CONVERGENCE ANALYSIS OF THE DIRECT ALGORITHM *

D. E. FINKEL† AND C. T. KELLEY *

Abstract. The DIRECT algorithm is a deterministic sampling method for bound constrained Lipschitz continuous optimization. We prove a subsequential convergence result for the DIRECT algorithm that quantifies some of the convergence observations in the literature. Our results apply to several variations on the original method, including one that will handle general constraints. We use techniques from nonsmooth analysis, and our framework is based on recent results for the MADS sampling algorithms.

Key words. DIRECT, Sampling Methods, Clarke derivative, Lipschitz Functions

AMS subject classifications. 65K05, 65K10

1. Introduction.

The DIRECT (DIviding RECTangles) algorithm [19] is a deterministic sampling method. The method is designed for bound constrained non-smooth problems in a small number of variables. Typical applications are engineering design problems, in which complicated simulators are used to construct the objective function [3–8, 16].

By a sampling method we mean that the progress of the optimization is governed only by evaluations of the objective function. No derivative information is needed. Unlike most classical direct search methods [21, 25] such as the Hooke-Jeeves [17], Nelder-Mead [22], implicit filtering [15, 20], or the variations [1, 2, 10] of the multidirectional search algorithm, DIRECT does not sample on a stencil about the current best point at any stage of the optimization, and is designed to perform a global search as the optimization progresses. `DIRECT` is a naturally parallel algorithm, in that a large number of calls to the objective function can be made simultaneously, and distributed to multiple processors.

In the limit, DIRECT will sample a dense grid of points in the feasible set. The behavior of the iteration in the early to middle phases is more important, however. The typical observations are that the sampling clusters near local optima [5–8, 12]. Figure 1.1 is an example of such clustering. The figure is a plot of the 521 sample points from the first first 28 iterations of DIRECT as as applied to the Branin function, a test problem used in [19].

The purpose of this paper is to analyze the behavior of DIRECT by showing that certain subsequences of the sample points converge to points that satisfy the necessary conditions for optimality in the sense of Clarke [9]. We do this using the framework from [1], in which the authors design a sampling algorithm with a very rich set of search directions. The richness of that set allows them to extend the results of [2] to show that all cluster points of the iterations for the new method satisfy certain nonsmooth necessary conditions for optimality. In this paper we observe that DIRECT, which is not based on search directions at all, can be analyzed with these techniques.

*Version of July 14, 2004.

† North Carolina State University, Center for Research in Scientific Computation and Department of Mathematics, Box 8205, Raleigh, N. C. 27695-8205, USA (definkel@ncsu.edu, Tim_Kelley@ncsu.edu). This research was supported in part by National Science Foundation grants DMS-0070641, DMS-0112542, and DMS-0209695.

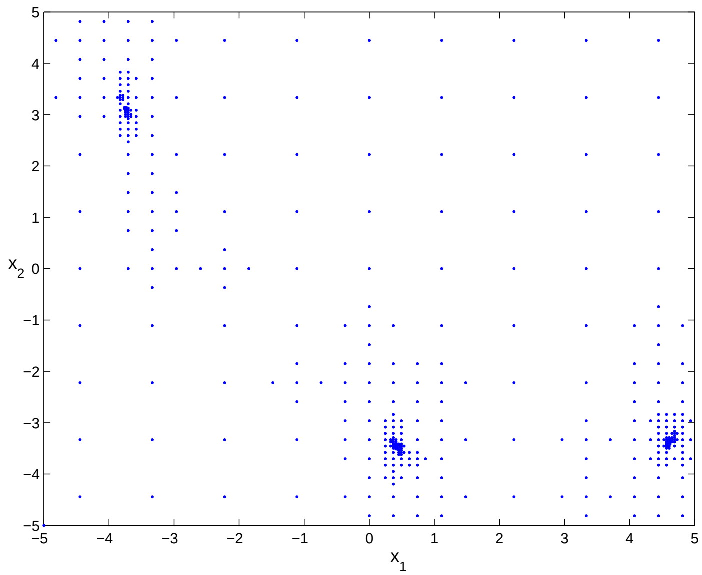

FIG. 1.1. *Branin function example*

We consider a bound-constrained optimization problem,

$$\min_{x \in \Omega} f(x), \qquad \text{where } f : \mathbb{R}^N \to \mathbb{R}, \tag{1.1}$$

where

$$\Omega = \left\{ x \in \mathbb{R}^N : l \leq x \leq u \right\}.$$

and f is Lipschitz continuous on Ω.

