A partition-based global optimization algorithm

Giampaolo Liuzzi · Stefano Lucidi · Veronica Piccialli

Abstract   This paper is devoted to the study of partition-based deterministic algorithms for global optimization of Lipschitz-continuous functions without requiring knowledge of the Lipschitz constant. First we introduce a general scheme of a partition-based algorithm. Then, we focus on the selection strategy in such a way to exploit the information on the objective function. We propose two strategies. The first one is based on the knowledge of the global optimum value of the objective function. In this case the selection strategy is able to guarantee convergence of every infinite sequence of trial points to global minimum points. The second one does not require any a priori knowledge on the objective function and tries to exploit information on the objective function gathered during progress of the algorithm. In this case, from a theoretical point of view, we can guarantee the so-called every-where dense convergence of the algorithm.

Keywords   Global optimization · Partition-based algorithm · DIRECT-type algorithm

G. Liuzzi
Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti", viale Manzoni 30, 00185 Rome, Italy
e-mail: Giampaolo.Liuzzi@iasi.cnr.it

S. Lucidi (✉)
Dipartimento di Informatica e Sistemistica "A. Ruberti", Sapienza Università di Roma, via Ariosto 25, 00185 Rome, Italy
e-mail: Stefano.Lucidi@dis.uniroma1.it

V. Piccialli
Dipartimento di Ingegneria dell'Impresa, Università degli Studi di Tor Vergata, viale del Politecnico 1, 00100 Rome, Italy
e-mail: piccialli@disp.uniroma2.it

1 Introduction

We consider the global optimization problem

$$\text{glob} \min_{x \in \mathcal{D}} f(x), \tag{1}$$

where $\mathcal{D}$ is a hyperrectangle in $\Re^n$, namely, given $l, u \in \Re^n$, $\mathcal{D} = \{x \in \Re^n : l \le x \le u\}$, and we assume the function f to be Lipschitz continuous over the feasible domain, that is, a constant $0 < L < \infty$ exists such that, for every $x, y \in \mathcal{D}$,

$$|f(x) - f(y)| \le L\|x - y\|.$$

Many deterministic approaches have been proposed in the literature for solving Problem (1), see in particular [8,9,16,20,22]. A particularly interesting class of methods is the one that searches for the global minimum points by using sequences of partitions of the feasible domain. For every iteration k, they consider a collection of sets $\{\mathcal{D}^i, i \in I_k\}$ which constitutes a partition of the feasible domain. Namely, $\{\mathcal{D}^i, i \in I_k\}$ is such that:

$$\mathcal{D} = \bigcup_{i \in I_k} \mathcal{D}^i, \quad \text{Int}(\mathcal{D}^i) \bigcap \text{Int}(\mathcal{D}^j) = \emptyset, \quad \forall\, i, j \in I_k, \ i \ne j. \tag{2}$$

The sequence of partitions of the feasible domain is generated by choosing, at every iteration, some sets to be further partitioned. These sets can be chosen according to different criteria. One possibility consists in selecting the sets on the basis of some a priori knowledge on the objective function (e.g. knowledge of the Lipschitz constant). We refer to [8,9,16,20,22] for some recent approaches. On the other hand, the selection strategy can be defined without requiring any a priori knowledge on the objective function. In particular, the DIRECT algorithm [10,11] is an example of a partition method that tries to compensate the lack of knowledge of the Lipschitz constant by choosing it from a set of values varying from zero to infinity. More recently, in [19] the use of a set of Lipschitz constants is proposed with the aim of accelerating the convergence of a partition-based algorithm. In [1,5,12,18,20], further assumptions on f and ∇f are exploited in order to define new methods for global optimization problems.

In the paper we propose two partitioning strategies. The first one is based on the knowledge of the global optimum value of the objective function. In this case the selection strategy is able to guarantee convergence of every infinite sequence of trial points to global minimum points. The second one does not require any a priori knowledge on the objective function and tries to exploit information on the objective function gathered during progress of the algorithm. In this case, from a theoretical point of view, we can guarantee the so-called every-where dense convergence of the algorithm.

The paper is organized as follows. In Sect. 2 we describe a general scheme of partition-based algorithms and give some theoretical properties. In Sect. 3 we introduce two new strategies for selecting the hyperintervals for further partitioning. Finally, in Sect. 4 we present some illustrative numerical results.

2 Partitioning-based algorithms

We define the set of global minimum points

$$X^* = \left\{x^* \in \mathcal{D} : f(x^*) \le f(x), \text{for all } x \in \mathcal{D}\right\}.$$

Given a hyperinterval $\mathcal{D}^i$, we denote by $d^i = \|u^i - l^i\|$ its diagonal and by $\mathcal{P}(\mathcal{D}^i) = x^i$ the representative point of $\mathcal{D}^i$, namely the point having the best objective function value among those associated to $\mathcal{D}^i$.

A general Partitioning-based algorithm model can be described by the following scheme.

Partitioning-based algorithm (PBA)

Step 0: Set $\mathcal{D}^0 = \mathcal{D}$, $l^0 = l$, $u^0 = u$, $I_0 = \{0\}$ and $k = 0$.

Step 1: Given the partition $\{\mathcal{D}^i : i \in I_k\}$ of $\mathcal{D}$ with

$$\mathcal{D}^i = \{x \in R^n : l^i \leq x \leq u^i\}, \quad \text{for all } i \in I_k,$$

choose a particular subset $I_k^* \subseteq I_k$; set $\bar{I}^0 = I_k$, $\hat{I}^0 = I_k^*$ and $\ell = 0$.

Step 2: Choose an index $h \in \hat{I}^\ell$ and partition the set $\mathcal{D}^h$ into $m \geq 2$ subintervals

$$\mathcal{D}^{h_1}, \mathcal{D}^{h_2}, \dots, \mathcal{D}^{h_m}.$$

Step 3: Set

$$\bar{I}^{\ell+1} = \bar{I}^\ell \bigcup_{j=1,\dots,m} \{h_j\} \setminus \{h\},$$

$$\hat{I}^{\ell+1} = \hat{I}^\ell \setminus \{h\},$$

if $\hat{I}^{\ell+1} \neq \emptyset$ set $\ell = \ell + 1$ and go to Step 2.

Step 4: Define the new partition $\{\mathcal{D}^i : i \in I_{k+1}\}$ with

$$I_{k+1} = \bar{I}^{\ell+1},$$

set $k = k + 1$ and go to Step 1.

