Another face of **DIRECT**

Lakhdar Chiter

Department of Mathematics, Setif University, Setif 19000, Algeria

E-mail address: chiterl@yahoo.fr

Abstract

It is shown that, contrary to a claim of [D. E. Finkel, C. T. Kelley, Additive scaling and the DIRECT algorithm, J. Glob. Optim. 36 (2006) 597-608], it is possible to divide the smallest hypercube which contains the low function value by considering hyperrectangles whose points are located on the diagonal of the center point of this hypercube using a division procedure which influences the slope to be below the the threshold fmin - ε |fmin| and thus reduces the influence of the parameter ε.

Keywords: DIRECT ; Global optimization; Geometrical interpretation

1. Introduction

The DIRECT (DIviding RECTangles) algorithm of Jones et al. [2] is a deterministic sampling method designed for bound constrained non-smooth problems. There are two main components in DIRECT: one is its strategy of partitioning the search domain, the other is the identification of potentially optimal hyperrectangles, i.e., having potential to contain good solutions. The relative potential of each partition is characterized by two attributes: the value of the objective function at the center of the partition, and its size. The former provides a measure of the potential of the partition with respect to local search, i.e., partitions with good function values at their center are more desirable than those with worse function values. The latter provides a measure of the partition's potential with respect to global search, that is, larger partitions contain more unexplored territory, and therefore provide a greater opportunity for further exploration. A parameter is used to control the balance between local and global search and protect the algorithm against excessive emphasis on local search. The effects of this parameter and its influence on the rapidity of the convergence were studied in a recent paper by Finkel and Kelley [1]. An algorithm that places too much emphasis on local search will easily be trapped near a local optimum. Conversely, an algorithm that spends too much time performing global search will converge very slowly. Therefore, a technique that does not depend on such parameter would be desirable. This paper is concerned with such techniques.

As reported by many authors, the principal disadvantage of DIRECT is its slow local convergence, i.e, when the size of hyperrectangles becomes too small. After many iterations of DIRECT the smallest hyperrectangle containing f_{min} (the current best function value) is always a hypercube, this hypercube will not be candidate for selection because of the parameter ε. Finkel and Kelley [1], in their paper (see thm. 3.1.), show that the hypercube, which contains the current minimal point will not become potentially optimal until all larger hyperrectangles having their centers on the stencil are the same size as of this hypercube, i.e., after all these hyperrectangles have been divided.

The aim of this paper is to show that the result of theorem 3.1 in [1] is not necessary relevant. The fact that the hypercube with the low function value is rejected from selection because of the influence of hyperrectangles whose points (centers) are on the stencil, it does not follow that this will happen with hyperrectangles whose points (quarters) are on the diagonal. This should not be necessary the case as we shall see in sect. 4. We propose a modification to DIRECT. The modified method produces more hyperrectangles than the original DIRECT by evaluating the objective function at the quarters of each hyperrectangle, thus minimizing the number of evaluations. Each hyperrectangle in the division will have, after division, two points. Each point can be seen as a quarter (1/4) or (3/4) in the other face according to each directions. Using this procedure, we can discard all potentially optimal hyperrectangles on the stencil which prevent the smallest hypercube for being optimal, by considering the points on the diagonal. We can seek the values of *f* that allows us to have a sufficient decrease on the slope to the left of the potentially optimal hyperrectangle and thus increases the slope to the right of the smallest hypercube. Modifications using space partitioning technique have been investigated by many authors, see for example, tree-Direct [5].

This paper is organized as follows: In the next section, we give a short description of DIRECT. In Section 3, we give the geometrical interpretation of condition 5 of theorem 3.1 of [1]. This allows us to describe our modification to DIRECT in section 4, and give a weaker condition which prevents the smallest hypercube for not being optimal. Then we conclude in section 5.

2. DIRECT

The DIRECT algorithm begins by scaling the design domain, Ω, to an *n*-dimensional unit hypercube. This has no effects on the optimization process. DIRECT initiates its search by evaluating the objective function at the center point of , $c = (1/2, ..., 1/2)$. is identified as the first potentially optimal hyperrectangle. The DIRECT algorithm begins the search process by evaluating the function f in all directions at the points $c \pm \delta e_i$ which are determined as equidistant to the center c. Where δ is the one third of the distance of the hypercube, and e_i is the i^{th} unit vector. The DIRECT moves to the next phase of the iteration, and divides the first potentially optimal hyperrectangle. The division procedure is done by trisecting in all directions. The trisection is based on the directions with the smallest function value. This is the first iteration of DIRECT. The second phase of the algorithm is the selection of potentially optimal hyperrectangles. A definition for this is is given below. Sampling of the maximum length directions prevents boxes from becoming overly skewed and trisecting in the direction of the best function value allows the largest rectangles to contain the best function value. This strategy increases the attractiveness of searching near points with good function values. More details about DIRECT can be found in [2].

