J Glob Optim (2010) 48:99–112
DOI 10.1007/s10898-009-9508-x

An information global minimization algorithm using the local improvement technique

Daniela Lera · Yaroslav D. Sergeyev

Received: 11 December 2009 / Accepted: 14 December 2009 / Published online: 30 December 2009

Abstract In this paper, the global optimization problem with a multiextremal objective function satisfying the Lipschitz condition over a hypercube is considered. An algorithm that belongs to the class of information methods introduced by R.G. Strongin is proposed. The knowledge of the Lipschitz constant is not supposed. The local tuning on the behavior of the objective function and a new technique, named the *local improvement*, are used in order to accelerate the search. Two methods are presented: the first one deals with the one-dimensional problems and the second with the multidimensional ones (by using Peano-type space-filling curves for reduction of the dimension of the problem). Convergence conditions for both algorithms are given. Numerical experiments executed on more than 600 functions show quite a promising performance of the new techniques.

Keywords Global optimization · Local information · Acceleration

This research was supported by the following grants: Italian PRIN project 2007-9PLLN7, Russian RFBR grant 07-01-00467-a, the Russian Federal Program "Scientists and Educators in Russia of Innovations", contract number 02.740.11.5018, and the grant 4694.2008.9 for supporting the leading research groups awarded by the President of the Russian Federation.

D. Lera
Dipartimento di Matematica e Informatica, Università di Cagliari, Cagliari, Italy
e-mail: lera@unica.it

Y. D. Sergeyev (✉)
Dipartimento di Elettronica, Informatica e Sistemistica, Università della Calabria
and the Institute of High Performance Computing and Networking of the National Research
Council of Italy, Via Pietro Bucci 42C, 87036 Rende, CS, Italy
e-mail: yaro@si.deis.unical.it

Y. D. Sergeyev
Software Department, University of Nizhni Novgorod, Gagarin Av. 23, Nizhni Novgorod, Russia

1 Introduction

Let us consider the following global optimization problem looking for the value F^* and the corresponding point y^* such that

$$F^* = F(y^*) = \min\{F(y) : \ y \in [a, b]\}, \quad a, b \in \mathbf{R}^N, \tag{1}$$

where the function $F(y)$ is a multiextremal function that satisfies the Lipschitz condition over the hyperinterval $[a, b]$:

$$|F(y') - F(y'')| \leq L\|y' - y''\|, \quad y', y'' \in [a, b], \tag{2}$$

where the Lipschitz constant $0 < L < \infty$ is generally unknown and $\|\cdot\|$ denotes the Euclidean norm.

Many numerical algorithms have been proposed for solving the problem (1), (2) (see, e.g., [1,7,11,13,17,21–23]). In [18] and [19], R.G. Strongin has introduced a powerful information approach derived as optimal statistical decision rules within the framework of a stochastic model representing the function to be optimized as a sample of some random function. In [14], a modified information algorithm has been proposed where the *local tuning* technique for the approximation of the local Lipschitz constants has been introduced.

This technique consists both of an adaptive estimating local Lipschitz constants L_i related to subregions D_i of the domain $[a, b] \in \mathbf{R}^N$ and of a smart balancing the global and various kinds of the local information during the optimization process. As a result, this procedure performs a local tuning on the behavior of the objective function over every subregion of $[a, b]$. It has been proved in [9,15,16,21] that algorithms using the local tuning technique converge more rapidly to the desired solution with respect to the global optimization algorithms that use in their work only global (adaptive or a priori given) estimates of the value L from (2) valid for the entire region $[a, b]$.

In this paper, we present a new acceleration tool that can be used together with the local tuning technique from [14] in the framework of the information algorithms. The new approach (called hereinafter the *local improvement*) forces the information global optimization method to make a local improvement of the best approximation of the global minimum immediately after a new approximation better than the current one is found. In Sect. 2, the one-dimensional algorithm is presented and sufficient convergence conditions are proved for it. In Sect. 3, the multidimensional algorithm based on the Peano-type space filling curves and its convergence conditions are described. It is worthy to emphasize that not only the global search works on the curve but the local improvement technique is executed on the curve, as well. Section 4 contains results of numerical experiments executed on more than 600 test functions taken from the literature. Finally, Sect. 5 concludes the paper.

2 The one-dimensional algorithm

Let us start by considering the problem (1), (2) for a function $f(x)$ in dimension $N = 1$, i.e.,

$$\min\{f(x) : \ x \in [a, b]\}, \quad a, b \in \mathbf{R}^1, \tag{3}$$

where

$$|f(x') - f(x'')| \leq L|x' - x''|, \quad x', x'' \in [a, b]. \tag{4}$$

We denote later on the one-dimensional version of the new algorithm by ODII (One-Dimensional Information algorithm with the local Improvement). By the term *trial* we denote the evaluation of the function $f(x)$ at a point x which is refereed to as the *trial point*.

Algorithm ODII.

STEP 0. Starting trial points $x^1, x^2, \ldots, x^m, m > 2$, are fixed in such a way that $x^1 = a, x^m = b$ and the other $m-2$ points are chosen arbitrarily. Values $f(x^1), \ldots, f(x^m)$ are calculated at these points. The point $x^{k+1}, k \geq m$, of the current $(k+1)$-th iteration is chosen as follows.