DIRECT begins by scaling the domain, Ω, to the unit hypercube. This transformation has no impact on the results, simplifies the description and analysis, and allows the implementation to pre-compute and store common values used repeatedly in calculations, thereby reducing the run-time of the algorithm. Thus, for the remainder of this paper, we will assume that

$$\Omega = \left\{ x \in \mathbb{R}^N : 0 \leq x_i \leq 1 \right\}.$$

DIRECT's sample points are centers of hyperrectangles. In each iteration, new hyperrectangles are formed by dividing old ones, and then the function is sampled at the centers of the new hyperrectangles. We refer to a divide-sample pair as an iteration. In the first, or division phase of an iteration, DIRECT identifies hyperrectangles that show the most potential to contain good, unsampled points. The second or sampling phase is to sample f at the centers of the newly created hyperrectangles. DIRECT typically terminates when a user-supplied budget of function evaluations is exhausted.

In this next section, we outline the original DIRECT algorithm, and discuss several modifications that have been done. Our results apply to very general constraints, if one takes the approach in [5, 6] of assigning an artificial value to an infeasible point.

In § 2 we give a more detailed description of the version of DIRECT from [19] and some of the variations of the algorithm that have appeared since [19] was published. In § 3 we review some basic ideas from nonsmooth analysis and state and prove our convergence results for the case of simple bound constraints. In § 3.3, we use more general concepts from [1, 2, 9] to prove the result for general constraints.

2. Description of DIRECT. DIRECT initiates its search by sampling the objective function at the center of Ω. The entire domain is treated as the first hyperrectangle, which DIRECT identifies as potentially optimal and divides.

`DIRECT` divides a potentially optimal hyperrectangle by trisecting the longest coordinate directions of the hyperrectangle. For example, in the first iteration when the hypercube Ω is potentially optimal, all directions are long, and `DIRECT` divides in every direction. Figure 2.1 illustrates this process.

The order in which sides are divided is important; the right-hand side of Figure 2.1 would look different had the vertical side of the rectangle been divided first. `DIRECT` employs a simple heuristic to determine the order in which long sides are divided. This process is explained in Table 2.1.

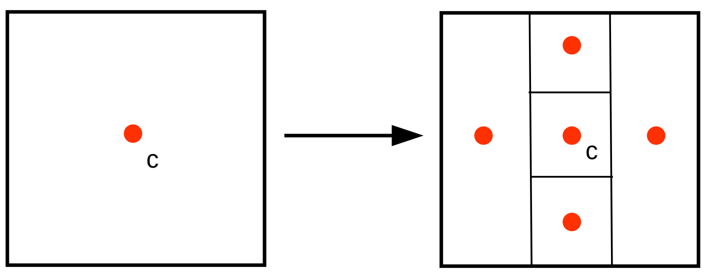

Fig. 2.1. *Two dimensional example of division by DIRECT*

Table 2.1
Division of a hyperrectangle h with center c

1:	Let h be a potentially optimal hyperrectangle with center c.
1:	Let ξ be the maximal side length of h.
2:	Let I be the set of coordinate directions corresponding to sides of h with length ξ.
3:	Evaluate the objective function at the points $c \pm \frac{1}{3}\xi e_i$, for all $i \in I$, where e_i is the ith unit vector
4:	Let $w_i = \min\left\{f(c \pm \frac{1}{3}\xi e_i)\right\}$
5:	Divide the hyperrectangle containing c into thirds along the dimensions in I, starting with the dimension with smallest w_i and continuing to the dimension with the largest w_i.