Definition 2.1. *Assuming that the unit hypercube with center c_i is divided into m hyperrectangles, a hyperrectangle j is said to be potentially optimal if there exists rate-of-change constant $\widetilde{K}$ such that*

$$f(c_j) - \widetilde{K}d_j \leq f(c_i) - \widetilde{K}d_i, \quad for\ i = 1,...,m \qquad (2.1)$$

$$f(c_j) - \widetilde{K} d_j \le f_{\min} - \varepsilon |f_{\min}| \qquad (2.2)$$

Where f_{min} is the best function value found up to now, d_i is the distance from the center point to the vertices, and the parameter ε is used here to protect the algorithm against excessive local bias in the search. The set of potentially optimal hyperrectangles are those hyperrectangles defining the bottom of the convex hull of a scatter plot of hyperrectangle diameter versus $f(c_i)$ for all rectangle centers c_i, see Fig.1. In this graph, the first equation (2.1), forces the selection of the rectangles in the lower right convex hull of dots. Condition (2.1) can be interpreted by the slopes of the linear curves represented to the right and to the left of the point $P(d_i, f(c_i))$. If the slopes of the curves passing through P and the points on the right of this one are all greater than those passing by P and the points on the left of this one, then there exists some $\widetilde{K}$ >0 verifying (2.1). The condition (2.2) forces more the choice of boxes in terms of size. In fact, the hyperrectangle i will be selected only if the slopes of curves on the right of P are greater than the line passing by P and f_{min}. This allows not to select very small boxes and so to stop the convergence earlier. The selection presented here allows to explore at the same time boxes with important sizes to realize a global search and boxes with small sizes to carry out a local search. The parameter influences the slope of the line passing by P and f_{min}. More this slope is weak ($\varepsilon = 0$), more hyperrectangle with small size are selected and thus we do a local search. If ε is close to 1, the slope of this curve is more strong and only few hyperrectangles of small size are selected. We have then a global search.

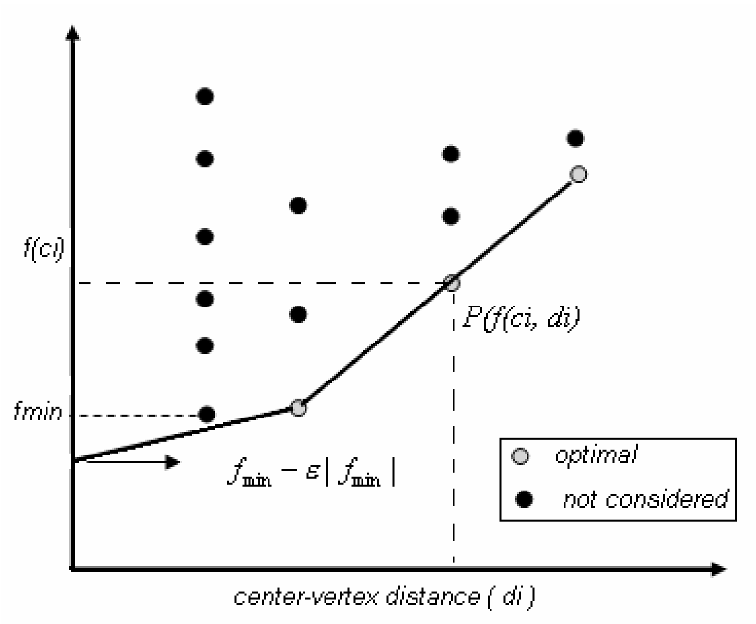

Fig1. Interpretation of definition 2.1.

3. Geometrical interpretation

In this section we describe how the smallest hypercube with the lowest function value, $f(c) = f_{min}$, can be discarded from being optimal, i.e., in the sens that it does not satisfy condition (2.2), the condition which uses the parameter ε. We will further give a geometrical interpretation of this from a nice theorem due to Finkel et and Kelley [1]. In this paper we are not concerned with this parameter. In the Fig. 2, we can see that the slopes are stronger to the right, this is the global part of the algorithm. As the algorithm continues, the search will be local, and the size of the hyperrectangles becomes too small and thus the slopes are too weaker. This prevents to not selecting hypercube with small function value. In the Fig. 2, the square dot alters the lower convex hull, and the small hyperrectangle, which contains the low function value is not potentially optimal. The line going through the points $(\alpha_T, f(c_T))$ and $(\alpha_R, f(c_R))$ cannot be below the quantity $f_{min} - \varepsilon |f_{min}|$. This is due to the larger hyperrectangle to the right having a comparable value of f at its center. For best understanding, the following theorem (see [1]) explains how a hyperrectangle containing f_{min} does not satisfy condition (2.2).