STEP 1. Renumber the trial points $x^1, x^2, \ldots, x^k$ of the previous iterations by subscripts so that

$$a = x_1 < x_2 < \ldots x_{k-1} < x_k = b. \tag{5}$$

STEP 2. Compute the value μ_i being an estimate of the Lipschitz constant of $f(x)$ over the interval $[x_{i-1}, x_i], i = 2, \ldots, k$, as follows

$$\mu_i = \max\{\lambda_i, \gamma_i, \xi\}, \quad i = 2, \ldots, k, \tag{6}$$

with

$$\lambda_i = \max\{\Lambda_{i-1}, \Lambda_i, \Lambda_{i+1}\}, \quad i = 2, \ldots, k-1, \tag{7}$$

where

$$\Lambda_j = \frac{|z_j - z_{j-1}|}{|x_j - x_{j-1}|}, \quad j = 2, \ldots, k, \tag{8}$$

and when $i = 1$ and $i = k$ consider only Λ_1, Λ_2, and Λ_{k-1}, Λ_k, respectively. The values $z_j = f(x_j), j = 1, \ldots, k$ and the value

$$\gamma_i = \Lambda^k |x_i - x_{i-1}| / X^{\max}, \tag{9}$$

where

$$\Lambda^k = \max\{\Lambda_i : i = 2, \ldots, k\}$$

and

$$X^{\max} = \max\{|x_i - x_{i-1}|, \quad i = 2, \ldots, k\}. \tag{10}$$

The parameter $\xi > 0$ is a small number that takes into account our hypothesis that the objective function $f(x)$ is not constant over the interval $[a, b]$.

STEP 3. For each interval $[x_{i-1}, x_i], i = 2, \ldots, k$, compute characteristics

$$R_i = r\mu_i(x_i - x_{i-1}) + \frac{(z_i - z_{i-1})^2}{r\mu_i(x_i - x_{i-1})} - 2(z_i + z_{i-1}), \tag{11}$$

where $r > 1$ is a reliability parameter of the method.

STEP 4. Select the interval $[x_{t-1}, x_t]$ for the next trial using the following designations:

$imin$ - the index corresponding to the current minimal value of the function;
z^k - the result of the last trial corresponding to a point x_j in (5), i.e., $x^k = x_j$;
$flag$ - a parameter initially equal to zero.

IF (flag=1)

IF $z^k < z_{imin}$ THEN $imin = j$.
Local improvement: Alternate the choice of the interval $[x_{t-1}, x_t]$ among $t = imin + 1$ and $t = imin$, if $imin = 2, \ldots, k-1$, (if $imin = 1$ or $imin = k$ take $t = 2$ or $t = k$, respectively) in such a way that for $\delta > 0$ it follows

$$|x_t - x_{t-1}| > \delta. \tag{12}$$

ELSE (flag=0)

$$t = argmax\{R_i \ : \ 2 \le i \le k\}. \tag{13}$$

ENDIF
flag=NOTFLAG(flag)

STEP 5. If the length of the interval to be subdivided is such that

$$|x_t - x_{t-1}| > \varepsilon, \tag{14}$$

where $\varepsilon > 0$ is a given search accuracy, then execute the next trial at the point

$$x^{k+1} = 0.5\left(x_t + x_{t-1} - \frac{z_t - z_{t-1}}{r\mu_t}\right) \tag{15}$$

and go to STEP 1. Otherwise, calculate the following estimate of the minimum as

$$f_k^* = \min\{z_i \ : \ 1 \le i \le k\}$$

and STOP.

Let us give some comments upon the scheme introduced above. In STEP 2, the local tuning technique is applied. The local Lipschitz constant, related to the i-th subinterval, is approximated by the value μ_i on the basis of the local estimate, λ_i, and the global one, γ_i, which controls authenticity of the local information in consideration. If the current subinterval, $[x_{i-1}, x_i]$, is "small", the local component λ_i becomes relevant; if the subinterval is "large" than the local information is not reliable and the global estimate, γ_i, is applied. Formula (9) is used to balance the impact of the global information on the estimate μ_i. The parameter ξ is introduced in order to avoid an estimate of the local Lipschitz constant equal to zero.

In STEP 3, the characteristic R_i can be interpreted (after normalization) as the probability that a global minimizer is located within the interval $[x_{i-1}, x_i]$, in the course of the $(k+1)$th iteration. Then, STEP 4 describes the local improvement technique choosing a subinterval for the next trial. In the original information algorithms [14,19] only criterion (13) is used, i.e., an interval corresponding to the maximal characteristic is selected at each iteration k. In the ODII, we alternate this choice with the local improvement technique that forces the global optimization method to make a local improvement of the best approximation of the global minimum found up to now. The parameter *flag* is used in order to alternate the two methods.

More precisely, when $flag = 1$, first we identify the index $imin$ corresponding to the current minimum among the found values of the objective function, and then we select the interval on the right of the current point of minimum x_{imin}, or the interval on the left of x_{imin}. The method keeps working alternatively to the right and the left of the current x_{imin} until a new trial point with value less than z_{imin} is found. The parameter δ defines the width of the

intervals subdivided during the phase of the local improvement. Since it is not sure that the found best approximation is really located in the neighborhood of the global minimizer, when $flag = 0$ the criterion (13) is used in order to turn the search to new subregions possibly containing the global solution.