At each iteration PBA produces a new partition of the set $\mathcal{D}$. The choice of the set I_k^* of the hyperintervals to be partitioned, can be driven by information on the objective function. The asymptotic properties of the algorithm can be described by the asymptotic behavior of the sets that it generates. PBA produces an infinite number of sequences of subsets $\{\mathcal{D}^{i_k}\}$. Each of these sequences $\{\mathcal{D}^{i_k}\}$ can be characterized by associating to every subset $\mathcal{D}^{i_k}$, with $i_k \in I_k$, a predecessor $\mathcal{D}^{i_{k-1}}$, with $i_{k-1} \in I_{k-1}$, in the following way:

- if the set $\mathcal{D}^{i_k}$ has been generated at the k-th iteration, then $\mathcal{D}^{i_{k-1}}$ is the set which has been partitioned at the k-th iteration and which has generated the subset $\mathcal{D}^{i_k}$;
- if the set $\mathcal{D}^{i_k}$ has not been generated at the k-th iteration than $\mathcal{D}^{i_{k-1}} = \mathcal{D}^{i_k}$.

By definition, the sequences $\{\mathcal{D}^{i_k}\}$ are *nested* sequences of subsets, namely sequences such that, for all k,

$$\mathcal{D}^{i_k} \subseteq \mathcal{D}^{i_{k-1}}.$$

Among these sequences an important role in the theoretical analysis of the previous algorithm is played by the *strictly nested* sequences which have the property that for infinitely many times it results

$$\mathcal{D}^{i_k} \subset \mathcal{D}^{i_{k-1}}. \tag{3}$$

In Algorithm PBA the choice of the set I_k^* of the indices of the subsets to be partitioned and the choice of the partitioning technique are not specified. In this section we identify some general assumptions that the sets I_k^* and the partitioning techniques must satisfy in order to guarantee some theoretical properties of the algorithm model.

First we consider the requirements on the partition techniques and then the ones concerning the sets I_k^*.

2.1 Choice of the partition technique

In order to guarantee some theoretical properties of Algorithm PBA, the partition technique must be sufficiently regular and this can be formally stated by the following property.

Property 1 *There exist two scalars ϵ_1 and ϵ_2, with $0 < \epsilon_1 < \epsilon_2 < 1$, such that, for all k, every subset $\mathcal{D}^h$, $h \in I_k^*$, selected at Step 2 of the algorithm is partitioned into m subintervals $\mathcal{D}^{hj}$, $j = 1, \ldots, m$ such that:*

$$\epsilon_1 \left\| u^h - l^h \right\| \leq \left\| u^{hj} - l^{hj} \right\| \leq \epsilon_2 \left\| u^h - l^h \right\|, \quad j = 1, \ldots, m. \tag{4}$$

Under Property 1 it is possible to characterize the sequences $\{\mathcal{D}^{i_k}\}$ strictly nested. In the following proposition we recall some important properties of strictly nested sequences.

Proposition 1 *If Property 1 is satisfied, a sequence of sets $\{\mathcal{D}^{i_k}\}$ produced by Algorithm PBA is strictly nested if and only if one of the following points holds:*

(i) $\lim_{k\to\infty} \|u^{i_k} - l^{i_k}\| = 0$;
 there exists a point $\bar{x} \in \mathcal{D}$ such that
(ii) $\bigcap_{k=0}^{\infty} \mathcal{D}^{i_k} = \{\bar{x}\}$ *which is equivalent to*

$$\lim_{k\to\infty} u^{i_k} = \bar{x}, \qquad \lim_{k\to\infty} l^{i_k} = \bar{x}; \tag{5}$$

(iii) *for every $\epsilon > 0$, an index $\bar{k}$ exists such that for all $k \geq \bar{k}$ we have:*

$$\mathcal{D}^{i_k} \subset B(\bar{x}; \epsilon). \tag{6}$$

Proof Point (i) Let $\{\mathcal{D}^{i_k}\}$ be a sequence produced by Algorithm PBA. By definition of the sequence $\{\mathcal{D}^{i_k}\}$ and by the instructions of the algorithm, every time that happens the following strict inclusion $\mathcal{D}^{i_k} \subset \mathcal{D}^{i_{k-1}}$ we have that, at the $k-1$-th iteration, the subset $\mathcal{D}^{i_k}$ has been generated by the partition of the set $\mathcal{D}^{i_{k-1}}$. Then Property 1 implies that

$$\epsilon_1 \left\| u^{i_{k-1}} - l^{i_{k-1}} \right\| \leq \left\| u^{i_k} - l^{i_k} \right\| \leq \epsilon_2 \left\| u^{i_{k-1}} - l^{i_{k-1}} \right\|, \tag{7}$$

Recalling again the definition of the sequence of sets $\{\mathcal{D}^{i_k}\}$ we have that, for $j = 1, \ldots, k, \mathcal{D}^{i_j} \subset \mathcal{D}^{i_{j-1}}$ or $\mathcal{D}^{i_j} = \mathcal{D}^{i_{j-1}}$. By applying repeatedly (7) we obtain

$$(\epsilon_1)^{p_k} \left\| u^0 - l^0 \right\| \leq \left\| u^{i_k} - l^{i_k} \right\| \leq (\epsilon_2)^{p_k} \left\| u^0 - l^0 \right\|, \tag{8}$$

where p_k indicates the number of iterations where (7) is verified at Step 2.

Now, if the sequence $\{\mathcal{D}^{i_k}\}$ is strictly nested we have, by definition, that $\lim_{k\to\infty} p_k = \infty$, which, by using (8) and $\epsilon_2 \in (0, 1)$, implies that

$$\lim_{k\to\infty} \left\| u^{i_k} - l^{i_k} \right\| \leq \lim_{k\to\infty} (\epsilon_2)^{p_k} \left\| u^0 - l^0 \right\| = 0.$$

On the other hand, if the sequence $\{\mathcal{D}^{i_k}\}$ satisfies point (i) of the proposition, by using (8) we obtain that

$$\lim_{k\to\infty} (\epsilon_1)^{p_k} \left\| u^0 - l^0 \right\| \leq \lim_{k\to\infty} \left\| u^{i_k} - l^{i_k} \right\| = 0,$$

which, taking into account that $\epsilon_1 \in (0, 1)$, yields $\lim_{k\to\infty} p_k = \infty$. This limit implies that the sequence $\{\mathcal{D}^{i_k}\}$ produced by the algorithm is strictly nested.

Point (ii) First we prove that if $\{\mathcal{D}^{i_k}\}$ is strictly nested and Property 1 is satisfied, then the limits (5) hold.