Theorem 3.1. (see [1]) Let $f: R^n \to R$ be a Lipschitz continuous function with Lipschitz constant K. Let S be the set of hyperrectangles created by DIRECT, and let R be a hypercube with a center c_R and side length 3^{-l}. Suppose that

(i) $\alpha_R \leq \alpha_T$, for all $T \in S$ (i.e. R is in the set of smallest hypercubes).

(ii) $f(c_R) = f_{min} \neq 0$ (i.e. $f(c_R)$ is the low function value found).

If

$$\alpha_R < \frac{\varepsilon |f(c_R)|}{2K} \left(\sqrt{n+8} - \sqrt{n} \right) \qquad (3.1)$$

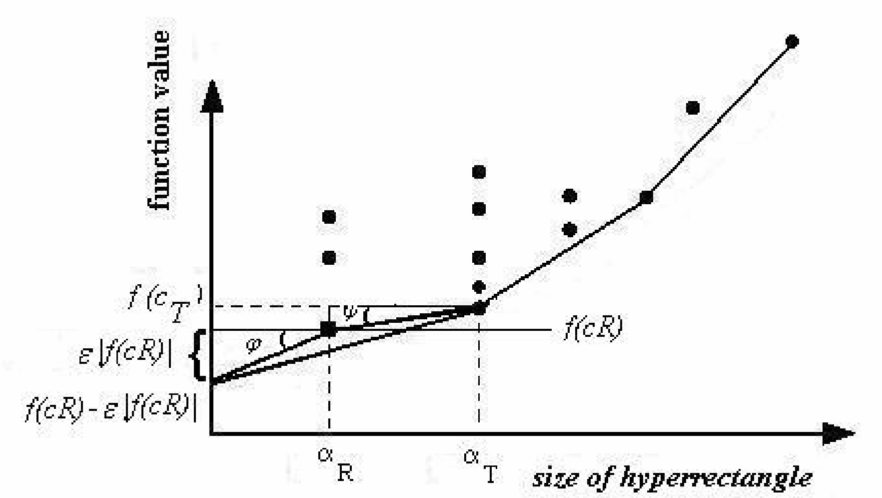

Fig.2. The geometrical interpretation of conditions (2.2) and (3.1).

then R will not be potentially optimal until all hyperrectangles in the "*neighborhood*" of R, i.e., all hyperrectangles whose centers are on the stencil $c_R \pm 3^{-l} e_i$ for $i = 1, ..., N$ are the same size as R.

Remark 3.1. Condition (3.1) in the above theorem can be interpreted by the the following inequality

$$\frac{\varepsilon|f(c_R)|}{\alpha_R} > \frac{f(c_{\tilde{T}}) - f(c_R)}{\alpha_{\tilde{T}} - \alpha_R} \qquad (3.2)$$

Where $\alpha_{\tilde{T}}$ is the size of a the smallest hyperrectangle, (see [1] for details). In fact, if condition (2.2) in the definition 2.1, is not satisfied, i.e.,

$$f(c_R) - \tilde{K}\alpha_R > f_{\min} - \varepsilon|f_{\min}|$$

Then

$$\tilde{K} < \frac{\varepsilon|f(c_R)|}{\alpha_R}$$

But, $$\tilde{K} = \frac{f(c_{\tilde{T}}) - f(c_R)}{\alpha_{\tilde{T}} - \alpha_R}, \text{ and } \tilde{K} \le \frac{2K}{\sqrt{n+8} - \sqrt{n}}$$

However, the condition if in the above theorem shows that if condition (3.2) is satisfied, then the hypercube R will not be potentially optimal. Geometrically, condition (3.2) is represented in Fig. 2, by the tangent of the angle ϕ which is greater than the tangent of the angle ψ, where

$$\tan\phi = \frac{\varepsilon|f(c_R)|}{\alpha_R}, \text{ and } \tan\psi = \frac{f(c_{\tilde{T}}) - f(c_R)}{\alpha_{\tilde{T}} - \alpha_R}$$

In their paper, Finkel and Kelley. [1], have chosen to reduce the influence of the parameter ε. Their modification to `DIRECT` relates to an update to the definition of potentially optimal hyperrectangles. Modifications related to the parameter ε can also be found in [3] for small values, i.e., $\varepsilon \to 0$. In her thesis, Runser [6] suggested to create an optimization process using `DIRECT`, but from a certain size of a hyperrectangle, we add a search method by descent to converge rapidly. Our modification is related to the potentially optimal hyperrectangle *T*, which influences the slope and then preventing hypercube *R* for being optimal, see Fig 2. We seek for values of *f* that allows us to have a sufficient decrease on the slope to the left of hyperrectangle *T* and thus increases the slope to the right of hypercube *R*.