Let us consider now convergence properties of the introduced algorithm. We study properties of an infinite sequence $\{x^k\}$, $x^k \in [a, b]$, $k > 1$, of trial points generated by the ODII algorithm in the course of minimization of $f(x)$.

Definition 1 Convergence to a point $x' \in (a, b)$ is said bilateral if there exist two subsequences of $\{x^k\}$ converging to x' one from the left, the other from the right.

Theorem 1 *Assume that the objective function $f(x)$ satisfies the condition (4), and let x' be any limit point of $\{x^k\}$ generated by the ODII. Then the following assertions hold:*

1. *convergence to x' is bilateral, if $x' \in (a, b)$;*
2. *$f(x^k) \geq f(x')$, for all trial points x^k, $k \geq 1$;*
3. *if there exists another limit point x'', $x'' \neq x'$, then $f(x'') = f(x')$;*
4. *if the function $f(x)$ has a finite number of local minima in $[a, b]$, then the point x' is locally optimal.*

Proof The proofs of assertions 1–4 are completely analogous to the proofs of Lemmas 1.1–1.4 from [14]. □

Theorem 2 (Sufficient conditions of global convergence). *Let $\delta > 0$ be a finite constant and $\varepsilon = 0$, where δ is from (12) and ε from (14). Let x^* be a global minimizer of $f(x)$ and $\{k\}$ be the sequence of all iteration numbers $\{k\} = \{1, 2, 3, \ldots\}$. If there exists an infinite subsequence $\{h\}$ of iteration numbers $\{h\} \subset \{k\}$ such that for the interval*

$$[x_{i-1}, x_i], \quad i = i(l), \quad l \in \{h\},$$

containing the point x^* at the l-th iteration, the inequality

$$r\mu_i \geq K_i + \sqrt{K_i^2 - M_i^2} \tag{16}$$

holds, then the set of limit points of the sequence $\{x^k\}$ coincides with the set of global minimizers of the function $f(x)$. In (16), r is from (11), μ_i is from (6), and

$$K_i = \max\left\{\big(z_{i-1} - f(x^*)\big)/(x^* - x_{i-1}), \big(z_i - f(x^*)\big)/(x_i - x^*)\right\},$$
$$M_i = |z_{i-1} - z_i|/(x_i - x_{i-1}).$$

Proof Since δ is a finite positive number and $\varepsilon = 0$, the algorithm uses the local improvement technique only at the initial stage of the search until the selected interval $[x_{t-1}, x_t]$ is greater than δ. When $|x_t - x_{t-1}| \leq \delta$ the interval cannot be divided by the local improvement procedure and the criterion (13) is used. Thus, since the one-dimensional search region has a finite length and δ is a fixed finite number, there exists a finite iteration number j such that at all iterations $k > j$ only criterion (13) will be used. It follows that, in order to complete the proof, it is sufficient to prove the assertions of the theorem by supposing that only the criterion (13) is considered in the course of the optimization. This is done by a complete analogy to the proof of Theorem 1.1 from [14]. □

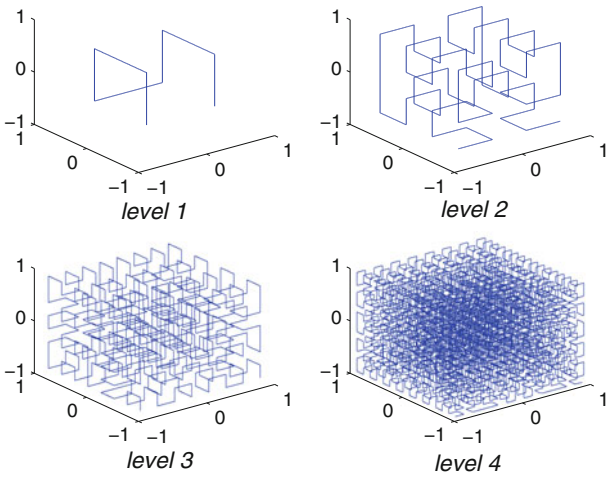

Fig. 1 Approximations of levels 1–4 of the three-dimensional Hilbert curve

3 The multidimensional algorithm

Let us consider now the multidimensional case, i.e., the problem (1), (2) with $N \geq 2$. The approach based on the reduction of the dimension by using Peano–Hilbert curves (see [2,12,18,21]) is applied. These curves are fractals generated by an iterative process, that fill in the hypercube $[a, b]$ in $\mathbf{R}^N$, i.e., they pass through every point of $[a, b]$. In Fig. 1 approximations of levels 1, 2, 3, and 4 of the Hilbert curve are given in the domain $[-1, 1] \in \mathbf{R}^3$.

