DIRECT algorithm : A new definition of potentially optimal hyperrectangles

Lakhdar Chiter
Laboratoire de Mathématiques Fondamentales et Numérique, Departement de Mathématiques, Université de Sétif, 19000, Algérie
E-mail : chiterl@yahoo.fr

Abstract

We propose a new version of potentially optimal intervals for the DIRECT algorithm. A two-points based sampling method is presented. The method starts from a distingished point (the peak point) by forming an initial triangle. The idea is to sample the midpoint of a specific interval: the basis of the resulting triangle. This specific interval is obtained by translating the initial interval towards the lowest function value : $\min\{f(c_i),f(c_{i+1})\}$ and then overcoming the disadvantage if the global minimum lies at the boundaries. Two-dimensional version of our subdivision and sampling method is also discussed.

Keywords: Global optimization; DIRECT algorithm; Two-points based sampling method; Potentially optimal triangle

1. Introduction

By DIRECT: ”DIviding RECTangles”, we mean the algorithm developed by Jones et al. [11], and [12]. The algorithm begins by scaling the domain into the unit hypercube by adopting a center-sampling strategy. The objective function is evaluated at the midpoint of the domain, where a lower bound is constructed. In one-dimension, the domain is trisected and two new center points are sampled. At each iteration (dividing and sampling), DIRECT identifies intervals that contain the best minimal value of the objective function found up to now. This strategy of selecting and dividing gives DIRECT its performance and fast convergence compared to other deterministic methods. As a young method, DIRECT is being enhanced with new ideas. Some modifications have been investigated by many authors: Gablonsky et al. [6-9] studied the behaviour of DIRECT. The algorithm was used for lower dimensional problems with few local minima. Finkel and Kelly, [3-5] studied the convergence of the algorithm. Huyer and Neumaier [10] used some ideas behind DIRECT to overcome some disatvantages that the algorithm encounters The algorithm converges to the global optimal function value, if the objective function is continuous or at least continuous in the neighborhood of a global optimum. This could be assured since, as the number of iterations goes to infinity. In this paper we introduce a different way for sampling and dividing the search domain. Since the midpoint sampling is the most efficient method in global optimization, as in DIRECT, our subdivision scheme maintains this property. Each interval contains a distinguished point, the peak point, whose function value is known. This point is nothing but the center of a specific interval. Like DIRECT, we derive two possibilities to define a potentially optimal interval. The paper is organized as follows. Section 2 is a short introduction to the subdivision and sampling method both for one and two dimensional versions. Details of this method are discussed in sections 3 and 4, while in section 5, we give an algorithm for this. We then conclude in the 6^{th} section.

2. A new variante of the DIRECT algorithm

In this section, we give an overview of the ideas behind our subdivision scheme and leave the details to the next section. The one-dimensional DIRECT algorithm can be described by the following steps: the first step in the algorithm is the initialization, it consists by evaluating the objective function at the midpoint of an initial interval. This value is taken as f_{min}. This interval is then divided into three subintervals. The point sampled before will be a midpoint of the center subinterval. So, two new points are added at every step. In each iteration (divide-sample), new intervals are created by dividing old ones, and then the function is sampled at new centers of the new intervals. During an iteration, DIRECT will identifie potentially optimal intervals. DIRECT will sample at the centers of the newely created intervals. We describe now the one and two dimensional versions of our method of subdividing and sampling. For a more detailed description of the one dimensional version, the reader is refered to [2]. We evaluate the objective function at two consecutive points : one-third and two-third of the length of an interval [a, b]. These two points are given by : c_1 =(2a + b)/3, and c_2 =(a + 2b)/3. We divide this interval into two subintervals (bisection). The sampled points (1/3 and 2/3) will be respectively 2/3 and 1/3 of the left and right subintervals. We add a new point for each subinterval. So we need two new points at each iteration. This is exactly the contrary of what is done in DIRECT, see figure (1). The two dimensional version consists by considering four points which form the center of the considered domain. The subdivision can be done similarly. In three dimensions, the points considered form a centered cube. For the N-dimensional case we will have 2^N points. Figure (2) is an illustration of a two-dimensional case. We describe how this subdivision is done in N-dimensional bisection either for a hypercube or a hyperrectangle. Recall that DIRECT evaluates the function at the points $c \pm \delta e_i$, where c is the center of a hypercube and δ equals 1/3 of the side length of the cube and e_i is the ith coordinate vector. The algorithm divides the hypercube in the order given by : $w_i = min\{f(c+\delta e_i), f(c-\delta e_i)\}$, first, perpendicular to the direction with the lowest w_i , and then divides the remaining domain perpendicular to the direction of the second lowest w_i . Figure (2) is an illustration of the division of a hypercube, where w_1=min{5, 8}=5, and w_2=min{6, 2}=2, i.e., the division is done, first, perpendicular to the y-axis, and then perpendicular to the x-axis. An analogue subdivision to our method is presented in figure (7). Note that this subdivision can be done using the lower bound in [2]. Recall that for a hyperrectangle, DIRECT trisectes the longest coordinate directions. Details are discussed in the next section.

