $(x_{t-1}, x_t), t = t(Q)$, containing a limit point $x^* \in [a, b]$ the relation (3.7) takes place (if $x^* = a$ or $x^* = b$, unilateral convergence is sufficient to fulfil (3.7)). The rules (2.4) and (2.7), together with (3.7), allow us to determine the stopping condition of the type

$$\min_{1\le q\le p} \rho(x_{i_q-1}, x_{i_q}) \le \varepsilon, \tag{3.11}$$

where $\rho(\bullet)$ is a continuous function possessing metrics properties, and $\varepsilon > 0$ is a preset search accuracy.

Let us now establish the truth of the theorem for concrete algorithms from Section 2.

The fulfillment of the conditions of the theorem with $\nu = \frac{1}{2}, \alpha = \mu = c = 0$ is evident for the scanning method (see Example 1).

Let us assume now that the function $\phi(x)$ meets the Lipschitz condition with a constant $L > 0$ within the interval $[a, b]$.

Taking into account the Lipschitz condition, and making use of the relation

$$\delta_i = \min\{z_{i-1}, z_i\} = \frac{1}{2}(z_{i-1} + z_i - |z_i - z_{i-1}|), \tag{3.12}$$

it is not difficult to show that Theorem 1 is true for the broken lines parallel method (see Example 2) with $\alpha = c = 0, \mu = 1, \nu = \frac{1}{2}\left(1 + \frac{L}{m}\right)$ if $m > L$.

Assuming that ω from (2.15) is bounded, the information algorithm with parallel iterations (see Example 3), insures the fulfillment of (3.3) with $\alpha = c = 0, \mu = 4$, and (3.5) with $\nu = \frac{1}{2}\left(1 + \frac{1}{r}\right)$. The truth of (3.4) for the characteristics (2.12) and (2.13) is evident. As for characteristic (2.14), let us present it in the following form

$$R(i) = r\omega\Delta_i\left(\beta^2 - \frac{2\beta}{r} + 1\right) - 4\delta_i$$

where $\beta = |z_1 - z_{i-1}|/(\omega\Delta_1)$. In so far as β satisfies inequality $0 \le \beta \le 1$, then

$$R(i) \ge r\omega\Delta_i(1 - r^{-2}) - 4\delta_i >= 4\delta_i.$$

THEOREM 2. *Under the conditions of Theorem 1 with $\mu = 0$ and $\alpha \ge 0$, any point of the interval $[a, b]$ is a limit point of the trial sequence $\{x^s\}$ generated by the algorithm.*

Proof. Let us assume that there exists a point $\bar{x} \in [a, b]$ which is not a limit point of the trial sequence. We designate as $j = j(Q)$ the number of an interval (x_{j-1}, x_j), such that $x_{j-1} \le \bar{x} \le x_j$. (If $\bar{x} \in (a, b)$ and for some $u \ge 1$ $\bar{x} = x^u$, then there exist two such intervals and one can take any of them). Starting from a certain step, the trials will not fall into this interval; therefore, according to (3.4)

$$\lim_{Q\to\infty} R(j) > c. \tag{3.13}$$

On the other hand, due to the boundness of the interval $[a, b]$, there exists at least one limit point x^* of the trial sequence $\{s^s\}$. For characteristics of the intervals $(x_{t-1}, x_t), t = t(Q)$, such that $x_{t-1} \le x^* \le x_t$ according to (3.3) we have

$$\lim_{Q\to\infty} R(i) = c. \tag{3.14}$$

Due to (3.13), (3.14) for a sufficiently large number Q inequality (3.10) holds. This fact contradicts our initial assumption because of (2.4), (2.7). ■

This theorem establishes the conditions of the so-called "everywhere dense" convergence of parallel characteristical algorithms. At the same time, these conditions are sufficient conditions of convergence to the global minimizer of the function $\phi(x)$, if convergence of the trial sequence $\{x^s\}$ to some point $\hat{x}$ should be defined as existence of a subsequence $\{x^v\}$, $\{x^v\} \subseteq \{x^s\}$, converging to $\hat{x}$.

The "everywhere dense" type of convergence is inherent, for example, to the scanning method and methods [14, 31]. The other type of convergence is established by the following theorem.

THEOREM 3. *Let: (i) the objective function $\phi(x), x \in [a, b]$, be continuous and the number of local extrema of $\phi(x)$ be finite;*

(ii) conditions (3.3)–(3.6) along with $\mu > 0$ and $\alpha \geq 0$ hold for a parallel characteristical algorithm;

(iii) the number $p(Q)$ of parallel trials be uniformly limited, i.e.

$$p(Q) \leq P, \quad Q > n, \tag{3.15}$$

where $P > 1$ is a preset constant;

(iv) x^* be a limit point of the trial sequence $\{x^s\}$ generated by the algorithm.

Then: (1) $\phi(x^s) \geq \phi(x^*), \quad s \geq 1$;

(2) if there exists another limit point x^{**} alongside with x^*, then $\phi(x^*) = \phi(x^{**})$;

(3) the point x^* is a local minimizer if the function $\phi(x)$ has a finite number of local extrema within the interval $[a, b]$.

Proof. Let us denote by $t = t(Q)$ the number of an interval (x_{t-1}, x_t) such that $x_{t-1} \leq x^* \leq x_t$. Point out that according to Theorem 1 (3.9) is true for the characteristic $R(t)$ of this interval.