It has been shown (see [2,18–21]) that, by using space filling curves, the multidimensional global minimization problem (1), (2) is turned into a one-dimensional problem. More precisely, Strongin proved that the global solution to the problem (1), (2) with the Lipschitz function $F(y)$, $y \in \mathbf{R}^N$, over a hypercube, may be obtained by minimizing the function $f(x)$ so defined

$$f(x) = F(p(x)), \quad x \in [0, 1], \tag{17}$$

where $p(x)$ is the Peano–Hilbert curve. Moreover:

$$|f(x') - f(x'')| \leq H|x' - x''|^{\frac{1}{N}}, \quad x', x'' \in [0, 1], \tag{18}$$

i.e., $f(x)$ is a Hölderian function, in the Hölder metric, and $H = 2L\sqrt{N+3}$, where L is the Lipschitz constant of the multidimensional function $F(y)$. Let us describe now the Multi-Dimensional Information algorithm with the local Improvement (MDII) for minimizing the function $f(x)$ from (17) over the interval $[0, 1]$.

Algorithm MDII.

STEP 0. Starting points $x^1, x^2, \ldots, x^k$, $k > 2$, are fixed in such a way that $x^1 = 0$, $x^k = 1$ and the other $k - 2$ points are chosen arbitrarily. Values $z^j = f(x^j) = F(p_m(x^j))$, $j = 1, \ldots, k$, are calculated, where $p_m(x)$ is the m-approximation of the Hilbert curve. After executing k trials the choice of the new trial point is done as follows.

STEP 1. Renumber the trial points $x^1, x^2, \ldots, x^k$ of the previous iterations by subscripts so that

$$0 = x_1 < x_2 < \ldots x_{k-1} < x_k = 1. \tag{19}$$

STEP 2. Evaluate the values μ_i according to (6) replacing $(x_i - x_{i-1})$ by $(x_i - x_{i-1})^{1/N}$ in (8), (9) and $X^{\max}$ by $(X^{\max})^{1/N}$ in (9). The values $f(x_j)$ are replaced by $F(p_m(x_j))$.

STEP 3. For each interval $[x_{i-1}, x_i]$, $i = 2, \ldots, k$, calculate characteristics R_i according to (11), replacing $(x_i - x_{i-1})$ by $(x_i - x_{i-1})^{1/N}$.

STEP 4. Execute STEP 4 of the ODII to select the index t.

STEP 5. If

$$|x_t - x_{t-1}|^{1/N} \leq \varepsilon \tag{20}$$

where $\varepsilon > 0$ is a given search accuracy, then calculate the following estimate of the global minimum of the objective function $F(y)$:

$$F_k^* = \min\{z_i \ : \ 1 \leq i \leq k\}$$

and STOP. Otherwise, execute the next trial at the point

$$x^{k+1} = 0.5(x_t + x_{t-1}) - (|z_t - z_{t-1}|/\mu_t)^N (2r)^{-1} \mathrm{sign}(z_t - z_{t-1}) \tag{21}$$

and go to STEP 1.

Let us study convergence conditions of the method introduced above.

Theorem 3 *Assume that the objective function $f(x)$ satisfies the condition (18), then the results described in Theorem 1 for the ODII take place for the MDII, as well.*

Proof The corresponding results for the MDII are obtained by repeating the proof of Theorem 1 introducing in the formulae (see [14]) the same changes that have been done in STEP2–STEP5 to pass from the ODII to the MDII. □

The following theorem generalizes Theorem 2 to the multidimensional case.

Theorem 4 *Let $\delta > 0$ be fixed finite, and $\varepsilon = 0$, where δ is from (12) and ε is from (20). Let x^* be a global minimizer of $f(x)$ and $\{k\}$ be the sequence of all iteration numbers $\{k\} = \{1, 2, 3, \ldots\}$. If there exists an infinite subsequence $\{h\}$ of iteration numbers $\{h\} \subset \{k\}$ such that for an interval*

$$[x_{i-1}, x_i], \quad i = i(l), \ l \in \{h\},$$

containing the point x^* at the l-th iteration, the inequality

$$r\mu_i \geq 2^{1-1/N} K_i + \sqrt{2^{2-2/N} K_i^2 - M_i^2} \tag{22}$$

holds, then the set of limit points of the sequence $\{x^k\}$ coincides with the set of global minimizers of the function $f(x)$. In (22), the values K_i and M_i are the following:

$$K_i = \max\left\{ \big(z_{i-1} - f(x^*)\big)/(x^* - x_{i-1})^{1/N}, \big(z_i - f(x^*)\big)/(x_i - x^*)^{1/N} \right\},$$
$$M_i = |z_{i-1} - z_i|/(x_i - x_{i-1})^{1/N}.$$

Proof The proof follows from the proof of Theorem 2 in Sect. 2, and the proof of Theorem 2.1 from [14]. □

Theorem 5 *For every function $f(x)$ satisfying (18) with $H < \infty$ there exists r^* such that for all $r > r^*$ the algorithm MDII determines all global minimizers of the function $f(x)$ over the search interval* $[0, 1]$.