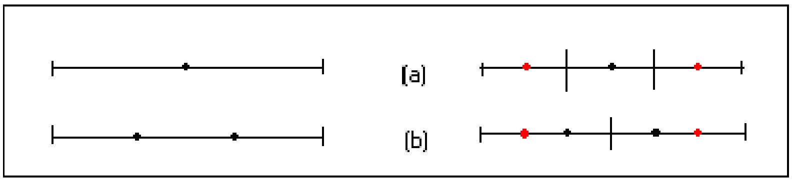
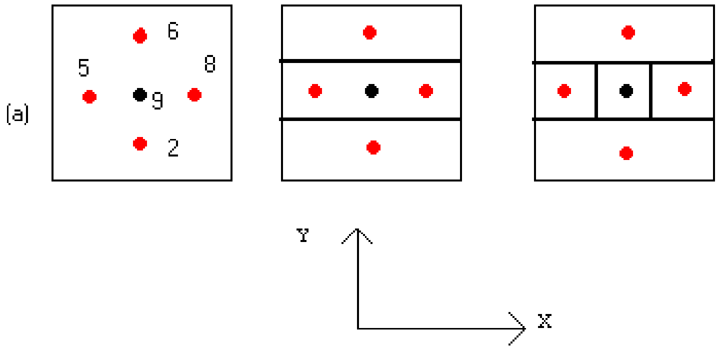


Figure 1. A comparison with DIRECT: (a) DIRECT ; (b) : the new sampling method. Points in (b) are c_1= (2a+b)/3; and c_2= (a+2b)/3 for an interval [a, b].