1. Proving the first assertion let us assume that as a result of an iteration with a number $d \geq 1$, the trial in a point $x^m \in T(d)$ results in getting value $\phi(x^m)$ which is such that

$$\phi(x^m) < \phi(x^*). \tag{3.16}$$

We designate by $j = j(Q)$ the number of the point x^m in the series (2.3) which corresponds to the iteration with a number $Q \geq d$, i.e. $x_j = x^m, z_j = \phi(x^m)$. Let us show that the point x_j is a limit point for the trial sequence $\{x^s\}$ also. If it is not true, then for the characteristic $R(j)$ of the interval (x_{j-1}, x_j) (if $j = 0$ it is necessary to take the interval (x_j, x_{j+1})) according to (3.4) we have

$$\lim_{Q \to \infty} R(j) > -\mu z_j + c. \tag{3.17}$$

However, taking into account (3.9) and (3.16), we obtain that starting from a certain step, inequality (3.10) holds. This fact contradicts to our assertion that x_j is not a limit point of $\{x^s\}$.

Since the objective function $\phi(x)$ is continuous, then inequality $\phi(x) < \phi(x^*)$, $x \in \Lambda$, is true in some neighbourhood Λ of the point x^*. As the point x^m is a limit

point, this neighborhood will contain more than $P+1$ points from $\{x^s\}$, i.e. more than P intervals formed by these points will have characteristics greater than the characteristic $R(t)$ from (3.9). This means that in accordance with the decision rule (2.4) the trials will fall into the interval (x_{t-1}, x_t); yet this is impossible because the point x^* is the limit one.

2. To prove the second assertion it is sufficient to assume that $\phi(x^*) < \phi(x^{**})$. In assuming this, however, we are immediately in contradiction with the first assertion.

3. Let us assume that the point x^* is not a local minimizer. Since the number of local extrema of the function $\phi(x)$ is finite, then there exists a neighbourhood of the point x^* within which the function $\phi(x)$ is strictly monotonous: i.e. either on the left or on the right of the point x^*, the inequality $\phi(x) < \phi(x^*)$ is true. As much as convergence to the point x^* is bilateral (see Theorem 1), however, there will be an obligatory point x^m satisfying (3.16). (In the cases $x^* = a$ or $x^* = b$, unilateral convergence is sufficient.) This fact contradicts the first assertion of the theorem. ■

Thus, when the conditions of Theorem 3 are fulfilled for a parallel characteristical algorithm, its trial sequence $\{x^s\}$ can have only local and global minimizers as limit points, and convergence to different-height minima is impossible. This type of convergence is assured by the Piyavskii method [19] and by the whole spectrum of information algorithms [24–26].

The assumptions of Theorem 3, providing local optimality of limit points, do not guarantee convergence to global minimizers. Such guarantees (sufficient conditions of global convergence) are given in the next section together with a scheme generalizing one-dimensional characteristical algorithms to the multi-dimensional case.

4. Multidimensional Optimization and Conditions of Global Convergence

There are several ways to extend the characteristical algorithms in order to solve multidimensional global optimization problems which employ, for example, the approaches proposed in [13, 18, 30]. In this section, we will consider another approach (see [5,21,25]) based on Peano-type space-filling curves.

Consider the problem of finding the global minimum of a function $\Phi(z)$ of N variables over a hyperinterval D:

$$\Phi(z^*) = \min\{\Phi(z) : \ z \in D\}, \tag{4.1}$$

$$D = \{z \in R^N : \ a_j \le z_j \le b_j, 1 \le j \le N\}, \tag{4.2}$$

where, in general, $\Phi(z)$ is multiextremal.

If $\Phi(z)$ is continuous, then for the function

$$\phi(x) = \Phi(z(x)), x \in [a, b], \tag{4.3}$$

where $z(x)$ is the continuous Peano-type mapping of closed interval $[a, b]$ onto the hyperinterval D, we have

$$\min\{\phi(x) : x \in [a, b]\} = \min\{\Phi(z) : z \in D\}. \tag{4.4}$$

Therefore, solving the multidimensional problem (4.1), (4.2) can be replaced by solving the following one-dimensional problem

$$\phi(x^*) = \min\{\phi(x) : x \in [a, b]\} \tag{4.5}$$

with $\phi(x)$ from (4.3).

As it has been shown in [25], if $\Phi(z)$ is Lipschitzean with a constant $K > 0$ then $\phi(x)$ satisfies the Hölder condition

$$|\phi(x') - \phi(x'')| \leq L|x' - x''|^{1/N}, \quad x', x'' \in [a, b], \tag{4.6}$$

with a constant $L \geq 0$ (Hölder constant). If $N = 1$ then (4.6) is the usual Lipschitz condition.

For parallel characteristical algorithms the following theorem establishes sufficient conditions of convergence to the global minima of the functions satisfying (4.6) and, therefore, substantiates the capacity of methods to be considered for solving the Lipschitzean multidimensional problems (4.1), (4.2) via the Peano-type mappings.

THEOREM 4. *Let a function, $\phi(x)$, satisfy the Gölder condition (4.6). Assuming that the parallel characteristical algorithm minimizing the function $\phi(x)$ over $[a, b]$ meets the conditions (3.3), (3.5), (3.6) and, for all the intervals (x_{i-1}, x_i) such that $(x_{i-1}, x_i) \cap \{x^s\} = \emptyset$, the relation*

$$\lim_{Q\to\infty} R(i) > \alpha\Delta_i^{1/N} - \mu\hat{\delta}_i + c \tag{4.7}$$

is fulfilled with

$$\alpha \geq 2^{-1/N}\mu L \tag{4.8}$$

if $x_{i-1} \in \{x^s\}$ and $x_i \in \{x^s\}$, and

$$\alpha \geq 2^{1-1/N}\mu L \tag{4.9}$$

in the opposite case.

Any global minimizer of the function $\phi(x)$ in the interval $[a, b]$ is then the limit point of the trial sequence generated by the algorithm.