Proof Since $H < \infty$ and any value of r can be chosen in the MDII, it follows that there exists r^* such that condition (22) will be satisfied for all global minimizers for $r > r^*$. This fact taken together with Theorem 4 concludes the proof. □

To conclude this section, let us comment upon a practical connection between convergence to a global minimizer x^* of the reduced problem (17), (18) and convergence to the solution y^* of the original problem (1), (2). Since the global minimizer in the N-dimensional space can have up to 2^N images on the curve (for more details see [14]) and in the process of optimization a curve $p_m(x)$ is used, in order to have convergence to the global minimizer y^* from (1) it is sufficient to have convergence to one of the images of y^* on the curve $p_m(x)$. Of course, in the limit case ($m \to \infty$ and $\varepsilon = 0$ in (20)) if condition (22) is satisfied for one of the images, all global minimizers will be found. However, in practice we work with a finite $m < \infty$ and $\varepsilon > 0$, i.e., with a finite trial sequence, the search can stop after finding the only image of y^* providing the required approximation of the global minimizer y^* and accelerating the search. For more details on this kind of convergence see [14] and [21].

4 Numerical experiments

In this section, numerical experiments containing comparison of the algorithms ODII and MDII with other global optimization methods taken from the literature. Three series of experiments have been executed in dimension $N = 1, 2, 3, 4$.

In the first series of experiments, we compare the performances of the one-dimensional algorithm ODII with the following global optimization methods: Kushner [8], Evtushenko [3], Pijavskii [10], Strongin [18,19], and the information algorithm from [14]. These algorithms are denoted in Table 1 as KM, EA, PA, SM, and ODV, correspondingly. All the methods considered for the comparison do not use derivatives. We emphasize that the local tuning technique has been used in both ODII and ODV algorithms. However, only the ODII uses the local improvement technique.

A set of 20 test functions proposed by Hansen, Jaumard and Lu [6] has been considered. The experiments have been performed by using the accuracy $\varepsilon = 0.0001$ from (14). In the methods KM, EA, and PA, the precise values of Lipschitz constant have been used. In the method KM, the parameter γ has been chosen equal to 1, and in the methods SM and ODV, the parameter $r = 2$ has been chosen. To choose the parameter r in the ODII we should use (16). Taking into account the fact that M_i may be equal to zero and μ_i can be close to K_i, we must choose $r \geq 2$. In our experiments, we have used $r = 2$. Finally, we have fixed the parameter $\xi = 10^{-8}$ from (6).

All the global minimizers have been found by all the methods except KM for the function 17. It can be seen from Table 1 that the number of trials executed by the ODII is on average very low with respect to other methods under consideration. In particular, it can be seen that only the ODV was able to produce a smaller number of trials and only once (see results for the function 7).

The remaining series of experiments deal with the multidimensional algorithms and involve a total of 600 test functions in dimensions $N = 2, 3, 4$. More precisely, six classes of 100 functions each have been considered. They have been generated by the GKLS-generator described in [4] (and downloadable for free from

Table 1 Numerical results for one-dimensional algorithms

Function	KM	EA	PA	SM	ODV	ODII
1	2,327	4,363	149	127	35	34
2	4,693	1,205	155	135	36	36
3	416	373	195	224	136	40
4	1,241	2,559	413	379	41	40
5	4,153	607	151	126	45	42
6	4,425	2,146	129	112	54	40
7	4,281	1,560	153	115	39	40
8	355	389	185	188	132	34
9	3,908	1,068	119	125	42	38
10	1,488	1,887	203	157	40	38
11	6,401	522	373	405	71	36
12	5,633	1,787	327	271	68	40
13	2,289	3,809	993	472	45	32
14	5,377	347	145	108	46	36
15	6,067	1,251	629	471	63	38
16	1,638	3,953	497	557	53	38
17	529	951	549	470	101	48
18	5,211	1,762	303	243	41	40
19	2,252	2,054	131	117	34	32
20	3,385	2,545	493	81	42	38
Average	3,303.45	1,756.90	314.60	244.15	58.20	38.00

http://wwwinfo.deis.unical.it/~yaro/GKLS.html). It allows one to construct classes of multidimensional and multiextremal test functions with known values and locations of local and global minima. The generator works by constructing in $\mathbf{R}^N$ a convex quadratic function, i.e., a paraboloid, systematically distorted by polynomials. Each test class includes 100 functions and is defined by the following parameters:

N - problem dimension;
M - number of local minima;
f^* - value of the global minima;
ρ^* - radius of the attraction region of the global minimizer;
r^* - distance from the global minimizer to the vertex of the paraboloid.

In our numerical experiments, we have considered classes of continuously differentiable test functions with $M = 10$ local minima. The global minimum value f^* has been fixed equal to -1.0 for all classes.

By changing the user-defined parameters, classes with different properties can be created. For example, a more difficult test class can be obtained either by decreasing the radius ρ^* of the attraction region of the global minimizer or by increasing the distance, r^*, from the global minimizer to the paraboloid vertex. In this paper, for each dimension $N = 2, 3, 4$, two test classes where considered: a simple one and a difficult one. Table 2 describes all the classes.