Let us consider the following sequences of scalars $\{(u^{i_k})_j\}$ e $\{(l^{i_k})_j\}$, with $j = 1, \ldots, n$. By the instructions of the algorithm we have that, for all k:

$$(l)_j \leq \left(u^{i_k}\right)_j, \quad \left(u^{i_{k+1}}\right)_j \leq \left(u^{i_k}\right)_j, \quad j = 1, \ldots, n, \tag{9}$$

$$\left(l^{i_k}\right)_j \leq (u)_j, \quad \left(l^{i_k}\right)_j \leq \left(l^{i_{k+1}}\right)_j, \quad j = 1, \ldots, n. \tag{10}$$

For every $j = 1, \ldots, n$, by (9) we get that the sequence $\{(u^{i_k})_j\}$ is non increasing and bounded from below, by (10) we obtain that the sequence $\{(l^{i_k})_j\}$ is non decreasing and bounded from above. Therefore there exist the following limits:

$$\lim_{k\to\infty} \left(u^{i_k}\right)_j = (\bar{u})_j, \quad \lim_{k\to\infty} \left(l^{i_k}\right)_j = (\bar{l})_j, \quad j = 1, \ldots, n,$$

which imply that two vectors $\bar{u}, \bar{l} \in R^n$ exist such that

$$\lim_{k\to\infty} u^{i_k} = \bar{u}, \quad \lim_{k\to\infty} l^{i_k} = \bar{l}. \tag{11}$$

Then, by recalling that

$$\left\|\bar{u} - \bar{l}\right\| \leq \left\|\bar{u} - u^{i_k}\right\| + \left\|u^{i_k} - l^{i_k}\right\| + \left\|l^{i_k} - \bar{l}\right\|, \tag{12}$$

by taking the limit for k tending to infinity, by using (11) and point (i) of the proposition, we obtain $\bar{u} = \bar{l} = \bar{x}$.

Now, if the limits (5) hold, we have:

$$\lim_{k\to\infty} \left\|u^{i_k} - l^{i_k}\right\| \leq \lim_{k\to\infty} \left\|u^{i_k} - \bar{x}\right\| + \lim_{k\to\infty} \left\|l^{i_k} - \bar{x}\right\| = 0$$

that, by exploiting point (i) of the proposition, implies that the sequence $\{\mathcal{D}^{i_k}\}$ is strictly nested.

Point (iii) The definition of norm implies that every $x \in \mathcal{D}^{i_k}$ satisfies

$$\left\|x - l^{i_k}\right\| \leq \left\|u^{i_k} - l^{i_k}\right\|. \tag{13}$$

By using (5) and point (i) of the proposition, for every $\epsilon > 0$, an index $\bar{k}$ exists such that for all $k \geq \bar{k}$ we have:

$$\left\|l^{i_k} - \bar{x}\right\| < \frac{\epsilon}{2}, \quad \left\|u^{i_k} - l^{i_k}\right\| < \frac{\epsilon}{2}. \tag{14}$$

Then (13) and (14) imply that, for all $k \geq \bar{k}$ and for every $x \in \mathcal{D}^{i_k}$, we have:

$$\|x - \bar{x}\| \leq \left\|x - l^{i_k}\right\| + \left\|l^{i_k} - \bar{x}\right\| \leq \left\|u^{i_k} - l^{i_k}\right\| + \left\|l^{i_k} - \bar{x}\right\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

which proves (6).

On the contrary, if for every $\epsilon > 0$ an index $\bar{k}$ exists such that for all $k \geq \bar{k}$ the inclusion (6) holds then it follows that $u^{i_k} \in B(\bar{x}; \epsilon)$ and $l^{i_k} \in B(\bar{x}; \epsilon)$, which show that the limits (5) hold and, hence, point (ii) of the proposition implies that the sequence $\{\mathcal{D}^{i_k}\}$ is strictly nested.   □

The next corollary shows that, under Property 1, Algorithm PBA produces at least a strictly nested sequence of sets.

Corollary 1 *If Property 1 is satisfied, Algorithm PBA produces at least a strictly nested sequence of sets $\left\{\mathcal{D}^{i_k}\right\}$.*

Proof Let us assume by contradiction that the algorithm produces no strictly nested sequence. By Property 1 and Proposition 1, a constant $\Delta > 0$ exists such that, for all k and all $i \in I_k$, it results:

$$\left\| u^i - l^i \right\| \geq \Delta. \tag{15}$$

By recalling compactness of $\mathcal{D}$ and by using (15) we obtain that a constant $\bar{N}$ exists such that

$$|I_k| \leq \bar{N}, \quad \text{for all } k. \tag{16}$$

The instructions of the algorithm and the choice that $m \geq 2$ yield that, at every iteration, the number $|I_k|$ of hyperintervals that constitute the current partition of the initial domain grows at least of two elements, that is, $|I_{k+1}| \geq |I_k| + 2$. Hence, we get $\lim_{k\to\infty} |I_k| = +\infty$, which contradicts (16) thus completing the proof.   □

Under Property 1, the next corollary states that, given any infinite subset of iterations of Algorithm PBA, the corresponding sequence of partitions of the feasible set contains a strictly nested sequence of hyperintervals.

Corollary 2 *If Property 1 holds, then for every infinite subset of iterations K, the sequence of partitions $\left\{\mathcal{D}^i : i \in I_k\right\}_K$ contains at least a sequence of hyperintervals $\left\{\mathcal{D}^{i_k}\right\}_K$ that is strictly nested.*

Proof By Corollary 1, we know that at least a strictly nested sequence of hyperintervals $\left\{\mathcal{D}^{i_k}\right\}$ exists. We proceed by contradiction. Let us suppose that an index set K exists such that no strictly nested sequence $\left\{\mathcal{D}^{i_k}\right\}_K$ exists. This means that the inclusion $\mathcal{D}^{i_h} \subset \mathcal{D}^{i_k}$, with $k \in K$ and h the smallest integer such that $h \geq k, h \in K$, holds a finite number of times. Hence, a constant $\Delta > 0$ exists such that, for every $i \in I_k, k \in K$, it results

$$\left\| u^i - l^i \right\| \geq \Delta.$$

Now the proof follows by analogous reasoning as in the proof of Corollary 1.   □

2.2 Choice of the set I_k^*

Unlike from the case of partition techniques, the choices of sets I_k^* can be dictated by completely different strategies. In this subsection we examine some of them. All the theoretical properties can be derived by following the same reasonings as in [17]. Anyway, alternative proofs are reported in [13].