After the first iteration, the design space, Ω, has been divided into $2N + 1$ hyperrectangles. We let $\mathcal{H}$ be the set of hyperrectangles created by DIRECT. In the next iteration, and those that follow, DIRECT chooses potentially optimal hyperrectangles from $\mathcal{H}$. Potentially optimal hyperrectangles either have low function values at the centers or are large enough to be good targets for global search. To quantify this, DIRECT collects the hyperrectangles into groups of the same size, and considers subdividing hyperrectangles in the subgroups with the smallest value of f at the center. Not all such hyperrectangles are divided; an estimate of the Lipshitz constant of f is used to complete the selection. The formal definition from [19] is

DEFINITION 2.1. *Let $\mathcal{H}$ be the set of hyperrectangles created by DIRECT after k iterations, and let $f_{\min}$ be the best value of the objective function found so far. A hyperrectangle $R \in \mathcal{H}$ with center c_R and size α_R is said to be potentially optimal if there exists $\hat{K}$ such that*

$$f(c_R) - \hat{K}\alpha_R \le f(c_T) - \hat{K}\alpha_T, \quad \text{for all } T \in \mathcal{H} \tag{2.1}$$

$$f(c_R) - \hat{K}\alpha_R \le f_{\min} - \epsilon|f_{\min}|. \tag{2.2}$$

In (2.2), $\epsilon \ge 0$ is a "balance parameter" which provides the user control of the balance between local and global search.

Figure 2.2 illustrates the definition. Each point on the graph represents a subgroup of hyperrectangles having equal sizes (horizontal axis) and equal function values at the centers (vertical axis). The hyperrectangles represented by points on the lower right convex hull of this graph satisfy Equations 2.1 and 2.2, and thus are potentially optimal. Note the role of the balance parameter.

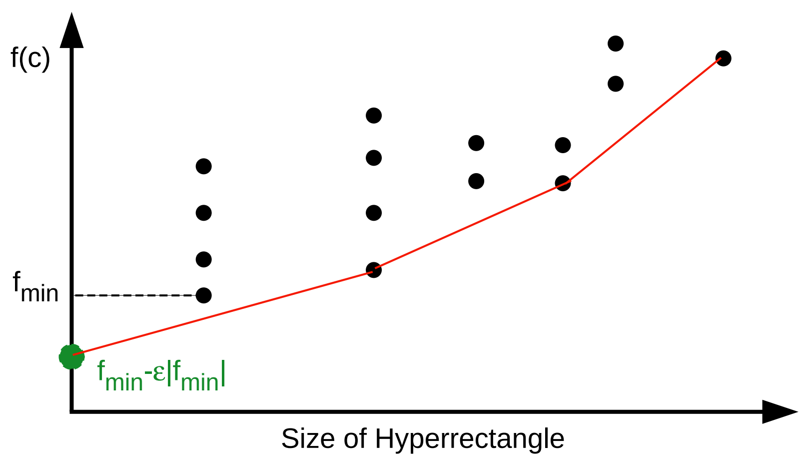

FIG. 2.2. *Hyperrectangles on the piecewise linear curve are potentially optimal*

In choosing the hyperrectangles from the lower right convex hull of Figure 2.2, the local and global search characteristics of DIRECT are illustrated. Hyperrectangles with low objective function values are inclined to fall on the convex hull of the set, as are (relatively) large hyperrectangles. One of the largest hyperrectangles will always be chosen for division [12, 19].

The parameter ϵ was introduced in [19] to balance the local and global search. In Figure 2.2, we see the effects of using the balance parameter. The point $(0, f_{\min} - \epsilon|f_{\min}|)$ alters the convex hull so that the hyperrectangle with the smallest objective function value need not be potentially optimal. In this way more sampling is done in larger, unexplored, hyperrectangles. The value of ϵ is not relevant to the results this paper.

In the original implementation of DIRECT, the size of a hyperrectangle, α_R, is defined to be the distance from the center to a corner. In [13], a modified version of DIRECT is derived which measures hyperrectangles by their longest side. In this way, the groups of hyperrectangles are larger, and consequently the iteration places a stronger emphasis on the value of the objective function at the center of a hyperrectangle. This tends to bias the search locally, as reported in [13].

In [16], the idea of potentially optimal hyperrectangles is discarded, and the implementation described divides hyperrectangles with the lowest function value from each size grouping. This is an aggressive version of DIRECT that was designed to exploit massive parallelism.