Table 2 Description of GKLS classes used in the experiments

Class		N	M	f^*	r^*	ρ^*
1	Simple	2	10	−1.0	0.66	0.33
2	Hard	2	10	−1.0	0.90	0.20
3	Simple	3	10	−1.0	0.66	0.33
4	Hard	3	10	−1.0	0.66	0.20
5	Simple	4	10	−1.0	0.66	0.33
6	Hard	4	10	−1.0	0.66	0.20

The numerical experiments have been carried out by using one of the following two stopping rules:

a. A value $\varepsilon > 0$ is fixed and the search terminates when the rule (20) is satisfied; then it is counted the number of functions of the class for which the method under consideration was able to put a point in the ball

$$B_i = \left\{ y \in \mathbf{R}^N : \|y - y_i^*\| \le d \right\}, \quad i = 1, \ldots, 100, \tag{23}$$

that is the ball having the center at the global minimizer y_i^* of the i-th considered function of the class, and a fixed radius $d > 0$.

b. The value $\varepsilon = 0$ is fixed in the stopping rule (20) and the search terminates when a trial point falls in the ball B_i, $i = 1, \ldots, 100$.

In the numerical experiments three algorithms have been compared with respect to several criteria: the original information method of Strongin (see [18,19]) that uses neither the local tuning nor the local improvement techniques; the information algorithm from [14]; and the new MDII method. Let T_s be the number of trials performed by a method to solve the problem number $s = 1, \ldots, 100$, from a fixed test class. Then we can start to introduce the criteria.

Criterion C1. The number of trials T_{s^*} required for a method to satisfy the fixed stopping criterion, for all 100 function of a particular test class, i.e.,

$$T_{s^*} = \max_{1 \le s \le 100} T_s, \quad s^* = \arg \max_{1 \le s \le 100} T_s. \tag{24}$$

Criterion C2. The average number of trials T_{avg} performed by the method during minimization of all 100 functions from a particular test class, i.e.,

$$T_{\text{avg}} = \frac{1}{100} \sum_{s=1}^{100} T_s. \tag{25}$$

Note that results based on Criterion C1 are mainly influenced by minimization of the most difficult functions of a class (that can be different for different methods); Criterion C2 considers average data of a class.

Let us describe the second series of experiments that has been performed in order to study the influence of the local improvement strategy in the situation when the stopping rule a) is used. In fact, in solving real-life problems we do not know a priori the global solution, thus it is very important to study how many trials should execute the methods to find the solution and to stop when the practical criterion a) is satisfied.

In Table 3, we summarize results of the numerical experiments. We denote the Strongin method with SM and the information algorithm from [14] with MDV. The "C1" columns show how many trials it is necessary to execute to solve all the problems of a class. The "C2"

Table 3 Results of the second series of numerical experiments

Class	d	C1			C2		
		SM	MDV	MDII	SM	MDV	MDII
1	$.01\sqrt{N}$	1,219	1,065	724	710.39	332.48	354.82
2	$.01\sqrt{N}$	4,678	2,699	2,372	1,705.89	956.67	953.58
3	$.01\sqrt{N}$	22,608	10,800	8,459	8,242.19	2,218.82	2,312.93
4	$.01\sqrt{N}$	70,492	47,456	37,688	20,257.50	12,758.14	11,505.38
5	$.02\sqrt{N}$	100,000(53)	100,000(2)	100,000(1)	83,362.00	23,577.91	23,337.03
6	$.02\sqrt{N}$	100,000(96)	100,000(24)	100,000(27)	99,610.97	61,174.69	61,900.93

columns give the average number of trials of the considered class. The parameter d is from (23). The value of ε from (20) has been fixed equal to 10^{-3} and the parameter $\xi = 10^{-8}$, where ξ is from STEP 2 of the MDII. The same value of ξ has been used in the MDV and the SM, as well.

The choices of the reliability parameter r from (22) for all the methods are given in "Appendix". Due to Theorem 5, every function optimized by the MDII has a crucial value r^* of this parameter. The same situation takes place for both the SM and MDV algorithms (see [21]). Since the algorithm SM uses a global estimate of the Lipschitz constant, the value r^* for this method is less variable for different functions of a fixed class. The algorithms MDII and MDV have been constructed in order to be tuned on each concrete function. Therefore, when one executes tests with a class of 100 different functions it becomes difficult to use specific values of r for each function and in our experiments only one or two values of this parameter have been fixed for the entire class. Clearly, such a choice does not allow the algorithms MDII and MDV to show their complete potential in the comparison with the SM. However, as it can be seen from Table 3, even under these unfavorable conditions, the algorithms show a very nice performance.

In the results described in Table 3, all the algorithms were able to find the solution to all 100 functions of each class. It can be seen that the MDV and the MDII were very fast to stop, whereas the SM executed a deeper global analysis of the whole domain of each objective function so that the stopping rule (20) was verified after a higher number of trials. In all the cases, the maximal number of function evaluations has been taken equal to 100,000 and in Table 3, in the C1 columns, the numbers in brackets present the number of functions for which the algorithm has reached this number.

In the third series of experiments, the efficiency of the local improvement technique was studied. For this purpose, the algorithm MDII has been compared with the algorithm MDV by using the stopping strategy b), i.e., the search went on until a point within the ball B_i has been placed. In solving many concrete problems very often it is crucial to find a good approximation of the global minimum in the lowest number of iterations. The most important aim of the local improvement is that of quicken the search: thus, the use of the stopping criterion b) that allows us to see which of the two methods faster approaches the global solution. In these experiments, we considered the criteria C1 and C2 previously described, and a new criterion defined as follows.