4. New interpretation of DIRECT

This section shortly describes some changes to `DIRECT`. We suggest to change the way a hypercube is divided. Instead of trisecting a hypercube according to the strategy of the lowest function values, we suggest to use a division as described below. A hyperrectangle is only bisected once along its longest side. This means that, this increases the number of hyperrectangles, therefore the search is done firstly more global. Once all hyperrectangles have been divided, the search becomes more local. This strategy does not places the lowest function values in the largest hyperrectangles. For a better understanding we start with an interval and then describe the more general case of a hyperrectangle. DIRECT can be seen as follows: Trisecting an interval is equivalent to divide it in two inequal subintervals, such that one is half of the other. This is done in the ratio *1/3* or *2/3* as illustred in Fig.3. Then we divide the large interval in two equal subintervals. In two dimensions, the corresponding division is shown in Fig. 4. Remark that there is no difference with the division procedure (a) or (b), since the value of *f* at this each point (except the center) is the same for the two shaded domains, as shown in Fig. 4, because we do not need to leave the best function value in a largest space as in DIRECT, and the size which is represented by the longest distance from the evaluation point to the vertices is the same. Each point, except the one of the center, belongs to *N* hyperrectangles in all directions, where *N* is the dimension of the domain. The size of a hyperrectangle is represented by the longest distance from the evaluation point which is located in the quarters of the rectangle to one of its corners. This measure can be interpreted as the half of a hyperrectangle for the DIRECT, which was called in [2-4] by the l^2 diameter. For more details about the size we refer to[4].

The points sampled are equidistant from the center. The distance from the center to each point in the stencil is represented by 2^{-l}, where 3.2^{-l} is the side lenght of a hypercube, as shown in Fig. 5. The function *f* is evaluated at $c \pm 2^{-l} e_i$, where e_i is the *i*th unit vector. In the Fig. 6 we show an example of the first three iterations.

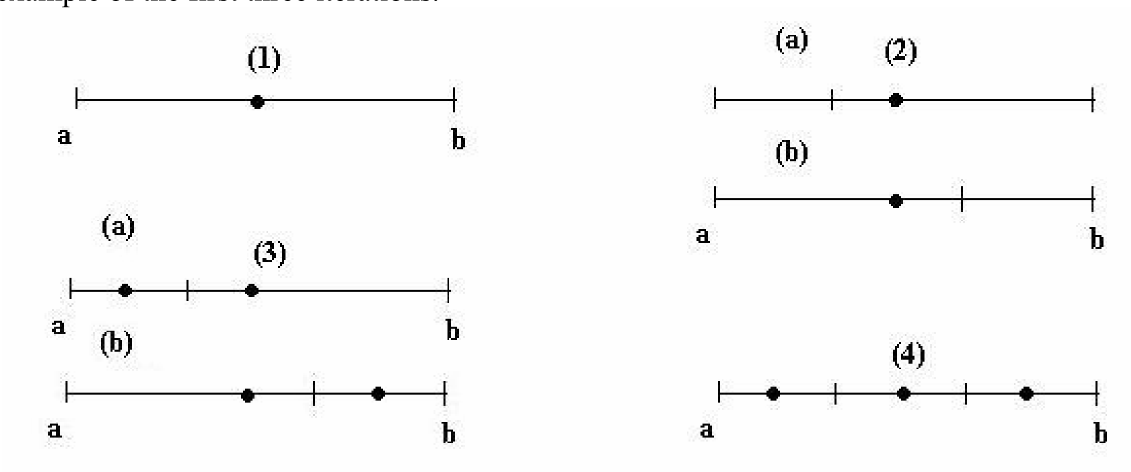

Fig.3. Division of an interval. In (2), it does not matter which division we choose (a) or (b), since the value of *f* at this point is the same.