The first choice requires the following notations.

$$d_k^{\max} = \max_{i \in I_k} \left\| u^i - l^i \right\|, \quad I_k^{\max} = \left\{ i \in I_k : \left\| u^i - l^i \right\| = d_k^{\max} \right\}.$$

Then we can introduce the following choice of I_k^*.

Choice 1 *The set I_k^* satisfies* $\quad I_k^{\max} \cap I_k^* \neq \emptyset.$

The next proposition describes the asymptotic properties of Algorithm PBA when the partition technique satisfies Property 1 and, infinitely many times, set I_k^* is chosen according to Choice 1.

Proposition 2 *If Property 1 holds and, for infinitely many iteration indices k, Choice 1 is used, then*

(i) *all the sequences of sets $\{\mathcal{D}^{i_k}\}$ produced by Algorithm PBA are strictly nested;*

(ii) *for every $\tilde{x} \in \mathcal{D}$, Algorithm PBA produces a strictly nested sequence of sets $\{\mathcal{D}^{i_k}\}$ such that*

$$\bigcap_{k=0}^{\infty} \mathcal{D}^{i_k} = \{\tilde{x}\}.$$

In order to introduce a different choice of I_k^*, we assume that for each hyperinterval of the partition $\{\mathcal{D}^i, i \in I_k\}$ a scalar R_k^i is computed which gives an estimate of the minimum value of the objective function on the hyperinterval.

Let

$$i_k^{\min} = \arg\min_{i \in I_k} R_k^i, \quad R_k^{i_k^{\min}} = \min_{i \in I_k} R_k^i. \tag{17}$$

Then we state the following choice of I_k^*.

Choice 2 *Let $I_k^* = I_k^S$ where*

$$I_k^S = \left\{ i \in I_k : R_k^i = R_k^{i_k^{\min}} \right\}. \tag{18}$$

The next propositions generalize the results described in [15] and characterize the asymptotic properties of Algorithm PBE where the set I_k^* is chosen according to Choice 2.

In particular, the following proposition guarantees the existence of a strictly nested sequence of hyperintervals converging to a global minimum, provided that Choice 2 is used for infinitely many iterations and the scalars R_k^i satisfy two reasonable assumptions.

Proposition 3 *Let $\{\mathcal{D}^{i_k}\}$ be the sequences of subsets produced by Algorithm PBA. Assume that Property 1 holds, that infinitely many times Choice 2 is used, and that the scalars R_k^i satisfy the following two assumptions*

(i) *for any strictly nested sequence of subsets $\{\mathcal{D}^{i_k}\}$, such that $\bigcap_{k=0}^{\infty} \mathcal{D}^{i_k} = \{\bar{x}\}$, it holds*

$$\lim_{k \to \infty} R_k^{i_k} = f(\bar{x});$$

(ii) *there exist a point $x^* \in X^*$ and an index $\bar{k}$ such that, for all $k \geq \bar{k}$, there exists a subset $\mathcal{D}^{j_k}$, where $j_k \in I_k$, such that $x^* \in \mathcal{D}^{j_k}$ and*

$$R_k^{j_k} \leq f(x^*).$$

Then, a strictly nested sequence of subsets $\{\mathcal{D}^{i_k}\}$ exists such that

$$\bigcap_{k=0}^{\infty} \mathcal{D}^{i_k} \subseteq X^*.$$

In case Choice 2 is used at every iteration k, then the next proposition shows that any strictly nested sequence of hyperintervals converges to a global minimum under the same assumptions on the scalars R_k^i.

Proposition 4 *Let $\left\{\mathcal{D}^{i_k}\right\}$ be the sequences of subsets produced by Algorithm PBA. Assume that Property 1 holds, that, for every k, Choice 2 is used, and that the scalars R_k^i satisfy the following two assumptions*

(i) *for any strictly nested sequence of subsets $\left\{\mathcal{D}^{i_k}\right\}$, such that $\bigcap_{k=0}^{\infty} \mathcal{D}^{i_k} = \{\bar{x}\}$, it holds*

$$\lim_{k\to\infty} R_k^{i_k} = f(\bar{x});$$

(ii) *there exist a point $x^* \in X^*$ and an index $\bar{k}$ such that, for all $k \geq \bar{k}$, there exists a subset $\mathcal{D}^{j_k}$, where $j_k \in I_k$, such that $x^* \in \mathcal{D}^{j_k}$ and*

$$R_k^{j_k} \leq f(x^*).$$

Then, for any strictly nested sequence of subsets $\left\{\mathcal{D}^{i_k}\right\}$, it holds

$$\bigcap_{k=0}^{\infty} \mathcal{D}^{i_k} \subseteq X^*. \tag{19}$$

Finally, by requiring stronger assumptions on the scalars R_k^i, the following proposition shows that for every global minimum point, the algorithm produces a strictly nested sequence of hyperintervals converging to it.

Proposition 5 *Let $\left\{\mathcal{D}^{i_k}\right\}$ be the sequences of subsets produced by the algorithm. Assume that, for every k, Choice 2 is used and that the scalars R_k^i satisfy the following two hypothesis*

(i) *for any strictly nested sequence of subsets $\left\{\mathcal{D}^{i_k}\right\}$, it holds that*

$$\lim_{k\to\infty} R_k^{i_k} = f(\bar{x})$$

where $\bigcap_{k=0}^{\infty} \mathcal{D}^{i_k} = \{\bar{x}\}$.

(ii) *there exists an index $\bar{k}$ such that, for all $k \geq \bar{k}$ and for every $x^* \in X^*$ there exists a subset $\mathcal{D}^{j_k}$, where $j_k \in I_k$, such that $x^* \in \mathcal{D}^{j_k}$ and*

$$R_k^{j_k} < f(x^*).$$

Then, for every $x^* \in X^*$ there exists a strictly nested sequence of subsets $\left\{\mathcal{D}^{i_k}\right\}$, such that

$$\bigcap_{k=0}^{\infty} \mathcal{D}^{i_k} = \{x^*\}. \tag{20}$$

3 An alternating selection strategy

3.1 A selection technique using the optimal function value

In this section we assume that the optimal function value f^* is known and that $f(x)$ is continuously differentiable. The aim of the algorithm to be defined is therefore to determine a point $\bar{x}$ with $f(\bar{x})$ as close as possible to f^*.

Under this assumption we can consider the following new objective function

$$\theta(x) = (f(x) - f^*)^p, \tag{21}$$

where $p \geq 1$ and $p > 1$ if a point $x^* \in X^*$ exists such that $x^* \in \partial \mathcal{D}$.