Criterion C3. The number p (number q) of functions from a class for which the MDV algorithm has executed less (more) function evaluations than the new algorithm MDII. If T_s is the number of trials performed by the MDII and $T_{s'}$ is the corresponding number of trials

Table 4 Results of the third series of numerical experiments

Class	C1		C2		C3		Ratio C1	Ratio C2
	MDV	MDII	MDV	MDII	MDV	MDII	MDV/MDII	MDV/MDII
1	668	434	153.72	90.73	20	79	1.5391	1.6942
2	1,517	1,104	423.39	198.82	22	77	1.3741	2.1295
3	7,018	5,345	1,427.06	838.96	25	75	1.3130	1.7008
4	40,074	15,355	6,162.02	2,875.06	25	75	2.6098	2.1432
5	67,017	36,097	10,297.14	6,784.37	36	64	1.8565	1.5177
6	76,561	73,421	21,961.91	16,327.21	40	60	1.0427	1.3451

performed by the MDV, then p and q are calculated as follows:

$$p = \sum_{s=1}^{100} \delta'_s, \quad \delta'_s = \begin{cases} 1, & T_{s'} < T_s, \\ 0, & \text{otherwise;} \end{cases} \tag{26}$$

$$q = \sum_{s=1}^{100} \delta_s, \quad \delta_s = \begin{cases} 1, & T_s < T_{s'}, \\ 0, & \text{otherwise.} \end{cases} \tag{27}$$

If $p + q < 100$, then both the methods solve the remaining $100 - (p + q)$ problems with the same number of function evaluations.

Table 4 presents results of numerical experiments in the third series. The "C1" and "C2" columns have the same meaning as before. The "C3" columns describe the results of comparison between the two methods in terms of this criterion: the MDV sub-column presents the number of functions, p, of a particular test class, for which MDV spent fewer trials than the new method. Analogously, the MDII sub-column shows the number of functions, q, for which the MDII executed less function evaluations with respect to the MDV (p and q are from (26) and (27), respectively). For example, in the line corresponding to the test class 1, for $N = 2$, we can see that the new method MDII was better (was worse) than the MDV on $q = 79$ ($p = 20$) functions, and for one function of this class the two methods generated the same number of function trials.

In all the cases, the maximal number of function evaluations has been taken equal to 100,000. The parameters d, ξ, and δ and the values of the reliability parameter r used in these experiments for the MDV and MDII methods are the same as in the second series of experiments.

It can be seen from Table 4 that on these test classes the new method MDII worked better than the information algorithm MDV. In particular, the columns "Ratio C1" and "Ratio C2" of Table 4 show the improvement obtained by the MDII with respect to Criteria C1 and C2. They represent the ratio between the maximal (and the average) number of trials performed by the MDV with respect to the corresponding number of trials performed by the new algorithm.

Figure 2 illustrates the behavior of the two methods for classes 1 and 4, respectively. The graphs describe the dynamic of the search inside each of the classes. For instance, it can be seen in Fig. 2-left that after 100 function evaluations the MDV has found the solution at 33 problems and the MDII at 68 problems.

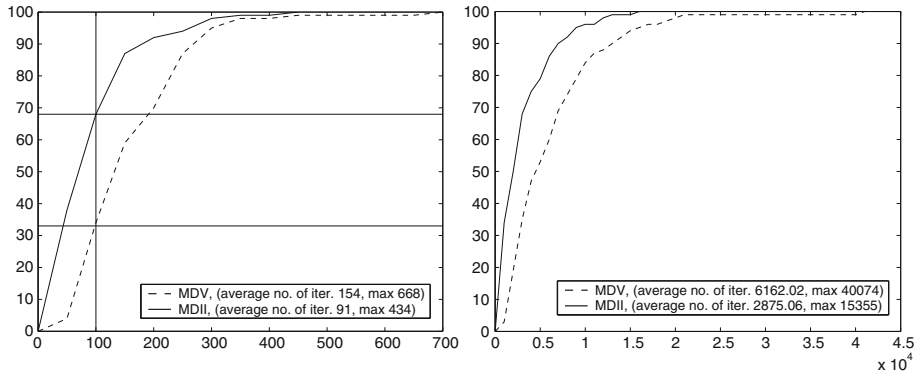

Fig. 2 Methods MDV and MDII, $N = 2$, class 1, *left*; $N = 3$, class 4, *right*

5 A brief conclusion

In this paper, the global optimization problem where the objective function satisfies the Lipschitz condition over a closed interval has been considered. For solving this problem, Strongin has introduced the powerful information approach (see [18] and [19]). It has been then shown (see [14] and [21]) that information algorithms can be accelerated by introducing the local tuning technique in the original information algorithms scheme. In this paper, the newly introduced local improvement strategy has been used, together with the local tuning, to accelerate the search. Two algorithms, the ODII and the MDII, have been described and their convergence conditions have been established. The methods have been tested on more than 600 functions taken from the literature. Numerical experiments have shown quite a satisfactory performance of the new information algorithms emphasizing that the usage of various kinds of a local information together with its accurate balancing with the global information can accelerate the global search significantly.