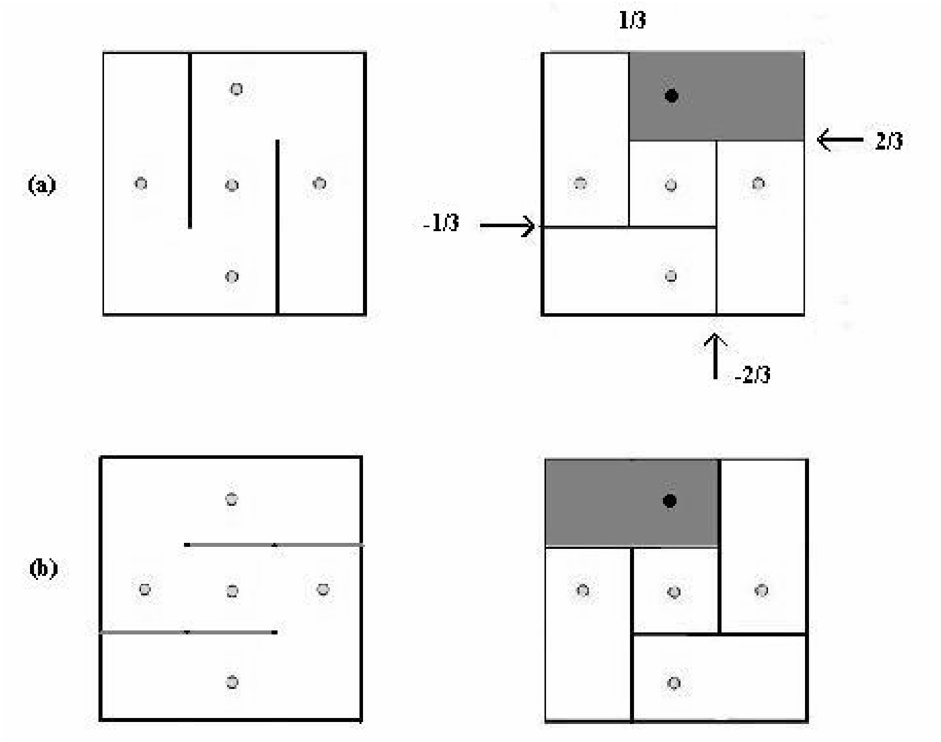

Fig.4. division of a square. As for one dimension, we can choose the division (a) or (b).

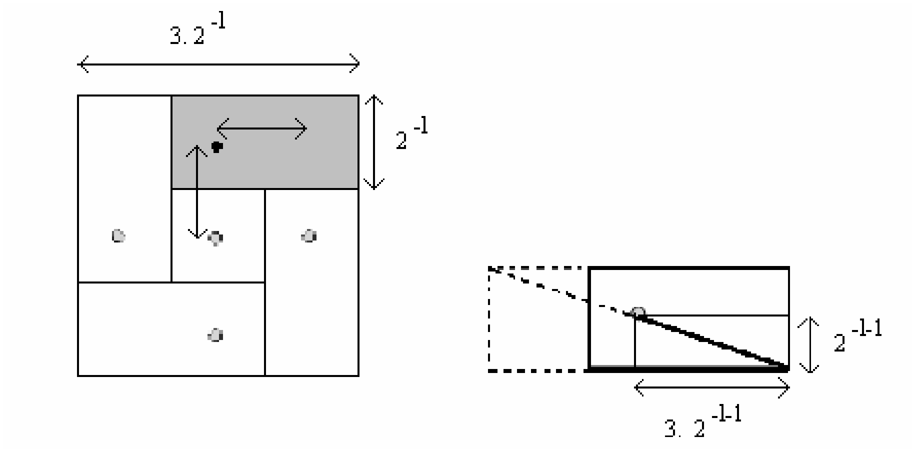

Fig.5. Size of a hyperrectangle. The distance from the sampled point to the vertex can be seen as the diameter of the small circonscribed rectangle.

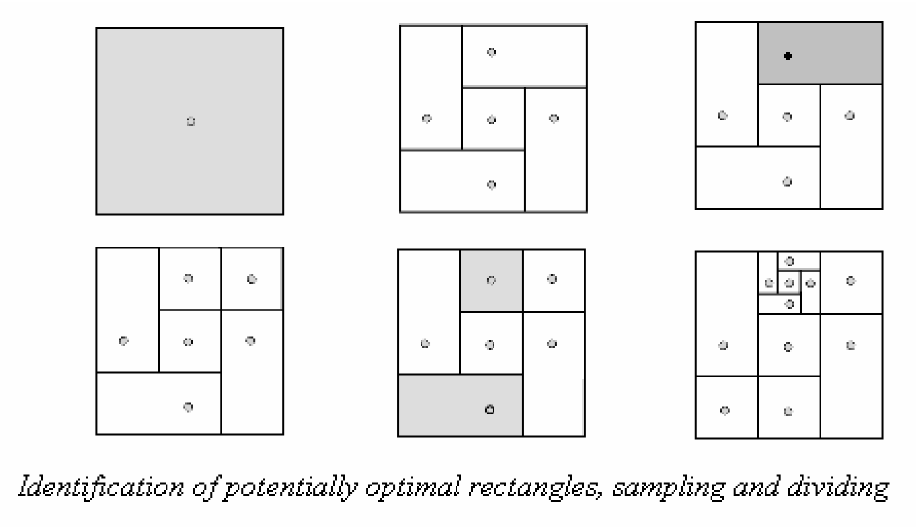

Fig. 6. An example of the first three iterations.

The division of a three dimensional rectangle is represented in Fig. 7. the only difference in the above examples, is that this division takes into account the fact that the lower function value is in a larger space, only in the first stage of the division.