Proposition 6 *Let f be a continuously differentiable function. For every constant $0 < \hat{L} < \infty$ and for all $x^* \in X^*$ a positive ϵ exists such that the constant $\hat{L}$ is a strictly overestimate of the local Lipschitz constant of the function $\theta(x)$ over the neighborhood $B(x^*; \epsilon)$, namely for all $x \in B(x^*; \epsilon)$*

$$|\theta(x^*) - \theta(x)| < \hat{L} \, \|x^* - x\| .$$

Proof By the Mean Theorem we have

$$\theta(x) = \theta(x^*) + \nabla\theta(\hat{x})^T(x - x^*) = \theta(x^*) + p(f(\hat{x}) - f^*)^{p-1}\nabla f(\hat{x})^T(x - x^*),$$

where $\hat{x} = x^* + \eta(x - x^*)$, with $\eta \in [0, 1]$.

From the preceding equality we obtain

$$|\theta(x) - \theta(x^*)| \leq p(f(\hat{x}) - f^*)^{p-1} \, \|\nabla f(\hat{x})\| \, \|x - x^*\| .$$

Therefore, by the continuity of f and ∇f and the compactness of the feasible set, for every $x^* \in X^*$ and for every $\hat{L} > 0$, there exists an $\epsilon(x^*)$ such that

$$\max_{x \in B(x^*; \epsilon(x^*))} p(f(x) - f^*)^{p-1} \, \|\nabla f(x)\| < \hat{L}.$$

Then we have that for all $x \in B(x^*; \epsilon(x^*))$, $|\theta(x^*) - \theta(x)| < \hat{L} \, \|x^* - x\|$. □

Under the assumption that the objective function value of the global minimum point is known "a priori", Proposition 6 states that every $\hat{L} > 0$ can be used as an estimate of the Lipschitz constant of $\theta(x)$ within a neighborhood of a global minimum point.

For every $\tilde{L} > 0$, and for every $i \in I_k$ let

$$R_k^i = \theta(x_k^i) - \tilde{L}d^i .$$

Choice 3 *If $R_k^{i_k^{\min}} < 0$ (where $R_k^{i_k^{\min}}$ is defined in (17)), then*

$$I_k^* = I_k^S,$$

otherwise I_k^* must be such that

$$I_k^* \cap I_k^{\max} \neq \emptyset.$$

The rationale behind Choice 3 is that of combining the selection Choices 1 and 2. Namely, when the test on $R_k^{i_k^{\min}}$ is satisfied we resort to Choice 2, otherwise we keep on partitioning according to Choice 1.

The Proposition that follows proves convergence of Algorithm PBA towards global minimum points under Choice 3 regardless of the value of $\tilde{L}$. This implies that no rule to estimate the Lipschitz constant of θ is needed.

Proposition 7 *Let Property 1 hold and assume that Choice 3 is used. Then, all the strictly nested sequences $\{D^{i_k}\}$ generated by Algorithm PBA satisfy*

$$\bigcap_{k=1}^{\infty} \{D^{i_k}\} = x^*,$$

where $x^* \in X^*$.

Proof First, we show that there exists $\bar{k}$ such that, for all $k \geq \bar{k}$, $R_k^{i_k^{\min}} < 0$. By contradiction, assume that there exists an infinite subset of indices K such that for all $k \in K$, $R_k^{i_k^{\min}} > 0$, namely

$$\theta\left(x_k^{i_k^{\min}}\right) \geq \tilde{L} d^{i_k^{\min}}. \tag{22}$$

Therefore, by Choice 3, we have that $I_k^* \cap I_k^{\max} \neq \emptyset$ infinitely many times. Hence, Proposition 2 implies that $\lim_{k \in K} d_k^{\max} = 0$. Let $x^* \in X^*$. By Proposition 6, there exists ϵ such that for all $x \in B(x^*, \epsilon) |\theta(x^*) - \theta(x)| < \tilde{L} \|x^* - x\|$. By point (ii) of Proposition 2, an index $h \in I_k$ exists such that $x_k^h \in B(x^*, \epsilon)$ and $D^h \subseteq B(x^*, \epsilon)$. Therefore, keeping into account that $\theta(x^*) = 0$, we get

$$\theta\left(x_k^h\right) < \tilde{L} d^h.$$

By the definition of $i_k^{\min}$, we have

$$\theta\left(x_k^{i_k^{\min}}\right) - \tilde{L} d^{i_k^{\min}} \leq \theta(x_k^h) - \tilde{L} d^h < 0,$$

that contradicts (22). Therefore, for k sufficiently large, we have $R_k^{i_k^{\min}} < 0$. Consider any strictly nested subsequence $\{D^{i_k}\}$ such that $\bigcap_{k=0}^{\infty} D^{i_k} = \{\bar{x}\}$. Let K be the index set of iterations such that $D^{i_{k+1}} \subset D^{i_k}$. Then we have, since $I_k^* = I_k^S$

$$\theta\left(x_k^{i_k}\right) < \tilde{L} d^{i_k}. \tag{23}$$

Moreover, by Proposition 1, we know that $\lim_{k \to \infty} d^{i_k} = 0$, and $\lim_{k \to \infty} x_k^{i_k} = \bar{x}$, which combined with (23) imply $\lim_{k \to \infty} f\left(x_k^{i_k}\right) = f(\bar{x}) = f^*$, that is, $\bar{x} \in X^*$. $\qquad \square$

3.2 A selection technique based on an adaptive estimate of the global minimum value

In this subsection we consider the case where the global minimum value is not known *a priori*. The hyperintervals generated at every iteration of a partition-based algorithm have associated one or more points which are its representatives.

Throughout this subsection we assume that every point $x^* \in X^*$ is such that $l < x^* < u$. Futhermore, let $\tilde{f}$ be an estimate of the global minimum value, i.e. $\tilde{f} = f(\tilde{x})$, for any $\tilde{x} \in \mathcal{D}$. Given an $\epsilon > 0$, we define

$$\theta_\epsilon(x) = f(x) - \left(\tilde{f} - \epsilon\right).$$

Partitioning-Based on Estimate Algorithm (PBE)

Data: $\tilde{L} > 0, \tau \in (0, 1), \epsilon_0 > 0$. Set $\mathcal{D}^0 = \mathcal{D}, x^0 = \mathcal{P}(\mathcal{D}^0), \tilde{f}_0 = f(x^0), l^0 = l, u^0 = u$, $I_0 = \{0\}$ and $k = 0$.