Appendix

The following values of the reliability parameter r have been used for the methods in the second series of experiments: for the test class 1 the value $r = 4.9$ in the MDV and MDII algorithms and the value $r = 3.1$ in the SM algorithm. For the class 2 the value $r = 5.4$ was used in the MDV and the MDII for 97 functions, and $r = 5.5$ for the remaining 3 functions of this class; in the SM the values $r = 4.1$ and $r = 4.3$ were used for 97 and 3 functions of the same class, respectively.

In dimension $N = 3$, the values $r = 5.5$ and $r = 5.7$ were applied in the MDV and the MDII for 97 and 3 functions of the class 3, respectively; the values $r = 3.2$ and $r = 3.4$ for 97 and 3 functions of this class when the SM algorithm has been used. By considering the test class 4 the following values of the parameter r have been used: $r = 6.5$ and $r = 6.6$ in the MDV and the MDII for 99 and 1 function, respectively; $r = 3.8$ for 99 functions in the SM and $r = 4.1$ for the remaining function.

In dimension $N = 4$, the value $r = 6.2$ was used for all 100 functions of test class 5 in the MDV and the MDII, and $r = 3.3$, $r = 3.5$ in the SM, for 96 and 4 functions, respectively. The value $r = 6.2$ was applied for 92 functions of test class 6 in the MDV and the MDII, and

the values $r = 6.6$ and 6.8 were used for 5 and 3 functions, respectively; in the SM algorithm the value $r = 3.8$ has been used for 98 functions of the class 6 and $r = 4.1$ for the remaining 2 functions.

Finally, the parameter δ from STEP4 of the MDII has been fixed equal to 10^{-6} for $N = 2$, and equal to 10^{-8} for $N = 3, 4$.

References

1. Bomze, I.M., Csendes, T., Horst, R., Pardalos, P.M.: Developments in Global Optimization. Kluwer, Dordrecht (1997)
2. Butz, A.R.: Space filling curves and mathematical programming. Inf. Contr. **12**(4), 313–330 (1968)
3. Evtushenko, Yu.G.: Numerical methods for finding global extrema of a nonuniform mesh. USSR Comput. Math. Math. Phys. **11**, 1390–1403 (1971)
4. Gaviano, M., Kvasov, D.E., Lera, D., Sergeyev, Ya.D.: Software for generation of classes of test functions with known local and global minima for global optimization. ACM TOMS **29**(4), 469–480 (2003)
5. Grishagin, V.A.: Operation characteristics of some global optimization algorithms. Prob. Stoch. Search **7**, 198–206 (In Russian) (1978)
6. Hansen, P., Jaumard, B., Lu, S.H.: Global optimization of univariate Lipschitz functions: 2. New algorithms and computational comparison. Math. Program. **55**, 273–293 (1992)
7. Horst, R., Pardalos, P.M.: Handbook of Global Optimization. Kluwer, Dordrecht (1995)
8. Kushner, H.: A new method for locating the maximum point of an arbitrary multipeak curve in presence of noise. J. Basic Eng. **86**, 97–106 (1964)
9. Lera, D., Sergeyev, Ya.D.: Global minimization algorithms for Holder functions. BIT **42**(1), 119–133 (2002)
10. Pijavskii, S.A.: An algorithm for finding the absolute extremum of a function. USSR Comput. Math. Math. Phys. **12**, 57–67 (1972)
11. Pintér, J.: Global Optimization in Action (Continuous and Lipschitz Optimization: Algorithms, Implementations and Applications). Kluwer, Dordrecht (1996)
12. Platzman, L.K., Bartholdi, J.J. III.: Spacefilling curves and the planar travelling salesman problem. J. ACM **36**(4), 719–737 (1989)
13. Ratscek, H., Rokne, J.: New Computer Methods for Global Optimization. Ellis Horwood, Chichester (1988)
14. Sergeyev, Ya.D.: An information global optimization algorithm with local tuning. SIAM J. Optim. **5**(4), 858–870 (1995)
15. Sergeyev, Ya.D.: Global one-dimensional optimization using smooth auxiliary functions. Math. Program. **81**, 127–146 (1998)
16. Sergeyev, Ya.D., Daponte, P., Grimaldi, D., Molinaro, A.: Two methods for solving optimization problems arising in electronic measurements and electrical engineering. SIAM J. Optim. **10**(1), 1–21 (1999)
17. Sergeyev, Ya.D., Kvasov, D.E.: Diagonal Global Optimization Methods. FizMatLit, Moscow (2008)
18. Strongin, R.G.: Numerical Methods in Multiextremal Problems. Nauka, Moscow (1978)
19. Strongin, R.G.: The information approach to multiextremal optimization problems. Stoch. Stoch. Rep. **27**, 65–82 (1989)
20. Strongin, R.G.: Algorithms for multi-extremal mathematical programming problems employing the set of joint space-filling curves. J. Glob. Optim. **2**, 357–378 (1992)
21. Strongin, R.G., Sergeyev, Ya.D.: Global Optimization with Non-Convex Constraints: Sequential and Parallel Algorithms. Kluwer, Dordrecht (2000)
22. Törn, A.A., Žilinskas, A.: Global Optimization. Lecture Notes in Computer Science, 350. Springer, Berlin (1989)
23. Zhigljavsky, A.A., Žilinskas, A.: Stochastic Global Optimization. Springer, New York (2008)