Proof. Let x^* be a global minimizer of the function $\phi(x)$, and $j = j(Q)$ be the number of an interval (x_{j-1}, x_j) such that $x_{j-1} \leq x^* \leq x_j$. Let us assume that x^* is not a limit point of the trial sequence $\{x^s\}$. Starting from a certain step of the search, then, trials will not fall into the interval (x_{j-1}, x_j).

Being within the interval (x_{j-1}, x_j), the point x^* can be represented as $x^* = \beta x_{j-1} + (1-\beta)x_j, 0 \le \beta \le 1$. Then, by virtue of (4.6) we obtain

$$z_{j-1} - \phi(x^*) \le L(x^* - x_{j-1})^{1/N} = L(1-\beta)^{1/N}\Delta_j^{1/N},$$
$$z_j - \phi(x^*) \le L(x_j - x^*)^{1/N} = L\beta^{1/N}\Delta_j^{1/N},$$

from where

$$\begin{aligned} z_j + z_{j-1} &\le 2\phi(x^*) + L(\beta^{1/N} + (1-\beta)^{1/N})\Delta_j \le \\ &\le 2\phi(x^*) + L\Delta_j^{1/N} \max\{\beta^{1/N} + (1-\beta)^{1/N}\} = \\ &= 2\phi(x^*) + 2^{1-1/N} L\Delta_j^{1/N} \end{aligned} \tag{4.10}$$

In the case of (4.8) $\hat{\delta}_j = \delta_j$, and taking (3.12) into account, we have

$$\lim_{Q\to\infty} R(j) > \left(\alpha\Delta_j^{1/N} - \frac{\mu}{2}(z_j + z_{j-1})\right) + c$$

from which the following inequality holds as a result of (4.8)–(4.10):

$$\lim_{Q\to\infty} R(j) > -\mu\phi(x^*) + c. \tag{4.11}$$

Let it now be $\hat{\delta}_j = z_j$ in situation (4.9) (the case $\hat{\delta}_j = z_{j-1}$ is considered analogously). Then, taking into account (4.10)

$$\begin{aligned} \lim_{Q\to\infty} R(j) &> \alpha\Delta_j^{1/N} - \mu z_j + c \ge \\ &\ge \alpha\Delta_j^{1/N} - 2\mu\phi(x^*) + \mu\phi(x_{j-1}) - 2^{1-1/N}\mu L\Delta_j^{1/N} \ge \\ &\ge (\alpha - 2^{1-1/N}\mu L)\Delta_j^{1/N} - \mu\phi(x^*) + c \ge \\ &\ge -\mu\phi(x^*) + c, \end{aligned}$$

i.e. (4.11) is again true.

At the same time there exists at least one limit point $\bar{x}$ of the trial sequence $\{x^s\}$ because the search interval $[a, b]$ is a bounded set. For the characteristic $R(t)$ of an interval (x_{t-1}, x_t) such that $x_{t-1} \le \bar{x} \le x_t$ we have

$$\lim_{Q\to\infty} R(t) = -\mu\phi(\bar{x}) + c. \tag{4.12}$$

Conditions (4.11) and (4.12) allow us to draw the conclusion that starting from a certain step of the search $R(j) > R(t)$ which brings the assertion about the absence of trials in the interval (x_{j-1}, x_j) to the contradiction with (2.4). ■

NOTE. If $\mu = 0$, then the conditions of Theorem 4 coincide with the assumptions of Theorem 2, and it is not necessary to suppose (4.6) for the function $\phi(x)$.

Thus, Theorem 4 formulates requirements upon whose completing the sequence of trials of the parallel characteristical algorithm have all global minimizers as limit points. If in this case $\mu > 0$, and the number of parallel processors is uniformly limited, then no other points can be limit points for this sequence. Note that in the set of sequential methods the similar property is assured: for instance, by the methods [13, 17, 19, 21, 22, 26].

The global convergence conditions for concrete known representatives of characteristical class are a partial case of the general results of Theorem 4. For illustration, assume that the function to be minimized meets the Lipschitz condition that corresponds to $N = 1$ in (4.6) and consider two examples.

First, we will deal with the algorithm [24] for which $\mu = 4$ and $c = 0$ in relations (3.3) and (4.7). Its characteristics (2.14)

$$R(i) \geq r\omega\Delta_i - \frac{\mu}{2}(z_i + z_{i-1}) \geq r\omega\Delta_i - \mu\delta_i$$

as from (3.12) $\delta_i \leq \frac{1}{2}(z_i + z_{i-1})$. Then (4.7) holds if

$$r\omega > 2L. \tag{4.13}$$

It is easy to derive the same inequality for characteristics (2.12) and (2.13). But (4.13) coincides with the global convergence condition from [24].

As the second example we consider the broken lines parallel method ($\mu = 1, c = 0$). In the sequential variant, it is similar to the Piyavskii algorithm [19]. Following the reasonings of the previous example we obtain for the characteristics of this method

$$R(i) \geq \frac{m}{2}\Delta_i - \mu\delta_i$$

from where (4.7) is true if $m/2 > \mu L/2$, or $m > L$. For the Piyavskii algorithm m has a geometrical interpretation as the slope of linear pieces of the support function. So, the basic condition of application for the method [19] can be considered as the consequence of characteristical theory.