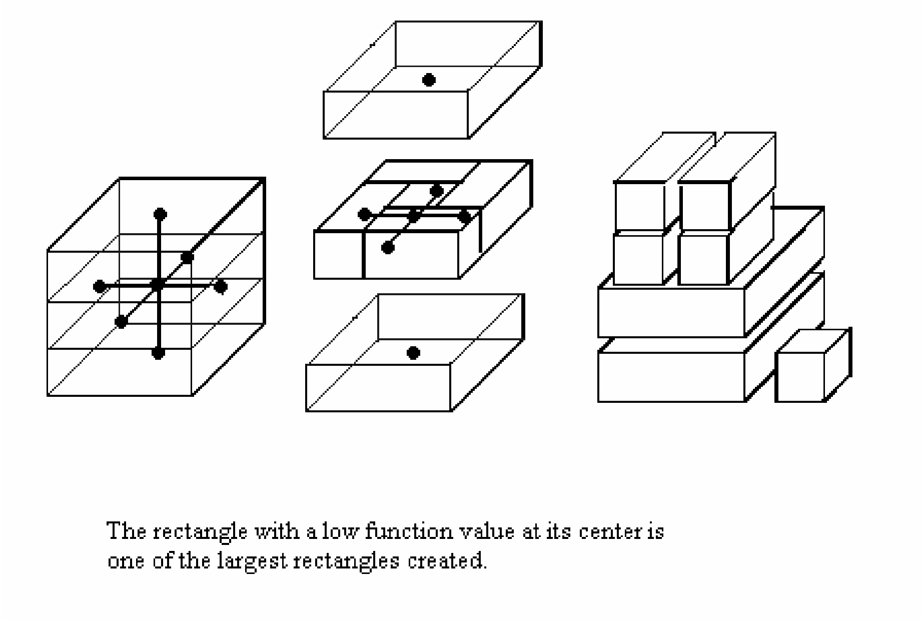

Fig. 7. Dividing a three dimensional rectangle

In what follows a similar result as in theorem 3.1 of [1], with a weaker condition. We adopt the notations of theorem 3.1, and we can choose either 2^{-l} or 3^{-l} as the size of the smallest hypercube. The following theorem shows that if we are in the conditions of theorem 3.1, i.e., a hypercube cannot be potentially optimal, it will not be necessary the case for hyperrectangles having their points on the diagonal. But if condition (4.1) is satisfied then condition (3.1) holds. This estabilishes the following theorem.

Theorem 4.1. Let $f: R^n \rightarrow R$ be a Lipschitz continuous function with Lipschitz constant K, Let S be the set of hyperrectangles created by DIRECT, and let R be a hypercube with a center c and side length 2^{-l}. Suppose that (i) and (2) of theorem 3.1. are satisfied.
If

$$\alpha_R < \frac{\varepsilon|f(c)|}{2\sqrt{n}K}\left(\sqrt{n+8}-\sqrt{n}\right) \qquad (4.1)$$

then R will not be potentially optimal until all hyperrectangles whose points are on the diagonal $c \pm 2^{-l}\sqrt{n}$ are the same size as R. i.e., R will neither be optimal for hyperrectangles on the stencil $c \pm 2^{-l} e_i$, nor for those on the diagonal.

Remark 4.1. Theorem 3.1 is still valid for our case, since the measure of a hyperrectangle is half of the diameter of the a hyperrectangle as shown in Fig. 5. Note that we can use either 3^{-l} or 2^{-l} as the smallest side length, and the same conclusion holds, for the simple raison that the terms 3^{-l} or 2^{-l} will be simplified as seen in theorem 3.1 of [1]. A hyperrectangle T 2 S will have n - 1 sides of length $(3^{-l})/2$ and one side of length $3^{-}l.(3/2)$, i.e.,

$$\alpha_T = \sqrt{(n-1)\left(\frac{3^{-l}}{2}\right)^2 + \left(3^{-l}\frac{3}{2}\right)^2} = \frac{3^{-l}}{2}\sqrt{n+8}$$

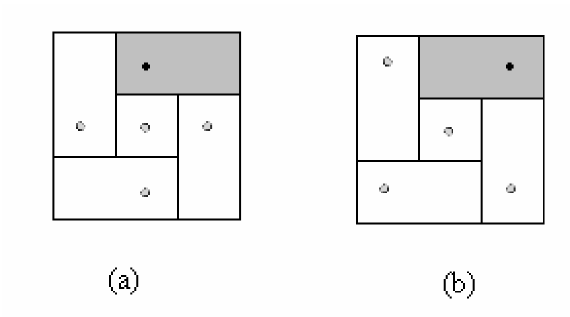

Fig.8. Points represented on the stencil (a), and in (b) on the diagonal.