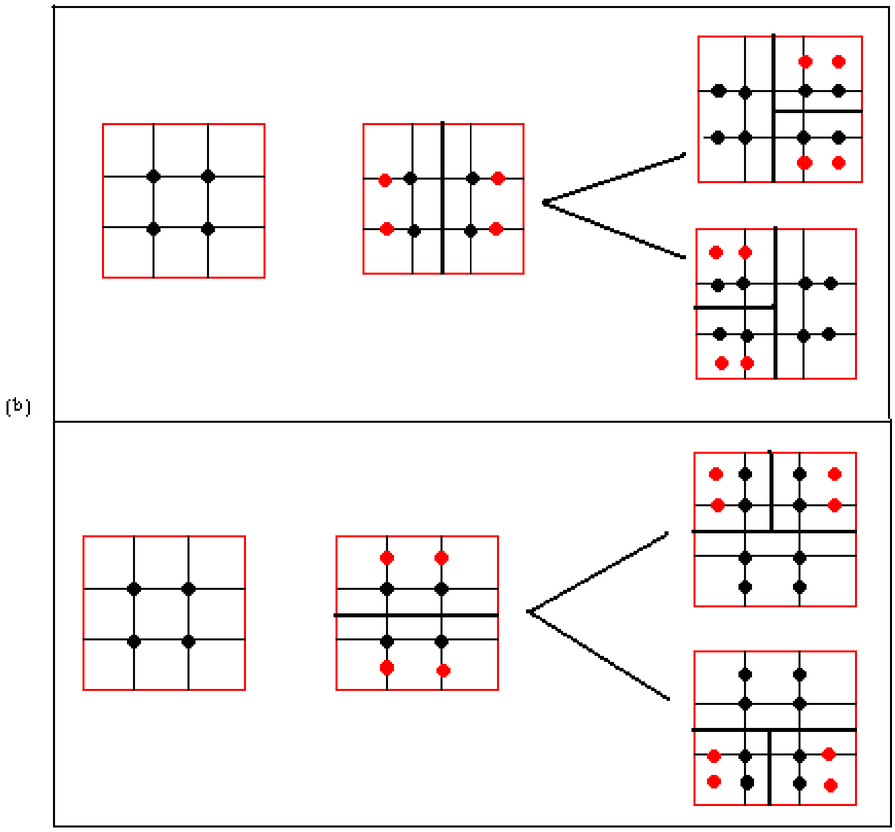


Figure 2. Two- dimensional version **:** (a) : DIRECT (points are taken from [7]); (b) : new sampling method. There is two possibilities of subdivision of a hypercube: vertically or horizontally.

3. Construction of a triangle

DIRECT deals with the following unbound-constrained optimization problem.

$$\min_{x \in \Omega} f(x)$$

where

$$f : R^N \rightarrow R$$

and

$$\Omega = \{x \in R^N : l \leq x \leq u\}$$

f is a Lipschitz continuous function on Ω, with Lipschitz constant K, i.e.

$$|f(x) - f(y)| \leq K\,|x - y| \quad \forall x, y \in \Omega \qquad (3.1)$$

For the clarity we consider an interval [a, b]. The concave function defined for any y $\in$ [a, b] by f(y) -K |x - y|, x $\in$ [a, b], underestimates *f*.
Let

$$c_1 = (2a + b)/3, \text{ and } c_2 = (a + 2b)/3$$

be two consecutive points in [a, b]. By setting y = c_1, and y = c_2 in (3.1), we get the following inequalities

$$f(x) \geq f(c_1) - K|x - c_1| \quad (3.2)$$

and

$$f(x) \geq f(c_2) - K|x - c_2| \quad (3.3)$$

The function

$$F(x) = \max_{k=1,2} f(c_k) - L|x - c_k| \quad (3.4)$$

formed by the intersection of (3.2) and (3.3) is an underestimator for *f*, its minimum is a lower bound on the least value of *f*. The restriction of *F* to the interval [c_1, c_2], (called tooth) attains its minimal value (downward peak) given by

$$L^* = \frac{f(c_1) + f(c_2)}{2} - K\frac{c_2 - c_1}{2}$$

and it occurs at the point

$$x^* = \frac{c_1 + c_2}{2} + \frac{f(c_1) - f(c_2)}{2K},$$

see figure (3). Note that the lower bounds in figure (3) can be considered as stronger bounds. As a consequence , strong bounds leads to weak optimality and weak bounds gives strong optimality, [11]. But we do not use these bounds as in DIRECT, this only will help us in the construction of a triangle whose basis is the interval in consideration whose length is exactly equals 1/3 of the length of the interval [a, b]. We have chosen this length to be equal 1/3 in order to guarantee that all triangles having the same basis have the same area. We can extend this length to cover the maximum possible length, but this will influence the potential optimality. We can restrict our lower bounds on the center subinterval [c_1, c_2]. Since the peak where *f* is minimal is not the lowest one , we can not see wich interval can be selected to be further explored. As observed before, we see that the downward peak is pointed in the direction of the minimum value: min{f(c_1),f(c_2)}, i.e., x* is close to the side where *f* is minimal at one of these two points. Hence, for each tooth the peak point will be the center of the subinterval located in the projection of the top of the triangle onto the interval, notice that this triangle have two equal sides, see figure (4). An idea to see weather an interval is potentially optimal or not, is to choose a triangle in each iteration. The triangle with the lowest function value at this point is optimal. The distance from the center of the interval [a, b] to the peak point helps us to construct a triangle. If the peak is at the right from the center (resp. to the left), then we should add this distance to the right (resp. to the left) point c_2 (resp. to the left point c_1). The whole interval $[c_1, c_2]$ moves (translation) to the right (resp. to the left), as illustrated in figure (4). Notice that all triangles have the same area. If the peak point coincide with the center, there is nothing to do, see figure (4). We compute the value of the objective function at this point. There is two possibilities to define a potentially optimal intervals, see figure (6). Note that the method of covering by triangles is also considered in [1], where the considered semplexes are regular. The method used a Branch and Bound technique to eliminate unecessary subregions (removal cone). In our method we don't need to remove those subintervals, this is because by the Lipschitz condition, we know that the objective function cannot lie in this region (see [13]).