The above methods can be generalized [23–25] for the case of Hölder functions (4.6). These modified algorithms being also the characteristical methods can be applied for solving the multidimensional problems (4.1) and (4.2) by means of the reduction scheme (4.3)–(4.5). Formally, the generalization consists in using the expression

$$\begin{aligned} x^{k+q} \;=\; & \frac{1}{2}(x_{i_q-1} + x_{i_q}) - \\ & -(2r)^{-1}(|z_{i_q} - z_{i_q-1}|\omega^{-1})\ \mathrm{sign}(z_{i_q} - z_{i_q-1}) \end{aligned} \tag{4.14}$$

instead of (2.11), and in the replacement of the Eucleadean length Δ_j from (2.9) by the Hölеerian length $\Delta_j = (x_j - x_{j-1})^{1/N}$ in the characteristics of the methods and in the estimate (2.15).

The analogous consideration of the generalized methods in the frame of characteristical theory leads to the same global convergence conditions which have been derived in [23–25].

5. Conditions of Nonredundant Parallelization

The parallelization of the sequential characteristical algorithms by the scheme (2.1)–(2.8) results in giving up a portion of the search information when choosing trial points. In fact, the values of the objective function at the points of all preceding trials are known for the sequential algorithm at the moment of selecting a point x^{k+1} for the next trial (see (2.6) with $q = 1$). In the parallel method, the choice of the points x^{k+j}, $1 < j \leq p = p(Q)$, from (2.6) at the Q-th iteration is made without taking into consideration the function values at the points $x^{k+1}, \ldots, x^{k+j-1}$ which have not been evaluated yet. The higher is the level of parallelization determined by conditions (2.2)–(2.8) and the function $p(Q)$, the more significant are the losses of information in planning trials, and these losses are most considerable in the case of complete parallelization when $p(Q) = \tau(Q)$.

Incomplete account of information can bring up the situation when the parallel algorithm produces the trials more densely in comparison with its sequential prototype, i.e. it generates *redundant* trials.

Following [25], let us introduce a number of notions. Let $\{x^k\}$ and $\{y^m\}$ be infinite (with $\varepsilon = 0$ in stopping condition (3.11)) sequences of trial points generated accordingly by a purely sequential characteristical algorithm and its parallel analog in case of minimizing the same function $\phi(x), x \in [a, b]$. Coincidence of these sequences, i.e.

$$\{x^k\} = \{y^m\}, \tag{5.1}$$

means that the parallel algorithm places trials at the same points where the purely sequential method executes trials. Note that (5.1) does not require the fulfillment of $x^s = y^s, s \geq 1$. When condition (5.1) takes place parallelization is called *nonredundant*. But if condition (5.1) is not observed, it means that the parallel scheme possesses some redundancy. Let us introduce a *redundancy coefficient* for its quantitative characteristic

$$\lambda(m, s) = T'(m, s)/(m - s), m > s, \tag{5.2}$$

where

$$T'(m, s) = \text{card}(\{y^{s+1}, \ldots, y^m\} \backslash \{x^k\}) \tag{5.3}$$

is the number of redundant points in $\{y^m\}$ from the $(s + 1)$-th to the m-th trial. This definition presupposes that inclusion $\{x^k\} \subset \{y^m\}$ takes place. It is evident that $\lambda(m, 0) = 0$ corresponds to the nonredundant case (5.1).

Let us consider parallel characteristical algorithms processing with two trials in every iteration after n initial steps (here in after we suppose that in the course of these n iterations K trials have been done), i.e.

$$p(Q) = 2, Q > n, \tag{5.4}$$

To continue we need a number of nonburdensome assumptions. We suppose that if trials have been executed in both the ends of a subinterval (x_{j-1}, x_j) then a new trial points $x^{k+q} \in (x_{j-1}, x_j)$ can be expressed as follows

$$x^{k+q} = c_j + \text{sign}(z_{j-1} - z_j)\xi_j, \tag{5.5}$$

$$c_j = (x_{j-1} + x_j)/2, \tag{5.6}$$

$$0 \leq \xi_j \leq \sigma\Delta_j, \tag{5.7}$$

where Δ_j is from (2.9) and $0 < \sigma \leq 0.5$. Note that (5.5)–(5.7) implicate (3.5) if $\nu = \frac{1}{2} + \sigma$.

If one of the ends of a subinterval (x_{j-1}, x_j) is not the trial point then we use

$$x^{k+q} = c_j. \tag{5.8}$$

Along with (5.5) and (5.8), assume that characteristics $R(i), 1 \leq i \leq \tau$, and values $\xi_j, 1 \leq j \leq p$, are completely determined by the points x_{j-1}, x_j of the corresponding subintervals and by the values of the objective function evaluated at these points, i.e.

$$R(i) = \Psi(x_{j-1}, x_j, z_{j-1}, z_j), \tag{5.9}$$

$$\xi_j = \Xi(x_{j-1}, x_j, z_{j-1}, z_j). \tag{5.10}$$

Let us also assume that

$$y^s = x^s, 1 \leq s \leq K, \tag{5.11}$$

i.e. initial steps of the search for the sequential and parallel methods not connected with characteristical decision rules (2.4)–(2.7), are identical.

THEOREM 5. *Let: (i) the objective function $\phi(x), x \in [a, b]$, meet Lipschitz condition with a constant L;*

(ii) conditions (5.4), (5.5), (5.7), (5.9)–(5.11), (3.3), $\mu > 0$ be fulfilled for sequential and parallel schemes of a characteristical algorithm;

(iii) for $\Delta_i > 0$ it follow

$$R(i) > \alpha\Delta_i - \mu\hat{\delta}_i + c. \tag{5.12}$$

Then with $\sigma \leq \frac{1}{6}$ and

$$\alpha > \mu L \tag{5.13}$$

the inequality

$$\lambda(m, K) \leq E[(m - K)/6]/(m - K) < 0.17 \tag{5.14}$$

takes place. Here $E[d]$ is integer part of d.