And if we use 2^{-l} as a size we will have $\alpha_T = \sqrt{(n-1)\left(2^{-l-1}\right)^2 + \left(3.2^{-l-1}\right)^2} = (\frac{2^{-l}}{2})\sqrt{n+8}$

Proof. The first affirmation is immediate since

$$\frac{\varepsilon|f(c)|}{2\sqrt{n}K}(\sqrt{n+8}-\sqrt{n}) \leq \frac{\varepsilon|f(c)|}{2K}(\sqrt{n+8}-\sqrt{n}), \quad for\ n \geq 1.$$

It is easy to remark that in two dimensions, the points on the diagonal are $c \pm 3^{-l}(e_1 \pm e_2)$, and the distance from c to these points is $2^{-l}\sqrt{2}$. For a three dimensional cube, this distance is $2^{-l}\sqrt{3}$, and for a hypercube we get $2^{-l}\sqrt{n}$. We adopt the same proof as in [1].

If all hyperrectangles in Fig.8 satisfy condition 3.1, we can use hyperrectangles having their points on the diagonal, since the size is the same in (a) or (b). We get the following corollary.

Corollary 4.2. *Let f and R be as in theorem 3.1. Suppose that conditions (i) and (ii) of theorem 3.1 holds. If*

$$\frac{\varepsilon|f(c)|}{2\sqrt{n}K}\left(\sqrt{n+8}-\sqrt{n}\right) \leq \alpha_R \leq \frac{\varepsilon|f(c)|}{2K}\left(\sqrt{n+8}-\sqrt{n}\right) \qquad (4.2)$$

then R will be potentially optimal for all hyperrectangles whose points are on the diagonal $c_R \pm 2^{-l}\sqrt{n}$.

By using the points on the diagonals for a potentially optimal hyperrectangle, we get a correction on the lower conex hull as seen in Fig. 9.

Proof. The right hand side of inequality means that R will not be potentially optimal for hyperrectangles on the stencil. While the left hand side means that, R will be potentially optimal, i.e., R satisfy condition (2.2) of definition 2.1. For hypercube R to be potentially optimal there must exist $\tilde{K}$ such that (2.1) and (2.2) hold. From condition (2.1) we get

$$\tilde{K} \geq \frac{f(c)-f(c_T)}{\alpha_R - \alpha_T}$$

we must choose

$$\tilde{K} = \max \frac{f(c)-f(c_T)}{\alpha_R - \alpha_T} = \frac{f(c)-f(c_{\tilde{T}})}{\alpha_R - \alpha_{\tilde{T}}}$$

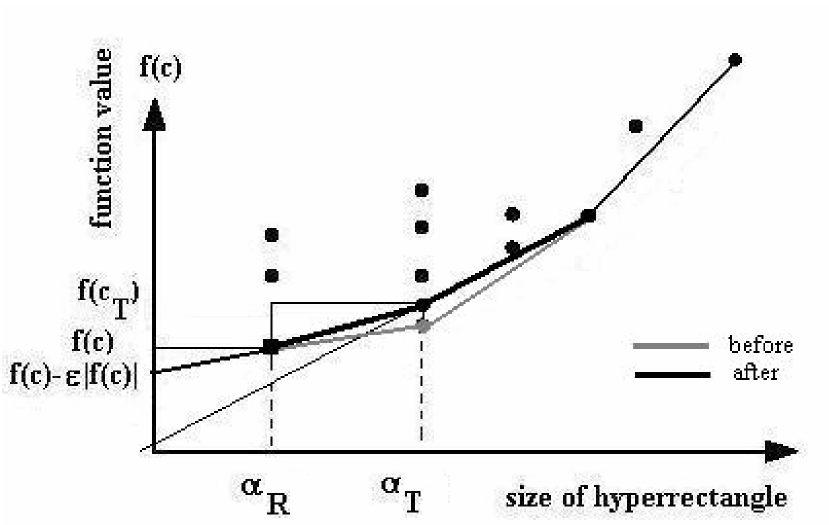

Fig.9. A correction on the lower convex hull. Hyperrectangle which satisfy condition (3.1) of theorem 3.1 is discarded and replaced by another hyperrectangle for which the value of *f* is much greater.

The left hand side of inequality (4.2) is equivalent to

$$\frac{\varepsilon|f(c)|}{\frac{2^{-l}\sqrt{n}}{2}} \le \frac{2\sqrt{n}K}{\sqrt{n+8}-\sqrt{n}}; \qquad (4.3)$$

and

$$\alpha_T - \alpha_{\tilde{T}} \le -\left(\frac{2^{-l}}{2}\right)\left(\sqrt{n+8}-\sqrt{n}\right)$$

The Lipschitz continuity of *f* is equivalent to $f(c) - f(c_{\tilde{T}}) \ge -K2^{-l}\sqrt{n}$.
Thus

$$\tilde{K} = \frac{f(c) - f(c_{\tilde{T}})}{\alpha_R - \alpha_{\tilde{T}}} \ge \frac{-K2^{-l}\sqrt{n}}{-\left(\frac{2^{-l}}{2}\right)\left(\sqrt{n+8}-\sqrt{n}\right)} = \frac{2\sqrt{n}K}{\sqrt{n+8}-\sqrt{n}}$$