Let $\delta=x^*-c=(f(c_1)-f(c_2))/2K$, if $f(c_1)>f(c_2)$, and $\delta'=c-x^*=(f(c_2)-f(c_1))/2K$, if $f(c_1)<f(c_2)$. By adding the distance δ, as shown in figure, to the right if $f(c_1)>f(c_2)$, the interval $[c_1, c_2]$ translates to the right, to become $[c_1+\delta, c_2+\delta]$. Similarly, by substracting δ' if $f(c_1)<f(c_2)$, the interval $[c_1, c_2]$ translates to the left, to become $[c_1-\delta', c_2-\delta']$. Note that if the peak point coincides with the center, then δ=0, and the distance is different for each subinterval. Figure (5) shows a two-dimensional translation of the domain. Note that the notion of enlarged rectangle has been used in [7].

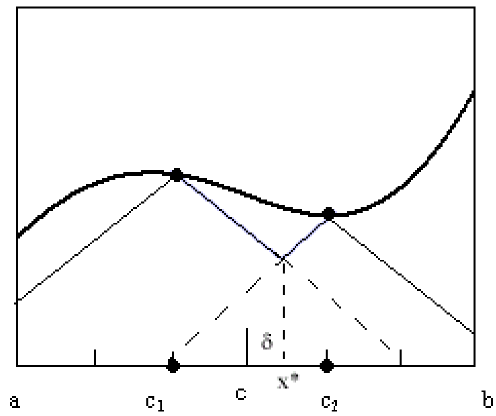


Figure 3. Lower bound with two points sampling method.

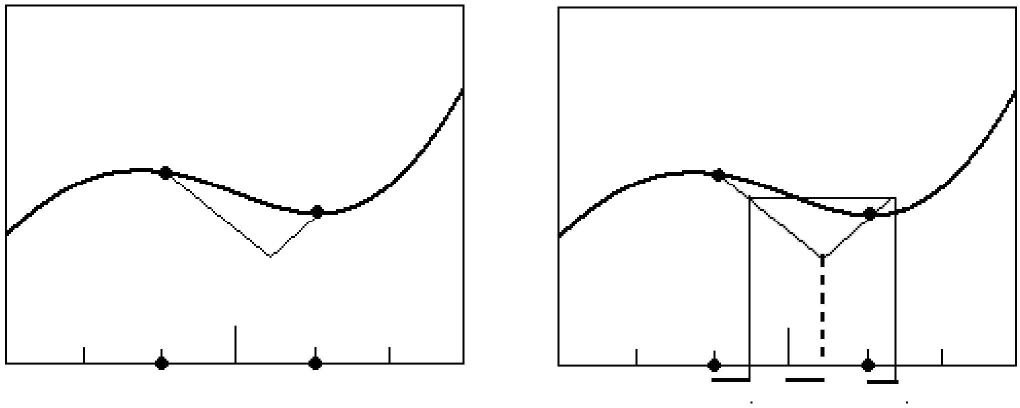


Figure 4. The resulting triangle.