Proof. If after the m-th trial of the parallel method and the k-th trial of the sequential method, the equality

$$(y_{t-1}, y_t) = (x_{q-1}, x_q), t = t(m), q = q(k), \tag{5.15}$$

is true and the next trials at points y^{m+1} and x^{k+1} respectively get into indicated intervals. Then on account of (5.9)

$$R(t(m)) = R(q(k)) \tag{5.16}$$

and according to (2.6), (5.5), (5.8), (5.10) we have

$$y^{m+1} = x^{k+1}. \tag{5.17}$$

From (5.11) with $m = k = n$ on account of (5.15)–(5.17) and decision rules (2.4)–(2.7) we obtain that

$$\{x^k\} \subseteq \{y^m\}. \tag{5.18}$$

Inclusion (5.18) makes it possible to evaluate redundancy with the help of the coefficient (5.2).

The truth of conditions (3.3), (3.13), (5.13) and (3.5) with $\nu = 0.5 + \sigma$ ensures for the accomplishment of the requirements of Theorem 4; therefore only the global minimizers of the function $\phi(x)$ can be limit points of the sequences $\{x^k\}$ and $\{y^m\}$. Thus, the set of limit points of the sequence $\{y^m\}$ coincides with the set of limit points of the sequence $\{x^k\}$.

Let the first k points of the sequence $\{x^k\}$ be arranged in accordance with (2.3) and $j = j(k)$ be the number of an interval (x_{j-1}, x_j) such that $x_{j-1} \leq x^* \leq x_j$, where x^* is a global minimizer of $\phi(x)$. Due to (5.12), (5.13) and the Lipschitz condition for $\phi(x)$, the relation

$$R(j) > -\mu\phi(x^*) + c \tag{5.19}$$

is true for the characteristic of this interval. As far as point x^* is the limit point, then because of (3.3)

$$R(j(k)) \to -\mu\phi(x^*) + c + 0,$$

if $k \to \infty$.

Taking into consideration (2.4) it follows from (5.19) that any interval (x_{i-1}, x_i), $i = i(k)$, whose characteristic satisfies (5.19) contains at least one point of the sequence $\{x^k\}$. At the same time any interval (x_{i-1}, x_i) for which

$$R(i) < -\mu\phi(x^*) + c \tag{5.20}$$

is true, does not contain points of the sequence $\{x^k\}$. Thus, we have obtained that redundant trials of the parallel method can be executed only in the intervals for which inequality (5.20) is true.

Let us again consider the interval $[x_{j-1}, x_j], j = j(k)$, containing the global minimizer x^*. The next trial executed at a point $x = x^{k+1}$ belonging to this interval generates two new subintervals

$$[x_{j-1}, x], [x, x_j] \tag{5.21}$$

one of which contains x^*. Let it be the first of them, i.e.

$$x^* \in [x_{j-1}, x]. \tag{5.22}$$

The truth of inequality (5.19) for the interval (x_{j-1}, x) from (5.21) comes out of (5.22). Let us show that this inequality is also true for the interval (x, x_j). We need the following designations

$$\Delta_x = x_j - x, \delta_x = \min\{z_j, z\}, z = z^{k+1}. \tag{5.23}$$

Suppose that the previous story of the search is such that the points x_{j-1} and x_j were the trials points: i.e. values z_{j-1} and z_j have been calculated. Let us evaluate the magnitude δ_x. Consider the case $z_{j-1} < z_j$. Then, according to (5.22), (5.5), (5.7), (5.23)

$$\begin{aligned} x^* &\le x = c_j - \xi_j \le c_j, \\ \delta_x &\le z \le \phi(x^*) + L(x - x^*) \le \phi(x^*) + L\Delta_x. \end{aligned} \tag{5.24}$$

If $z_{j-1} \ge z_j$, we consider two cases. In the first of them when $x^* > c_j$, we have

$$\begin{aligned} \delta_x &\le z \le \phi(x^*) + L(x - x^*) < \phi(x^*) + L(x - c_j) = \\ &= \phi(x^*) + L\xi_j \le \phi(x^*) + L\sigma\Delta_j \le \phi(x^*) + L\Delta_x, \end{aligned}$$

as far as $\sigma \le \frac{1}{6}$ and $\Delta_x \le (0.5 - \sigma)\Delta_j$.

In the second case when $x^* \le c_j$, we obtain

$$\begin{aligned} \delta_x &= 0.5(z_j + z - |z_j - z|) \le 0.5(z_j + z) \le \\ &\le 0.5(z_{j-1} + z) \le \phi(x^*) + 0.5L(x - x_{j-1}) \le \\ &\le \phi(x^*) + \frac{L(1 + 2\sigma)}{2(1 - 2\sigma)}\Delta_x \le \phi(x^*) + L\Delta_x. \end{aligned}$$

This inequality allows us to estimate characteristic R_x of the interval (x, x_j):

$$R_x > \alpha\Delta_x - \mu\delta_x + c \ge (\alpha - \mu L)\Delta_x - \mu\phi(x^*) + c$$

from where, taking into account (5.13), it comes out that

$$R_x > -\mu\phi(x^*) + c. \tag{5.25}$$

Inequality (5.25) is also true if $x_j \notin \{x^s\}$, since in this case, according to (5.8), $x = c_j$. As the Lipschitz condition is satisfied for $\phi(x)$, we have $z \leq \phi(x^*) + L\Delta_x$; therefore

$$R_x > \alpha\Delta_x - \mu z + c \geq (\alpha - \mu L)\Delta_x - \mu\phi(x^*) + c.$$

Thus, if after k trials there exists a pair of intervals of type (5.21), then the choice of two trials simultaneously in accordance with decision rule (2.1)–(2.8) and (5.4) cannot generate redundant trials. After K initial trials (see 5.11) during n iterations, the worst situation (in the sense of generating redundant trials during the next step) can be the situation when there exists only one interval satisfying condition (5.19): the interval of type (5.22). Consequently, one redundant point will be obtained at the $(n+1)$-th iteration.