Step 1: Given the partition $\{\mathcal{D}^i : i \in I_k\}$ of $\mathcal{D}$ with

$$\mathcal{D}^i = \left\{x \in R^n : l^i \leq x \leq u^i\right\}, \quad \text{for all } i \in I_k,$$

compute $R_k^i = \theta_{\epsilon_k}\left(\mathcal{P}\left(\mathcal{D}^i\right)\right) - \tilde{L} d^i$, for all $i \in I_k$,

if $R_k^{i_k^{\min}} < 0$, then choose I_k^* s.t. $I_k^* \subseteq \left\{ i \in I_k : R_k^i < 0 \right\}$
else choose I_k^* s.t. $I_k^* \cap I_k^{\max} \neq \emptyset$.
Set $\bar{I}^0 = I_k$, $\hat{I}^0 = I_k^*$ and $\ell = 0$.

Step 2: Choose an index $h \in \hat{I}^\ell$ and partition the set $\mathcal{D}^h$ into $m \geq 2$ subintervals

$$\mathcal{D}^{h_1}, \mathcal{D}^{h_2}, \ldots, \mathcal{D}^{h_m}.$$

Step 3: Set

$$\bar{I}^{\ell+1} = \bar{I}^\ell \bigcup_{j=1,\ldots,m} \{h_j\} \setminus \{h\}, \quad \hat{I}^{\ell+1} = \hat{I}^\ell \setminus \{h\},$$

if $\hat{I}^{\ell+1} \neq \emptyset$ set $\ell = \ell + 1$ and go to Step 2.

Step 4: Define the new partition $\left\{ \mathcal{D}^i : i \in I_{k+1} \right\}$ with $I_{k+1} = \bar{I}^{\ell+1}$.
Set $\tilde{f}_{k+1} = \min_{i \in I_{k+1}} f\left(\mathcal{P}(\mathcal{D}^i)\right)$
If $\tau \left(d_{k+1}^{\max}\right)^2 < \epsilon_k$, then set $\epsilon_{k+1} = \tau \left(d_{k+1}^{\max}\right)^2$ else $\epsilon_{k+1} = \epsilon_k$.
Set $k = k + 1$ and go to Step 1.

Algorithm PBE shifts the difficulty of the problem from the estimate of the Lipschitz constant to the estimate of the objective function value of the global minimum. Indeed, the main difference between Algorithm PBE and the methods proposed in references [3,6,10,11,19], consists in PBE iteratively estimating the objective function global minimum value rather than the Lipschitz constant.

We note that definition of Algorithm PBE should be completed by specifying the following aspects.

–   The operator $\mathcal{P}$. Two examples of possible choices for $\mathcal{P}$ are those proposed in [10] and [15]. The former reference associates to each hyperinterval its centroid. Whereas, the latter reference associates to each hyperinterval the two extreme points on the diagonal.
–   The choice of set I_k^* of hyperintervals to be divided. In particular, when $R_k^{i_k^{\min}} \geq 0$, I_k^* can be chosen as in [10] which guarantees that $I_k^* \cap I_k^{\max} \neq \emptyset$.
–   The partitioning scheme that defines Step 2. This can be any scheme that is able to guarantee Property 1.

The theoretical properties of Algorithm PBE can be derived without specifying the above points. The exact definition of our implementation of Algorithm PBE is postponed to the numerical results section.

The first result shows that the test at Step 1 of Algorithm PBE does not preclude the everywhere dense convergence property.

Proposition 8 *If Property 1 holds, then*

(i)   *all the sequences of sets $\left\{\mathcal{D}^{i_k}\right\}$ produced by Algorithm PBE are strictly nested;*
(ii) *for every $\tilde{x} \in \mathcal{D}$, Algorithm PBE produces a strictly nested sequence of sets $\left\{\mathcal{D}^{i_k}\right\}$ such that*

$$\bigcap_{k=0}^{\infty} \mathcal{D}^{i_k} = \{\tilde{x}\}.$$

Proof In order to prove the result, we show that for infinitely many iteration indices k, Algorithm PBE chooses I_k^* such that $I_k^* \cap I_k^{\max} \neq \emptyset$. We proceed by contradiction. Assume that an index $\bar{k}$ exits such that for all $k \geq \bar{k}$, $R_k^{i_k^{\min}} < 0$ and $I_k^* \cap I_k^{\max} = \emptyset$. Namely,

$$f\left(\mathcal{P}\left(\mathcal{D}^{i_k^{\min}}\right)\right) - \tilde{f}_k < -\epsilon_k + \tilde{L}d^{i_k^{\min}}. \tag{24}$$

Since $d_k^{i_k^{\min}} \to 0$, we have that for $k \geq \bar{k}$ sufficiently large, $d_k^{i_k^{\min}} < d_k^{\max}$, so that $\epsilon_k = \bar{\epsilon} \leq \epsilon_{\bar{k}}$,

$$f\left(\mathcal{P}\left(\mathcal{D}^{i_k^{\min}}\right)\right) - \tilde{f}_k < -\bar{\epsilon} + \tilde{L}d^{i_k^{\min}} < -\frac{\bar{\epsilon}}{2}.$$

This implies that $\tilde{f}_{k+1} \leq \tilde{f}_k - \frac{\bar{\epsilon}}{2}$ infinitely many times for k sufficiently large from which we would obtain $\tilde{f}_k \to -\infty$. Hence, Algorithm PBE chooses I_k^* according to Choice 1 infinitely many times. Therefore, the assumptions of Proposition 2 are satisfied and this concludes the proof. □

The next proposition explains the role of Choice 3 in the algorithm. Namely, Choice 3 promotes selection and partitioning of those hyperintervals containing the global minima.

Proposition 9 *Let Property 1 hold.*

(i) *For every $\bar{x} \in \mathcal{D} \setminus X^*$, an iteration index $\bar{k}$ exists such that, for every $k \geq \bar{k}$, if $\mathcal{D}^{i_k}$ satisfies $\bar{x} \in \mathcal{D}^{i_k}$, then $R_k^{i_k} \geq 0$.*

(ii) *For every $x^* \in X^*$, an iteration index $\bar{k}$ exists such that, for every $k \geq \bar{k}$, if $\mathcal{D}^{i_k}$ satisfies*

$$x^* \in \mathcal{D}^{i_k} \quad and \quad d^{i_k} \geq \frac{2\tau}{\tilde{L}} \left(d_k^{\max}\right)^2,$$

then $R_k^{i_k} < 0$.

Proof Point (i) By Proposition 8 it follows that for any $x \in \mathcal{D}$ Algorithm PBE produces a strictly nested sequence such that $\bigcap_{k=0}^{\infty} \mathcal{D}^{i_k} = \{x\}$. Therefore the statement is well posed.