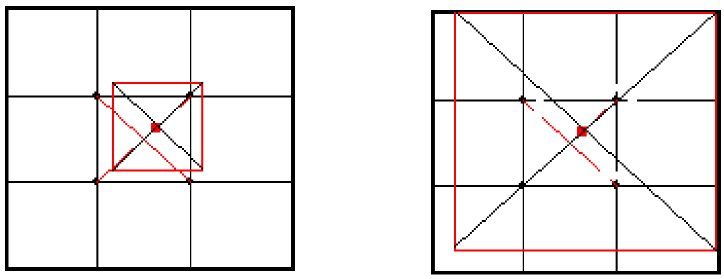


Figure 5. A two-dimensional translation of the domain.

4. Potentially optimal intervals

From the Lipschitz condition (3.1) in one dimensional formulation

$$|f(x) - f(y)| \leq K\,|x - y| \quad \forall x, y \in [c_1 \pm \delta, c_2 \pm \delta]$$

Let $x^* = c \pm \delta$, where $c = (c_1+c_2)/2 = (a+b)/2$, and $\delta = |x^*-c| = |f(c_1)-f(c_2)|/2K$,
By setting $y=x^*$ in (3.1), we get the following inequalities

$$y_r^U = f(x^*) - K(x - x^*) \leq f(x), \text{ if } x \geq x^*$$

$$y_l^U = f(x^*) + K(x - x^*) \leq f(x), \text{ if } x < x^*$$

By setting $x=c_1 \pm \delta$ or $x=c_2 \pm \delta$ in these two inequalities we get a lower bound for f in this interval. This lower bound is given by

$$f(x^*) - K(c_2 - c_1)/2 = f(x^*) - K(b-a)/6$$

The first version of potentially optimal intervals, which is the same as in DIRECT, and it is the most natural one, is then given by following defintion.

Definition 4.1. Let f_{min} be the best function value found up to now. An interval i =[a_i, b_i] is said to be potentially optimal if there exists a subinterval with length equals 1/3 of the length of [a_i, b_i] , and if there exists some rate of change $\widetilde{K}>0$ such that

$$(4.1) \qquad f(x_i^*)-\widetilde{K}(b_i-a_i)/6 \le f(x_j^*)-\widetilde{K}(b_j-a_j)/6, \forall j$$

$$(4.2) \qquad f(x_i^*)-\widetilde{K}(b_i-a_i)/6 \le f_{\min}-\varepsilon|f_{\min}|$$

These conditions in the definition are the same as in DIRECT . Condition (4.2) forces the lower bound for the interval, to exceed the current best solution by a nontrivial amount, ε>0, is a balance parameter between local and global search. In some calculations ε is proposed between 10^{-3} and 10^{-7}.

Remark 4.1. The above definition can be formulated in the sens of potentially optimal triangles.

The second version of potentially optimal can be obtained in the restricted domain [x_i, x_j] as illustrated in figure (6). We determine x_i, by finding where y_l^U (upper left line) and y_l^L (left lower line) interstects. Similarly, x_j is found where y_r^U (upper right line) and y_r^L (lower right line) intersects. These lines are given by

$$y_l^U = f(x^*)+K(x-x^*), \quad y_r^U = f(x^*)-K(x-x^*)$$

and

$$y_l^L = f(c_1)+K(c_1-x); \quad y_r^L = f(c_2)+K(x-c_2)$$

yielding

$$x_i = \frac{f(c_1)-f(x^*)}{2K}+\frac{(x^*+c_1)}{2}$$

and

$$x_j = \frac{f(x^*)-f(c_2)}{2K}+\frac{(x^*+c_2)}{2}$$

A similar definition to 3.1, can be given with the same conditions (4.1) and (4.2).