The second trial of this iteration will result in the appearing of the pair (5.21). In this case, no other intervals for which inequality (5.19) is true can be outside the interval (5.22). This means that the next points $y^m \in \{x^k\}$ may lay only in the interval (5.22) which already contains one trial and that the $(n+2)$-th iteration will take place in it. If one of trials of the $(n+3)$-th pair falls into (5.22), it will result in the appearing of a new pair of type (5.21). Consequently, the source of redundant trials can only be the situation when the next pair contains points from (x, x_j) but does not contain points from (5.22). If it happens at the $(n+3)$-th iteration, then the $(n+4)$-th iteration can place one point in interval (5.22), and another one in some interval (x_{i-1}, x_i) whose characteristic satisfies condition (5.20). The last point will be redundant. It will be followed by creating the situation identical to the position after executing the $(n+1)$-th iteration.

Thus, not more than one redundant trial can be obtained during six trials. This means that inequality (5.14) is true. ■

COROLLARY 1. *If the objective function $\phi(x)$ has H global minimizers and between every pair of them at least one point from $\{y^m\}$ has been placed in the course of starting n iterations, then, given conditions of theorem 6, a parallel characteristical algorithm with $2H$ parallel processors provides fulfillment of (5.14).*

Proof. The proof is obvious and we omit it. ■

6. Numerical Experiments

As an illustration we adduce results of numerical experiments executed on an ALLIANT FX/80 Departmental Mini-Supercomputer [1] with 4 vector processors. Three series of experiments have been done for testing one and multidimensional parallel characteristical algorithms.

We consider the case when the time required for executing trials is much greater than all the other times required for the algorithms functioning. Thus, we can use the speed up in iterations calculated as

$$s(p) = n(1)p/n(p), p > 1,$$

for estimating efficiency of the methods. Here $n(1)$ is the number of trials done by a sequential characteristical algorithm (after executing the starting trials) for solving a problem and $n(p)$ is one for solving the same problem by the parallel method with p parallel processors for every iteration.

In the first case we use known test functions (see [12]) for testing one-dimensional parallel characteristical algorithms. We construct parallel characteristical algorithms on the basis of the purely sequential methods proposed by Kushner (Method 1), Piyavskii (Method 2). The parallel method proposed in [24] generalizing Strongin's algorithm is the third method tested. We have taken three functions from [12] and the following starting points for all the methods:

PROBLEM 1. Function 2, starting points 3, 4, 5.

PROBLEM 2. Function 9, starting points 5, 8, 11.

PROBLEM 3. Function 19, starting points 1, 2, 3.

In methods 1 and 2, starting trials have been also done at the margin points a and b of the corresponding search regions $[a, b]$. For these methods we have used the exact Lipschitz constants in opposite with method 3 where the adaptive estimate ω and the parameter $r = 2$ (see (2.12)–(2.15)) have been used. For all experiments we have taken accuracy $\varepsilon = 0.0001(b - a)$ in the stopping rule (3.11). Results of the experiments are shown in Table I. Global solutions have been found in all the cases. Sometimes for methods 1 and 3 we have speed up greater than the number of the parallel processors used. This situation is possible as in these algorithms the behavior of the objective functions is adaptively estimated. For example, if in method 3 the Lipschitz constant is estimated better by the parallel version, the search is accelerated. Note that the following situation can take place due to the stopping rule (3.11):

$$n(1) - n(p) < p, \quad p > 1, \tag{6.1}$$

See, for example, the results of method 2 for Problem 2.

The second and third series of the experiments deal with testing the multidimensional parallel characteristical algorithms created in accordance with the approach of Section 4. In the second series we maximize 100 two-dimensional functions

$$\Phi(z) = \left\{ \left(\sum_{i=1}^{7} \sum_{j=1}^{7} (A_{ij} g_{ij}(z) + B_{ij} h_{ij}(z) \right)^2 + \right.$$

Table I. Numerical results for univariate functions.

NP	P	Method 1		Method 2		Method 3	
		Trials	Speed up	Trials	Speed up	Trials	Speed up
1	1	4252	–	151	–	120	–
	2	4353	1.95	151	2.0	119	2.02
	3	4352	2.93	155	2.92	162	2.21
	4	4353	3.91	157	3.83	163	2.93
2	1	4622	–	124	–	125	–
	2	4621	2.0	123	2.02	125	2.0
	3	4622	3.0	122	3.05	126	2.98
	4	4373	4.24	125	3.97	127	3.94
3	1	2780	–	126	–	148	–
	2	2779	2.0	127	1.98	153	1.93
	3	2762	3.02	128	2.95	159	2.79
	4	2769	4.02	133	3.78	163	3.63

NP – number of problem
P – number of parallel processors

$$+\left(\sum_{i=1}^{7}\sum_{j=1}^{7}(C_{ij}g_{ij}(z)-D_{ij}h_{ij}(z)\right)^2\Bigg\}^{1/2},$$

where $z=(z_1,z_2)\in R^2, 0\le z_s\le 1, \quad s=1,2,$

$$g_{ij}(z)=\sin(i\pi z_1)\sin(j\pi z_2),$$

$$h_{ij}(z)=\cos(i\pi z_1)\cos(j\pi z_2),$$

$A_{ij}, B_{ij}, C_{ij}, D_{ij}$ are random coefficients uniformly distributed on the interval $[-1,1]$.