2.1. `DIRECT`'s Grid. DIRECT samples from a dense set of points in Ω [12]

$$\mathcal{S} = \left\{ x \mid x = \sum_{i=1}^{N} \frac{2n_i + 1}{2 \cdot 3^{k_i}} e_i \right\}$$

where e_i is the ith coordinate vector, $k_i \geq 0$, and $0 \leq n_i \leq 2 \cdot 3^{k_i} - 1$.

We let $\mathcal{S}_k$ be the set of points that have been sampled after k iterations, and let $\mathcal{B}_k$, the set of best points be

$$\mathcal{B}_k = \{ x \in \mathcal{S}_k \mid f(x) \leq f(z) \text{ for all } z \in \mathcal{S}_k \ \}.$$

Since at least one of the largest hyperrectangles will be potentially optimal at each iteration [19], any given in point $\mathcal{S}$ will be sampled after finitely many iterations, *i. e.*

$$\mathcal{S} = \cup_k \mathcal{S}_k.$$

Hence, DIRECT is, in the limit, an exhaustive search and will, if f is continuous, find an arbitrarily accurate approximation to the global minimum [24]. DIRECT has more structure than this, however.

Let $\mathcal{B}$ be the set of best points

$$\mathcal{B} = \cup_k \mathcal{B}_k.$$

The objective of this paper is to study the cluster points of $\mathcal{B}$.

2.2. General Constraints. Bound constraints are part of DIRECT's sampling strategy, and are incorporated automatically into the optimization. DIRECT can address more general constraints by assigning an artificial value to the objective function at an infeasible point. An obvious way to do this is to assign a large value to such a point. However, such an approach will bias the search away from the boundary of the feasible region. If the solution is on or near that boundary, this approach will significantly degrade the performance of the algorithm.

The results in [5] use a different method, which is also incorporated into the code [14]. Here the artificial value assigned to an infeasible point changes as the optimization progresses. The process is to expand the hyperrectangle containing the infeasible point by doubling the lengths of the sides. If none of the samples in this lager hyperrectangle are feasible, then we flag the point as infeasible, and assign a (relatively) large value to it. Otherwise, we assign an artificial value of $f_* + \delta|f_*|$, where f_* is the minimum value of f over the feasible points in the larger hyperrectangle. In [14], $\delta = 10^{-6}$. This strategy, while non-trivial to implement [14], does not bias the sampling away from the boundary of the feasible region.

DIRECT makes no distinction between inequality constraints that are given directly as formulae and "hidden" constraints, which can only be detected when f fails to return a value.

We let $\mathcal{D}$ denote the feasible set. We will assume very little about the specific way in which artificial values are assigned to infeasible points. We will assume only that if $z \notin \mathcal{D}$, then we assign a function value to z that is larger than f_{min}, where f_{min} is the current best value of f, ie

$$\mathcal{B}_k \subset \mathcal{D}, \text{ for all } k. \tag{2.3}$$

3. Convergence Results: Simple Bound Constraints. In this section we state and prove a convergence result for the case of simple bound constraints. To do this we require a small subset of the results from [9] that were used in [1, 2]. The fully general results will require more of that machinery.

3.1. Results from Nonsmooth Analysis. In this section, we review the tools from nonsmooth analysis [9] that we will need to state and prove the result for simple bound constraints. Throughout we will assume that f is a Lipschitz continuous real-valued function on $X \subset \mathbb{R}^N$. In the context of this paper, $X = \Omega$ if there are no constraints other than simple bounds, and $X = \mathcal{D}$ if there are more general constraints.

Following [9, 18], we define the *generalized directional derivative* of f at $x \in X$ in the direction v as

$$f^o(x; d) = \limsup_{\substack{y \to x,\ y \in X \\ t \downarrow 0,\ y+tv \in X}} \frac{f(y + td) - f(y)}{t}. \tag{3.1}$$

We seek to show that if x^* is a cluster point of $\mathcal{B}$, then the necessary conditions for optimality hold, *i. e.*

$$f^o(x^*; v) \geq 0 \tag{3.2}$$

for all $v \in T_\Omega^{Cl}(x^*)$, the Clarke cone of directions pointing from x^* into Ω.