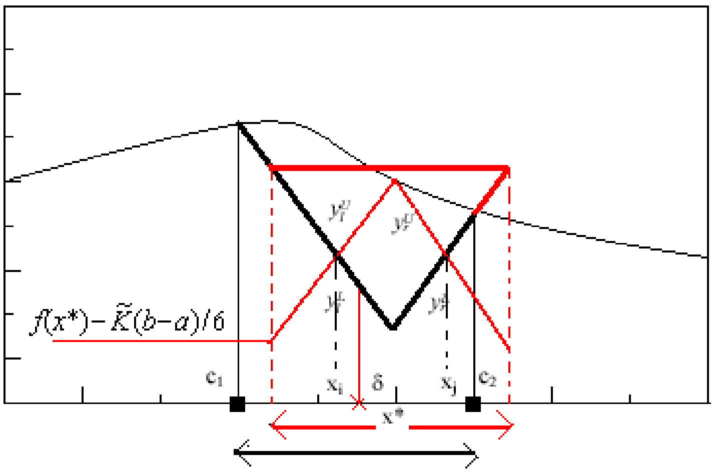


Figure 6. Fonction value corresponding to the peak, with an iIllustration of how to derive two possibilities of potentially optimal intervals.

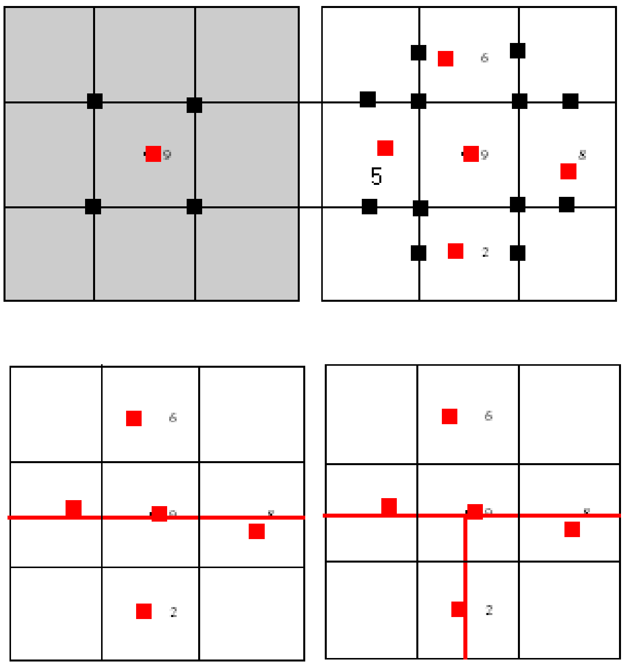


Figure 7. Illustration of the division of a hypercube, in the order given by : $w_i=min\{f(x^*e_i), f(y^*e_i)\}$, first, perpendicular to the direction with the lowest w_i, and then divides the remaining domain perpendicular to the direction of the second lowest w_i. Here $w_1=\min\{5, 8\}=5$, and $w_2=\min\{6, 2\}=2$, i.e., the division is done, first, perpendicular to the y-axis, and then perpendicular to the x-axis

We now formally state a one-dimensional algorithm for the first version.

5. Univariate algorithm.

Step 1. Set m = 1 (evaluation number), $[a_1, b_1] = [l, u]$, $c_1 = (2a_1 + b_1)/3$,
$c'_1 = (a_1 + 2b_1)/3$, and evaluate $f(x^*_1)$, Set $\delta = |x^*-c| = |f(c_1)-f(c_2)|/2K$
Let t = 0 (iteration counter). Set $f_{min} = f(x^*_1)$

Step 2. Identify the set S of all potentially optimal intervals.

The selection of all potentially optimal intervals is done using definition (4.1) by the construction of a triangle.

Step 3. Select any interval $i \in S$, with two consecutive points c_i and c'_i.

Step 4. Let $\alpha = (b_i-a_i)/2$, and set $c_{m+1} = c'_i - \alpha = c_{m+2} - \alpha$, and $c'_{m+2} = c_i + \alpha = c'_{m+1} + \alpha$. Then we have two subintervals, the left one with two consecutive points c_{m+1} and c'_{m+1}. The right subinterval have c_{m+2} and c'_{m+2} as consecutive points. Then, evaluate $f(c_{m+1})$ and $f(c'_{m+2})$, i.e., evaluate $f(x^*_{m+1})$, and $f(x^*_{m+2})$, in the resulting triangle, where $x^*_{m+1} \in [c_{m+1} \pm \delta), c'_{m+1} \pm \delta]$, and $x^*_{m+2} \in [c_{m+2} \pm \delta', c'_{m+2} \pm \delta']$. Then update f_{min}.

