New Interpretation of the DIRECT algorithm

Lakhdar Chiter

Department of Mathematics, Setif University
19000 Setif, Algeria
chiterl@yahoo.fr

Abstract

In this note we show that a modification in the division procedure for the DIRECT algorithm can remove the sensitivity, which affects the smallest hypercube containing the low function value, for being potentially optimal.

Mathematics Subject Classifications: 65K10, 90C30

Keywords: DIRECT algorithm; Global optimization; division procedure

1 Introduction

The DIRECT algorithm of Jones et al.[2] is an acronyme for DIviding RECTangles. It is a deterministic sampling method designed for finding the global minima for bound constrained non-smooth problems. The main components of DIRECT are the identification of potentially optimal hyperrectangles, i.e., those hyperrectangles having potential to contain good solutions, and its strategy of division. Two conditions regulate this selection by using a parameter whose role is to prevent the algorithm of becoming too local. A disadvantage of this algorithm is its slow local convergence, i.e, when the diameter 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 this parameter. It is observed that this hypercube depends on the hyperrectangles whose centers are on the neighborhood of this hypercube, and thus on the value of $f$ on these centers. Finkel and Kelley [1], have chosen to reduce the influence of the parameter $\varepsilon$. Their modification to DIRECT relates to an update to the definition of potentially optimal hyperrectangles. In this paper we are not concerned with $\varepsilon$. To overcome this situation, we propose a modification to DIRECT which produces more hyperrectangles than the original DIRECT. We can discard the potentially optimal hyperrectangle, which influences the slope
and then preventing the smallest hypercube for being optimal, by considering the points located on the diagonal. This 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.

2 DIRECT overview

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 Ω, \( c = (1/2, ..., 1/2) \). Ω 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 \( \delta \) 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 process 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 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

\[
\begin{align*}
    f(c_j) - \widetilde{K}d_j & \leq f(c_i) - \widetilde{K}d_i \quad \text{for } i = 1, ..., m \\
    f(c_j) - \widetilde{K}d_j & \leq f_{\min} - \varepsilon |f_{\min}|
\end{align*}
\]

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 \( \varepsilon \) 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 Figure1.

In this graph, the first equation (1), forces the selection of the rectangles in the lower right convex hull of dots. Condition (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 $\tilde{K} > 0$ verifying (1). The condition (2) forces more the choice of boxes in term 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 $\varepsilon$ 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.

3 A geometric 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), the condition which uses the parameter $\varepsilon$. We will further give a geometrical interpretation of this from a theorem due to Finkel and Kelley [1]. 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 Figure 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. The following theorem (see theorem 3.1., [1]) explains how a hyperrectangle containing $f_{\min}$ does not satisfy condition (2). For best understanding, this theorem is given below.

**Theorem 1 (1)** Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 value found).

If

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

(3)
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 1** Condition (3) in the theorem can be interpreted by the following inequality

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

(4)

Where $\alpha_\tilde{T}$ is the size of the smallest hyperrectangle, (see [1] for details). In fact, if condition (2) in the definition 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

$$\tilde{K} = \frac{f(c_\tilde{T}) - f(c_R)}{\alpha_\tilde{T} - \alpha_R}$$

and

$$\tilde{K} \leq \frac{2K}{\sqrt{n + 8} - \sqrt{n}}$$

This means that if condition (4) is satisfied, then the hypercube $R$ will not be potentially optimal. Geometrically, condition (4) is represented in Figure 2, by the tangent of the angle $\phi$ which is greater than the tangent of the angle $\psi$, 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 shown how DIRECT is sensitive to the parameter $\varepsilon$, and how it is affected by additive scaling. The modification suggests then an update to the definition of potentially optimal hyperrectangles. Our modification is related to the potentially optimal hyperrectangle $T$, which influences the slope and then preventing hypercube $R$ for being optimal. 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 dimensions 1 and 2). For better understanding, this division is shown in Figure 3. in two dimensions, and in Figure 4. for one dimension. 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 Figure 3, 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. We will come back about this modification in a forthcoming work.

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 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 length of a hypercube, as shown in Figure 5. The function $f$ is evaluated at $c \pm 2^{-l}e_i$, where $e_i$ is the $i$th unit vector.

In what follows a similar result as in theorem 3.1 of [1], with a weaker condition similar to (3). We adopt the notations, and we can choose either $2^{-l}$ or $3^{-l}$ as the size of the smallest hypercube.

**Theorem 2** Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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) $\alpha_R \leq \alpha_T$, for all $T \in S$ (i.e. $R$ is in the set of smallest hypercubes).
(ii) $f(c) = f_{\min} \neq 0$ (i.e. $f(c)$ is the low value found).