By point (i) of Proposition 1 and by Proposition 8 it follows that $\lim_{k \to \infty} d^{i_k} = 0$. Therefore by Step 4 of Algorithm PBE

$$\lim_{k \to \infty} \epsilon_k = 0. \tag{25}$$

By the updating rule defining $\tilde{f}_k$ and by (25), it follows that

$$\lim_{k \to \infty} \tilde{f}_k = f(x^*), \quad x \in X^*. \tag{26}$$

For every $\bar{x} \in \mathcal{D} \setminus X^*$, let $\{\mathcal{D}^{i_k}\}$ be the strictly nested sequence produced by Algorithm PBE such that

$$\bigcap_{k=0}^{\infty} \mathcal{D}^{i_k} = \{\bar{x}\}.$$

By the definition of $R_k^{i_k}$, it follows that

$$R_k^{i_k} = f\left(x_k^{i_k}\right) - \tilde{f}_k - \tilde{L}d^{i_k} + \epsilon_k = f\left(x_k^{i_k}\right) - f^* + f^* - \tilde{f}_k - \tilde{L}d^{i_k} + \epsilon_k,$$

Since $\lim_{k \to \infty} f\left(x_k^{i_k}\right) = f(\bar{x}) > f^*$, taking the limit we get by (25) and (26)

$$\lim_{k \to \infty} R_k^{i_k} > 0,$$

and hence the thesis follows.

Point (ii) Let us consider the hyperinterval $\mathcal{D}^{i_k}$ such that $x^* \in \mathcal{D}^{i_k}$. Since $x^* \in X^*$, it results

$$f\left(x_k^{i_k}\right) - \tilde{f}_k - \tilde{L}d^{i_k} + \epsilon_k \le f\left(x_k^{i_k}\right) - f\left(x^*\right) - \tilde{L}d^{i_k} + \epsilon_k.$$

By the updating rule of ϵ_k, we have that

$$f\left(x_k^{i_k}\right) - f\left(x^*\right) - \tilde{L}d^{i_k} + \epsilon_k \le f\left(x_k^{i_k}\right) - f\left(x^*\right) - \tilde{L}d^{i_k} + \tau\left(d_k^{\max}\right)^2$$

$$= f\left(x_k^{i_k}\right) - f\left(x^*\right) - \frac{\tilde{L}}{2}d^{i_k} + \left(\tau\left(d_k^{\max}\right)^2 - \frac{\tilde{L}}{2}d^{i_k}\right).$$

Recalling Proposition 6, we have that, for k sufficiently large, $\tilde{L}/2$ is an overestimate of the Lipschitz constant of $\theta(x) = f(x) - f^*$. Hence, for k sufficiently large, and by the stated assumptions, we get

$$f\left(x_k^{i_k}\right) - f(x^*) - \frac{\tilde{L}}{2}d^{i_k} + \left(\tau\left(d_k^{\max}\right)^2 - \frac{\tilde{L}}{2}d^{i_k}\right) < 0,$$

and the result follows. □

The above proposition stress the role played by the scalars R_k^i in the selection strategy. Roughly speaking, they can help to produce sequences of hyperintervals which more rapidly concentrate around the global minima. Indeed, for k sufficiently large, the value of R_k^i for hyperintervals not containing a global minimum is positive or equal to zero, whereas, those hyperintervals containing a global minimum and having a diagonal "sufficiently" large, have $R_k^i < 0$.

4 Preliminary numerical results and conclusions

In this section we preliminarily describe an implementation of Algorithm PBE and then we report and comment some numerical results on a class of well-known global optimization problems.

The algorithm that we propose is a modification of the DIRECT Algorithm [10] for global optimization. The main characteristics of DIRECT are: (a) set I_k^* is the set of *potentially optimal* (see [10]) hyperintervals which is in agreement with Choice 1; (b) the adopted partitioning procedure satisfies Property 1.

If $R_k^{i_{\min}} \ge 0$, then the algorithm exactly follows DIRECT. Otherwise, I_k^* is chosen to be the set of potentially optimal hyperintervals among those with $R_k^i < 0$. Moreover, according to Proposition 9, we know that the role of scalars R_k^i is significant for k sufficiently large. Hence, in the implementation of Algorithm PBE we select I_k^* using scalars R_k^i only when the current number of hyperintervals is greater than $(10n)^2$. As concerns the values of the constants defining Algorithm PBE, we chose $\tilde{L} = 0.5$, $\tau = 0.9$ and $\epsilon_0 = 10^{-3}\max\left\{1, \left|\tilde{f}_0\right|\right\}$.

In Table 1, we report a comparison between the original DIRECT algorithm and the proposed algorithm PBE on a set of well-known global optimization problems. Both the algorithms were stopped either when a point $\bar{x}$ is generated such that

$$\frac{f(\bar{x}) - f^*}{\max\{1, |f^*|\}} \le 10^{-4} \tag{27}$$

Table 1 Comparison between DIRECT and PBE

Problem	n	DIRECT		PBE	
		$f(\bar{x})$	n.int.	$f(\bar{x})$	n.int.
Schubert [7]	2	−186.7215373	2967	−186.7215373	3181
Schub. pen. 1 [14]	2	−186.7215352	2379	−186.7215352	2353
Shub. pen. 2 [14]	2	−186.7215331	1595	−186.7215331	1569
S-H. Camel B. [4]	2	−1.031529633	119	−1.031529633	119
Goldstein-Price [2]	2	3.000090378	191	3.000090378	191
Treccani mod. [22]	2	7.67E-05	111	7.67E-05	111
Quartic [14]	2	−0.352366398	133	−0.352366398	133
Shekel $m = 5$ [4]	4	−10.15234984	153	−10.15234984	153
Shekel $m = 7$ [4]	4	−10.40196762	145	−10.40196762	145
Shekel $m = 10$ [4]	4	−10.53539008	145	−10.53539008	145
Espon. mod. [1]	2	−0.999920033	89	−0.999920033	89
Espon. mod. [1]	4	−0.999919526	567	−0.999919526	567
Cos-mix mod. [1]	2	−0.199986472	111	−0.199986472	111
Cos-mix mod. [1]	4	−0.399972944	417	−0.399972944	417
Hartman [7]	3	−3.862452145	199	−3.862452145	199
Hartman [7]	6	−3.3220738	571	−3.3220738	571
Griewank mod. [7]	2	7.88E-07	41005	7.88E-07	41151
Rotated Griewank [21]	2	−179.9856895	129	−179.9856895	129
Ackley [7]	2	5.65E-05	561	5.65E-05	561
Ackley [7]	10	**9.4929E-02**	800000	**1.95E-02**	800000
Dixon Price [7]	2	6.26E-05	443	6.26E-05	423
Easom [7]	2	−0.999989985	7019	−0.999989985	6673
Michalewics [7]	2	−1.801272488	67	−1.801272488	67
5n loc-min [14]	2	1.23E-06	129	1.23E-06	129
5n loc-min [14]	5	6.41E-05	361	6.41E-05	361
5n loc-min [14]	10	4.70E-05	1773	4.70E-05	1773
10n loc-min [14]	2	1.97E-05	265	1.97E-05	265
10n loc-min [14]	5	2.72E-05	2765	2.72E-05	2765
10n loc-min [14]	10	2.97E-05	21601	2.97E-05	11203
15n loc-min [14]	2	9.30E-05	137	9.30E-05	137
15n loc-min [14]	5	9.61E-05	903	9.61E-05	903
15n loc-min [14]	10	9.44E-06	16959	9.44E-06	11579
Pinter [16]	2	1.66E-05	105	1.66E-05	105
Pinter [16]	5	3.16E-05	1613	3.16E-05	1613
Pinter [16]	10	7.48E-05	8265	7.48E-05	8265
Rastrigin [7]	2	5.46E-05	893	5.46E-05	831
Rastrigin [7]	10	**9.9497110**	800000	2.64E-08	22201