From the right hand side of (4.2) and inequality (4.3), we get

$$\tilde{K} \ge \frac{\varepsilon|f(c)|}{\frac{2^{-l}\sqrt{n}}{2}}, \text{ i.e., } f(c) - \tilde{K}\alpha_R \le f_{\min} - \varepsilon|f_{\min}|.$$

Remark 4.2. The set of points $f(c_T)$ such that the hypercube R must be potentially optimal are the points $(f(c_T), \alpha_T)$ which are between the lines (D) and (D') as seen in Fig. 10.
Let $y = ax + b$, be the equation of (D). Then

$$y = \frac{\varepsilon|f(c)|}{\alpha_R}x + f(c) - \varepsilon|f(c)|$$

If $(f(c_T), \alpha_T) \in (D)$, we get

$$\frac{f(c_T) - f(c)}{\alpha_T - \alpha_R} = \frac{\varepsilon|f(c)|}{\alpha_R},$$

then the points for which the hypercube R must be potentially optimal are such that

$$\frac{f(c_{\tilde{T}}) - f(c)}{\alpha_T - \alpha_R} \geq \frac{f(c_T) - f(c)}{\alpha_T - \alpha_R} \geq \frac{\varepsilon|f(c)|}{\alpha_R}.$$

The right hand side of the above inequation is to satisfy condition (2.2) of definition 2.1, while the left inequality is due to the convexity of the lower convex hull.

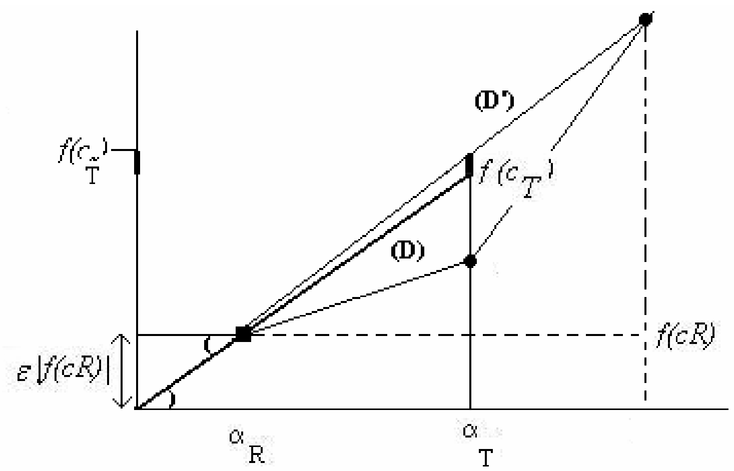

Fig. 10. The points for which the hypercube R must be potentially are those points such The right hand side of the above inequation is to satisfy condition (2.2) in definition 2.1, and the second inequality to the left, because of the convexity of the lower convex hull.

5. Conclusion

In this paper we have presented a modification in the division procedure for the DIRECT algorithm. The hyperrectangles created can have many faces depending on the dimension. For each potentially optimal hyperrectangle the size is measured as the longest distance from the evaluation points, which are located on the quarters, to its vertices. DIRECT uses a parameter which influences the smallest hyperrectangle containing f_{min} (the current best function value) for not being potentially optimal, and thus may influences the convergence of the algorithm. The smallest hyperrectangle is always a hypercube, which depends on the hyperrectangles whose centers are on the stencil of the center point of this hypercube. We can discard the potentially optimal hyperrectangle, which influences the slope and then preventing the smallest hypercube for being optimal, by considering hyperrectangles having their points on the diagonal. Then it is possible to find a potentially hyperrectangle for which the smallest hypercube will be potentially optimal.

References

[1] D. E. Finkel, C. T. Kelley, Additive scaling and the DIRECT algorithm, J. Glob. Optim. (2006) 36:597-608, Springer, 2006.
[2] D.R. Jones, C.D. Perttunen, B.E. Stuckman, Lipschitzian Optimization without Lipschitz Constant, Journal of Oprtimization Theory and Applications, Vol. 79, No. 1, October 1993.
[3] J. M. Gablonsky, C. T. Kelley, : A locally-biased form of the DIRECT algorithm. J. Global. Optim. 21, 27-37, 2001.
[4] J.M. Gablonsky, Modifications of the DIRECT Algorithm, Ph.D. Thesis, North Carolina State University, Raleigh, North Carolina, 2001.
[5] R. E. Haskell, C. A. Jackson, "Tree-Direct: An Efficient Global Optimization Algorithm," Proc. International ICSC Symposium on Engineering of Intelligent Systems, University of La Laguna, Tenerife, Spain, February 11-13, 1998.
[6] K. J. Runser, Méthodologies pour la planification de réseaux locaux sans-fil, PhD Thesis, France: INSA de Lyon, 27 octobre 2005, 230p. (in French), downloadable from : http://katia.runser.free.fr/references.html