If

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

(5)

then $R$ will not be potentially optimal until all hyperrectangles in the “neighborhood” of $R$, i.e., 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 2** Theorem 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 Figure
5. Note that we can use either $3^{-l}$ or $2^{-l}$ as the smallest side length, and the same conclusion holds, because the terms $3^{-l}$ or $2^{-l}$ will be simplified as seen in theorem 3.1 [1]. A hyperrectangle $T \in S$ will have $n - 1$ sides of length $3^{-l}/2$ and one side of length $3^{-l}3/2$, i.e.,

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

And if we use $2^{-l}$ as a size we will have

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

**Proof.** The first affirmation is immediate since, for $n \geq 1$

$$\frac{\varepsilon |f(c)|}{2\sqrt{nK}}(\sqrt{n+8} - \sqrt{n}) \leq \frac{\varepsilon |f(c)|}{2K}(\sqrt{n+8} - \sqrt{n}) \quad (6)$$

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 $2^{-l}\sqrt{3}$. For a hypercube we have $2^{-l}\sqrt{n}$. We adopt the same proof as in theorem 3.1 of [1]

If all hyperrectangles in Figure 6 satisfy condition (3), we can use hyperrectangles having their points on the diagonal, since the size is the same in (a) or (b).

**Corollary 1** Let $f$ and $R$ be as in theorem 1. Suppose that conditions (i) and (ii) of theorem 1 holds. If

$$\frac{\varepsilon |f(c)|}{2\sqrt{nK}}(\sqrt{n+8} - \sqrt{n}) \leq \frac{\varepsilon |f(c)|}{2K}(\sqrt{n+8} - \sqrt{n}) \quad (7)$$

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

By using the points on the diagonals for a potentially optimal hyperrectangle, we get a correction on the lower convex hull as seen in Figure 7.

**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) of definition 1. For hypercube $R$ to be potentially optimal there must exist $\tilde{K}$ such that (1) and (2) hold. From condition (1) we get

$$\tilde{K} \geq \frac{f(c) - f(c_T)}{\alpha_R - \alpha_T}$$
we must choose
\[
\tilde{K} = \max_{T \in S} \frac{f(c) - f(c_T)}{\alpha_R - \alpha_T} = \frac{f(c) - f(c_T)}{\alpha_R - \alpha_{\tilde{T}}}
\]

The left hand side of inequality (7) is equivalent to
\[
\varepsilon \frac{|f(c)|}{2^{-l}\sqrt{n}} \leq \frac{2\sqrt{n}K}{\sqrt{n + 8} - \sqrt{n}}
\]

and
\[
\alpha_R - \alpha_{\tilde{T}} \leq -\left(\frac{2^{-l}}{2}\right)\left(\sqrt{n + 8} - \sqrt{n}\right)
\]

The Lipschitz continuity of \( f \) is equivalent to \( f(c) - f(c_T) \geq -K \) \( 2^{-l}\sqrt{n} \)

Thus
\[
\tilde{K} = \frac{f(c) - f(c_T)}{\alpha_R - \alpha_{\tilde{T}}} \geq \frac{-K2^{-l}\sqrt{n}}{-\left(\frac{2^{-l}}{2}\right)(\sqrt{n + 8} - \sqrt{n})} = \frac{2\sqrt{n}K}{\sqrt{n + 8} - \sqrt{n}}
\]

From the right hand side of (7) and inequality (8) we get
\[
\tilde{K} \geq \frac{\varepsilon |f(c)|}{2^{-l}\sqrt{n}}
\]

i.e., \( f(c) - \tilde{K}\alpha_R \leq f_{\text{min}} - \varepsilon |f_{\text{min}}| \).

**Remark 3** The right hand side of (7) is much smaller than the one in inequation (8), i.e., the smallest hyperrectangle begins being ignored when inequality (7) is satisfied.

As in theorem 3.2 of [1], we can obtain a similar result by choosing \( \tilde{K} \) as small as possible, i.e.,
\[
f(c_T) = \max_{T \in S, \alpha_T > \alpha_R} f(c_T)
\]

**5 Conclusion**

In this paper we have presented a new interpretation of the DIRECT algorithm. The importance of this modification is related to the division procedure. First, it uses less evaluations points in the global search of the algorithm. The hyperrectangles created can have many faces depending on the dimension. DIRECT uses a parameter which influences the smallest hyperrectangle containing \( f_{\text{min}} \) (the current best function value) for not being potentially optimal, and thus may influences the convergence of the algorithm. This hyperrectangle,
which is always a hypercube, depends on the hyperrectangles whose centers are on the neighborhood of this hypercube, and thus on the value of $f$ on these centers. We can discard the potentially optimal hyperrectangle, which influences the slope and then preventing the smallest hypercube for being optimal, by considering the points on the diagonals, since for each potentially optimal hyperrectangle the longest distance from the evaluation points, which are located to the quarters, to its vertices is the same. We seek for 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 which possibly will be potentially optimal.

References


Received: August 6, 2007
Figure 1: The geometric interpretation of conditions (2) and (3)
Figure 2: Example of a division of a square
Figure 3: Division of an interval. We can choose either (a) or (b).
Figure 4: Size of a hyperrectangle. The distance from the sampled point to the vertex can be seen as the diameter of the small circumscribed rectangle.
Figure 5: Points represented on the stencil (a), and in (b) on the diagonal
Figure 6: A correction on the lower convex hull. Hyperrectangle which satisfy condition 3 of theorem 1 is discarded and replaced by another hyperrectangle for which the value of $f$ is much greater.
New interpretation of the DIRECT algorithm

Figure 7: Interpretation of definition 1