All the experiments have been executed with the accuracy $\varepsilon=0.001$ from (3.11) in Hölder's metrics. Two parallel characteristical algorithms have been tested. The first one is the parallel algorithm proposed in [22] where (as in Piyavskii's method) an estimate of the Lipschitz constant taken a priori has been used. This estimate has been obtained by evaluating the objective functions at 2000 points taken on a uniform mesh. Numerical results for this method (see Method 4) are shown in Table II. The global solutions have been found for 99 functions. In the unique case it has not been done by both the sequential and the parallel methods because the estimate was less than the real Lipschitz constant. Method 5 presented in Table II is the parallel information algorithm (see [25]). The global solutions have been found for all 100 functions.

Table II. Average results for 100 two-dimensional functions

Number of processors	Method 4		Method 5	
	Trials	Speed up	Trials	Speed up
1	1013.25	–	1575.12	–
2	1012.82	2.001	1596.08	1.974
3	1012.33	3.003	1562.61	3.024
4	1012.00	4.005	1599.92	3.938

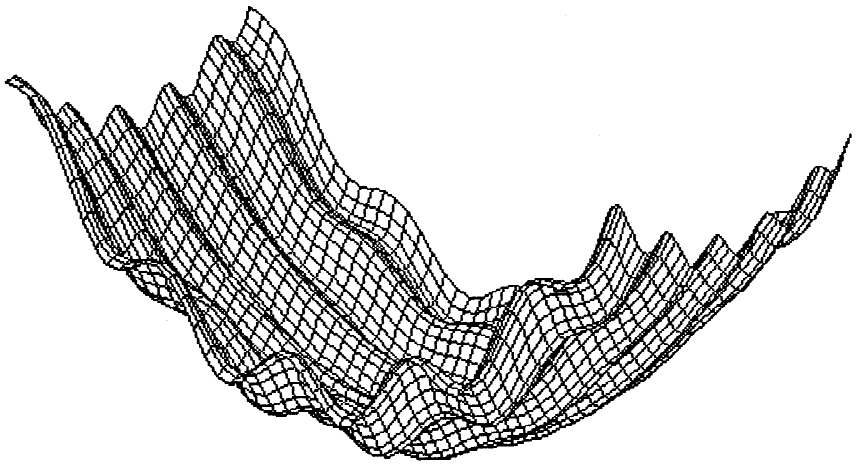

Figure 1. Two-dimensional section of the test function (6.2).

The third group of the test experiments consists in the minimization of the multidimensional functions

$$f(z) = (\pi/N)\{\sin^2(\pi y_1) + 5(y_N - 1)^2 + \\ + \sum_{i=1}^{N-1} [(y_i - 1)^2(1 + \sin^2(\pi y_{i+1}))]\} \tag{6.2}$$

where

$$y_i = 1 + 0.25(z_i - 1) \text{ and } -10 \le z_i \le 10, \quad 1 \le i \le N,$$

for the dimensions $N = 2, 3, 4$. This function is the modified test function from [15]. Figure 1 shows a two-dimensional section of the function (6.2) depending on 4 variables. This section corresponds to fixed coordinates $z_1 = -8, z_2 = 9$ and to free coordinates z_3 and z_4.

We have considered two characteristical algorithms: parallel technique [25] (Method 5) and the broken lines parallel algorithm generalized for the multidimensional optimization according to the scheme of Section 4 (see (4.14)) and further referred to as Method 6. For both the methods we have taken accuracy $\varepsilon = 10^{-5}$

Table III. Numerical results for Problem (6.2)

Dimension	Number of processors	Method 5		Method 6	
		Trials	Speed up	Trials	Speed up
2	1	1082	–	1027	–
	2	1084	1.996	1028	1.998
	3	1083	2.997	1050	2.934
	4	1136	3.810	1224	3.356
3	1	4784	–	4649	–
	2	4776	2.003	5018	1.853
	3	4776	3.005	4647	3.001
	4	4768	4.016	4480	4.151
4	1	7958	–	6274	–
	2	7876	2.021	5238	2.396
	3	6750	3.537	6486	2.902
	4	7820	4.072	6468	3.880

in (3.11) and have used the adaptive estimate of the Lipschitz constant $m = r\omega$, where ω from (2.15) and $r > 1$ is a parameter of methods. The initial stages of the search while solving the problem (4.5) were the same and included trials at the points 0.0, 0.2, 0.4, 0.6, 0.8, 1.0 of the interval [0,1]. Having taken into account the condition (4.9) we have chosen parameter $r = 3.5$ for the Method 5 and $r = 1.7$ for Method 6 in order to provide the sufficient conditions of convergence to global minima which have been found in all the searches.

The results of the experiments are contained in Table III. We have already noted above that adaptive estimation of the behavior of the function can lead to the situation when the speed up exceeds the number of parallel processors used (see Table III).

On the whole, the numerical experiments confirm the basic theoretical results of the paper concerning the convergence and efficiency of parallelization and demonstrate the applicability of the parallel characteristical algorithms for solving the multidimensional multiextremal problems.

7. Conclusion

In this paper a class of parallel characteristical algorithms for global optimization of one-dimensional multiextremal functions has been introduced. To illustrate the general approach examples of the algorithms have been presented. General conditions of “everywhere dense”, local and global convergence have been established for the class. Efficiency conditions for the parallel algorithms of the introduced class in comparison with their sequential versions have been obtained. A generalization for multidimensional case has been done. Numerical experiments executed

on a parallel computer with one and multidimensional methods on test functions taken from literature have been also presented.

Acknowledgements

The authors would like to thank the anonymous referees for their useful suggestions. The second author thanks the Systems Department, University of Calabria (Italy), where he was a visitor. The work was partially supported by the grant No. 95-01-01073 of the Russian Fund of Basic Researches.