or when the number of generated hyperintervals exceeds the prescribed limit of 800000, in which case the best funtion value obtained by the algorithm is reported in boldface.

We point out that for those functions (`Griewank`, `Cos-Mix`) having the global minimum point in the centroid of the feasible set (so that DIRECT would find them with one

objective function evaluation), we made the change of variables $y_i = x_i + 0.5\sqrt{2}$. For the same reason, for the `Treccani` and `Esponential` test functions, we applied a slightly different change of variables, that is $y_i = x_i + 0.25\sqrt{2}$, since the previous one would have moved the global minimum point out of the suggested bounds.

The table shows that the test introduced in Algorithm PBE on the selection strategy can help improving convergence of the algorithm. More in particular, PBE is more efficient on 9 problems out of 37 (meaning that either the number of function evaluations or the estimate of the global minimum value has been improved). The original DIRECT is better only on two problems. Furthermore, we stress that, for the `Rastrigin` test function with $n = 10$, PBE is able to find the global minimum point while DIRECT fails. All this considered, these results encourage to further investigate hyperintervals selection strategies based on scalars R_k^i. Reasonably, the numerical behavior of the method could benefit from an accurate choice and update of the estimate of the Lipschitz constant $\tilde{L}$. In particular, this could be achieved by associating to every hyperinterval a scalar L_k^i that can be evaluated according to the approach proposed in [19].

References

1. Breiman, L., Cutler, A.: Deterministic algorithm for global optimization. Math. Program. **58**, 179–199 (1993)
2. Dixon, L.C.W., Szegö, G.P.: Towards Global Optimization 2. North Holland (1975)
3. Finkel, D.E., Kelley, C.T.: Additive scaling and the DIRECT algorithm. J. Glob. Optim. **36**, 597–608 (2006)
4. Floudas, C.A., Pardalos, P.M., Adjiman, C.S., Esposito, W.R., Gümüs, Z., Harding, S.T., Klepeis, J.L., Meyer, C.A., Schweiger, C.A.: Handbook of Test Problems for Local and Global Optimization. Kluwer, Dordrecht (1999)
5. Gergel, V.P.: A global optimization algorithm for multivariate function with Lipschitzian first derivatives. J. Glob. Optim. **10**, 257–281 (1997)
6. Gablonsky, J.M., Kelley, C.T.: A locally-biased form of the DIRECT algorithm. J. Glob. Optim. **21**, 27–37 (2001)
7. Hedar, A.: http://www-optima.amp.i.kyoto-u.ac.jp/member/student/hedar/Hedar_files/TestGO.htm
8. Horst, R., Pardalos, P.M., Thoai, N.V.: Introduction to Global Optimization. Kluwer, Dordrecht (2000)
9. Horst, R., Tuy, H.: Global Optimization: Deterministic Approaches. Springer, Berlin (1990)
10. Jones, D.R., Perttunen, C.D., Stuckman, B.E.: Lipschitzian optimization without the Lipschitz constant. J. Optim. Theory Appl. **79**(1), 157–181 (1993)
11. Jones, D.R.: The DIRECT global optimization algorithm. In: Floudas, C., Pardalos, P. (eds.) Encyclopedia of Optimization, pp. 431–440. Kluwer, Dordrecht (2001)
12. Kvasov, D.E., Sergeyev, Y.D.: A univariate global search working with a set of Lipschitz constants for the first derivatives. Optim. Lett. **3**, 303–318 (2009)
13. Liuzzi, G., Lucidi, S., Piccialli, V.: A partition-based global optimization algorithm. Technical Report IASI (2009)
14. Lucidi, S., Piccioni, M.: Random Tunneling by Means of Acceptance-Rejection Sampling for Global Optimization. J. Optim. Theory Appl. **62**(2), 255–279 (1989)
15. Molinaro, A., Pizzuti, C., Sergeyev, Y.D.: Acceleration tools for diagonal information global optimization algorithms. Comput. Optim. Appl. **18**, 5–26 (2001)
16. Pintér, J.D.: Global Optimization in Action, Continuous and Lipschitz Optimization: Algorithms, Implementations and Applications. Nonconvex Optimization and Its Applications, Vol. 6. Kluwer, Dordrecht (1996)
17. Sergeyev, Y.D.: On convergence of divide the best global optimization algorithms. Optimization **44**, 303–325 (1998)

18. Sergeyev, Y.D.: Global one-dimensional optimization using smooth auxiliary functions. Math. Program. **81**, 127–146 (1998)
19. Sergeyev, Y.D., Kvasov, D.E.: Global search based on efficient diagonal partitions and a set of Lipschitz constants. SIAM J. Optim. **16**, 910–937 (2006)
20. Strongin, R.G., Sergeyev, Y.D.: Global Optimization with Non-convex Constraints. Kluwer, Dordrecht (2000)
21. Suganthan, P.N., Hansen, N., Liang, J.J., Deb, K., Chen, Y.P., Auger, A., Tiwari, S.: Problem definitions and evaluation criteria for the cec 2005 special session on real-parameter optimization. Technical Report, Nanyang Technological University, Singapore (2005)
22. Törn, A., Žilinskas, A.: Global Optimization. Springer, Berlin (1989)