Step 5. In the partition, we have the left subinterval $[a_{m+1}, b_{m+1}] = [a_i, a_i + \alpha]$, with two successive points c_{m+1} and c'_{m+1}, and the right subinterval $[a_{m+2}, b_{m+2}] = [b_{m+1}, b_{m+2}] = [a_i + \alpha, b_i]$, with two successive points c_{m+2} and c'_{m+2}. Then modify interval i to be the subinterval with the lowest function value $f(x^*_{m+1})$ or $f(x^*_{m+2})$. i.e., the left or the right subinterval.

Set m = m+ 2.

Step 6. Set $S = S-\{i\}$. If $S \neq \emptyset$, go to step 3.

Step 7. Set t = t+1 . If t < iteration number, then stop. Otherwise, go to step 2.

Remark 5.1. Step (4) can be done in the following way: if we set $\alpha = (b_i-a_i)/6$, and set $c_{m+1} = c_i - \alpha = c'_{m+1} - \alpha$, and $c'_{m+2} = c_{m+2} + \alpha = c'_i + \alpha$.

6. Conclusions

This paper gives a new method for sampling and subdividing the domain search, and presents a simple definition of potentially optimal hyperrectangles. There is a distingished point: the peak point, which is the center of a specific interval. This interval is obtained by a construction of a triangle using the distance from the peak point to the center. The result is a translation of the initial interval towards the lowest function value, and then, overcoming the disadvantage if the global minimum lies at the boundaries. The idea is to evaluate the objective function at this point. A new version of potentially optimal intervals is given. Future work should be done on numerical tests to compare both with DIRECT, and MCS [10], for functions for which global minimizers are at the boundaries.

References

[1] J. Clausen, A. Žilinskas, Subdivision, Sampling, and Initialization Strategies for Simplical Branch and Bound in Global Optimization, Computers and Mathematics with Applications 44 (2002) 943-955

[2] L. Chiter, (2005), Towards a new DIRECT algorithm: A two-points based sampling method (e-print): http://www.optimization-online.org /DB-FILE/2005/03/1077. pdf.

[3] D. E. Finkel, DIRECT Optimization Algorithm User Guide, Center for Research in Scientific Computation, North Carolina State University, Raleigh, march (2003).
[4] D. E. Finkel, C. T. Kelley, Convergence Analysis of the DIRECT algorithm, North Carolina State University, Center for Research in Scientific Computation, Raleigh, version of july 14, (2004).
[5] D. E. Finkel, C. T. Kelley, New Analysis of the DIRECT Algorithm, Copper Mountain Conference On Iterative Methods, April 2, (2004).
[6] J.M. Gablonsky, An implementation of the DIRECT Algorithm, North Carolina State University, Center for Research in Scientific Computation, Raleigh, version of August 20, (1998).
[7] J.M. Gablonsky, DIRECT version 2.0 userguide, Technical Report, CRSC-TR01-08, Center for Research in Scientific Computation, North Carolina State University, april (2001).
[8] J.M. Gablonsky, Modifications of the DIRECT Algorithm, PhD Thesis, North Carolina State University, Raleigh, North Carolina, (2001).
[9] J.M. Gablonsky, C.T. Kelley, A locally-biased form of the DIRECT algorithm. Journal of Global Optimization, 21 (2001) 27-37.
[10] W. Huyer, A. Neumaier, Global optimization by multilevel coordinate search, J. Global Optim.14 (4) 331-355, (1999), Kluwer Academic Publisher.
[11] D.R. Jones, C.D. Perttunen, B.E. Stuckman, Lipschitzian optimization without the lipschitz constant, Journal of Optimization Theory and Application, 79:157 (1993).
[12] D. R. Jones, DIRECT, Encyclopedia of optimization, (2001), to be published
[13] Z. B. Zabinsky, R. L. Smith, B. P. Kristinsdottir, Optimal Estimation of Univariate Black Box Lipschitz Functions with Upper and Lower Error Bounds.