References

1. Alliant. Product Summary (1989), Alliant Computer Systems Corporation, Littletown.
2. Archetti, F., and Betrò, B. (1979), A Probabilistic Algorithm for Global Optimization, *Calcolo* **16**(3), 335–343.
3. Avriel, M., and Wilde, D.J. (1966), Optimal Search for a Maximum with Sequences of Simultaneous Function Evaluations, *Management Science* **12**, 722–731.
4. Bertocchi, M. (1990), A Parallel Algorithm for Global Optimization, *Optimization* **62**(3), 379–386.
5. Butz, A.R. (1968), Space Filling Curves and Mathematical Programming, *Inform. Control* **12**(4), 314–330.
6. Chazan, D., and Miranker, W.L. (1970), A Non-gradient and Parallel Algorithm for Unconstrained Minimization, *SIAM Journal on Control and Optimization* **70**, 207–217.
7. Eriksen, O., and Staunstrup, J. (1983), Concurrent Algorithms for Root Searching, *Acta Informatica* **18**, 361–376, 501–511.
8. Evtushenko, Yu.G. (1971), Numerical Method of Global Extremum Search. *Journal of Computational Math. and Math. Physics* **11**(6), 1390–1403 (in Russian).
9. Gergel, V.P., Sergeyev, Ya.D., and Strongin, R.G. (1992), A Parallel Global Optimization Method and Its Implementation on a Transputer System, *Optimization* **26**, 261–275.
10. Grishagin, V.A. (1979), On Convergence Conditions for a Class of Global Search Algorithms. Theses of report to 3rd All-Union Seminar *Numerical Methods of Nonlinear Programming*, Kharkov, 82–84 (in Russian).
11. Grishagin, V.A. (1983), On Properties of One Class of Optimization Algorithms. Transactions of 3rd conference of young scientists of Applied Mathematics and Cybernetics Research Institute of Gorky University, Gorky, Nov. 24–26, 1982. Part 2, Gorky University, Gorky, 50–58. Deposited with VINITI, Aug. 14, 1984. No.5836-84 Dep.
12. Hansen, P., Jaumard, B., and Lu, S.-H. (1989), Global Optimization of Univeriate Lipschitz Functions: 2. New Algorithms and Computational Comparison, *Math. Programming* **55**, 273–293.
13. Horst, R., Pardalos, P.M. (Eds.) (1995), *Handbook on Global Optimization*, Kluwer Academic Publishers, Dordrecht.
14. Kushner, H. (1964), A New Method of Locating the Maximum Point of an Arbitrary Multipeak Curve in the Presence of Noise, *Transactions ASME, Ser. D, J. Basic Eng.* **86**(1), 97–106.
15. Lucidi S. and Piccioni M. (1989) Random tunneling by means of acceptance-rejection sampling for global optimization, *JOTA* **62**(2), 255–277.
16. Pardalos, P.M., Phillips, A.T., and Rosen, J.B. (1992), *Topics in Parallel Computing in Mathematical Programming*, Science Press.
17. Pintér, J. (1983), A Unified Approach to Globally Convergent One-dimensional Optimization Algorithms, *Technical report I AMI-83.5*, Istituto per le Applicazioni della Matematica e dell'Informatica CNR, Milan.
18. Pintér, J. (1992), Convergence Qualification of Adaptive Partition Algorithms in Global Optimization, *Math. Programming* **56**, 343–360.

19. Piyavskii, S.A. (1972), An Algorithm of Function Absolute Extremum Search, *Journal of Computational Math. and Math. Physics* **12**(4), 888–896 (in Russian).
20. Schendel, U. (1984), *Introduction to Numerical Methods for Parallel Computers*, Ellis Horwood, Chichester, England.
21. Sergeyev, Ya.D. (1995), An information global optimization algorithm with local tuning, *SIAM Journal on Optimization* **5**(4), 615–631.
22. Sergeyev, Ya.D., and Grishagin, V.A. (1994), A Parallel Method for Finding the Global Minimum of Univariate Functions, *JOTA* **80**(3), 513–536.
23. Sergeyev, Ya.D., and Grishagin, V.A. (1994), Sequential and Parallel Algorithms for Global Optimization, *Optimization Methods and Software* **3**, 111–124.
24. Sergeyev, Ya.D., and Strongin, R.G. (1989), An Algorithm of Global Minimization with Parallel Iterations, *Journal of Computational Math. and Math. Physics* **29**(2), 7–15.
25. Sergeyev, Ya.D., and Strongin, R.G. (1992), Global Multidimensional Optimization on Parallel Computer, *Parallel Computing* **18**, 1259–1273.
26. Strongin, R.G. (1978), *Numerical Methods in Multiextremal Problems (Information-Statistic Algorithms)*, Nauka, Moskow (in Russian).
27. Strongin, R.G., and Sergeyev, Ya.D. (1988), *Effective Algorithm for Global Optimization with Parallel Computations*, in Lecture Notes in Economics and Mathematical Systems, vol. 304, Springer, Berlin, 158–162.
28. Sutti, C. (1984), Local and Global Optimization by Parallel Algorithms for MIMD Systems, *Annals of Operations Research* **1**, 151–164.
29. Van Laarhoven, P.J.M. (1985), Parallel Variable Metric Algorithm for Unconstrained Optimization, *Mathematical Programming* **33**, 68–87.
30. Zhiglyavsky, A.A. (1991), *Theory of Global Random Search*, Kluwer Academic Publishers, Dordrecht.
31. Zilinskas, A.G. (1975), One-Step Bayesian Method for Finding the Extremum of One-dimensional Functions, *Cybernetics* **1**, 139–144 (in Russian).