The cone of directions is easy to describe if there are only simple bound constraints. If $x \in \Omega$, the Clarke tangent cone at x is

$$T_\Omega^{Cl}(x) = \{v \in R^N | x + tv \in \Omega \text{ for all } t > 0 \text{ sufficiently small}\}.$$

3.2. Statement and Proof of the Convergence Theorem. The formal statement of our convergence result is

THEOREM 3.1. *Let f be Lipschitz continuous on Ω and let x^* be any cluster point of $\mathcal{B}$. Then* (3.2) *holds.*

Proof. We will show that (3.2) holds with an indirect proof. Assume that $f^o(x^*; v) < 0$ for some $v \in T_\Omega^{Cl}(x^*)$. We will exhibit K and $\Delta > 0$ such that

$$\inf_{k \geq K} \operatorname{dist}(x^*, \mathcal{B}_k) \geq \Delta, \tag{3.3}$$

contradicting the assumption that x^* is a limit point of $\mathcal{B}$.

Since $f^o(x^*;v) < 0$ and $v \in T_\Omega^{Cl}(x^*)$, there is $\delta > 0$ such that

$$y^* = x^* + \delta v \in \Omega,$$

and $f(y^*) < f(x^*)$.

Let L_f denote the Lipschitz constant of f. Let

$$\Delta = \min\left\{\delta/2, \frac{f(x^*) - f(x*+\delta v)}{2L_f}\right\} \tag{3.4}$$

and let

$$\mathcal{N} = \{x \,|\, \|x - x^*\| \le \Delta\} \cap \Omega. \tag{3.5}$$

For all $x \in \mathcal{N}$,

$$f(x) - f(y^*) \ge f(x^*) - L_f\Delta - f(y^*) \ge \frac{f(x^*) - f(y^*)}{2} > 0. \tag{3.6}$$

Since $\mathcal{S}$ is dense in Ω, there is $K > 0$ and $\hat{x} \in \mathcal{S}_K$ such that

$$\|\hat{x} - y^*\| \le \Delta/2,$$

and hence, for all $x \in \mathcal{N}$,

$$f(\hat{x}) \le f(y^*) + L_F\Delta/2 \le f(y^*) + \frac{f(x^*) - f(y^*)}{4} < f(x). \tag{3.7}$$

Hence,

$$\mathcal{N} \cap \mathcal{B}_k = \emptyset$$

for all $k \ge K$, as asserted. □

3.3. Convergence Results: General Constraints. In this section, we extend our nonsmooth results to a more general design space, $\mathcal{D} \subset \Omega$. Again, `DIRECT` attempts to minimize f over Ω, and assigns artificial values to any $x \notin \mathcal{D}$ (that is, $x \in \mathcal{D}^C$).

As we said in § 2.2, the only requirement on the artificial assignment heuristic is (2.3). Examples of such assignment strategies include the already discussed approach of [14], the barrier approach used in [1], a traditional penalty function approach [11], and others.

The strength of the results in this section are dependent upon assumptions about properties of $\mathcal{D}$ at the cluster points of $\mathcal{B}$, a fact first pointed out in [23]. Our analysis follows the program from [1].

We first define a hypertangent cone.

DEFINITION 3.2. *A vector $v \in \mathbb{R}^N$ is said to be a hypertangent vector to the set $\mathcal{D} \subset \mathbb{R}^N$ at the point $x \in \mathcal{D}$ if there exists a scalar $\epsilon > 0$ such that*

$$y + tw \in \mathcal{D} \quad \text{for all } y \in \mathcal{D} \cap B_\epsilon(x), \quad w \in B_e(v), \ \text{ and } \ 0 < t < \epsilon. \tag{3.8}$$

The set of hypertangent vectors to $\mathcal{D}$ at x is called the hypertangent cone to $\mathcal{D}$ at x and is denoted by $T_\mathcal{D}^H(x)$.

If x^* is a cluster point of $\mathcal{B}$, then (3.2) holds for all $v \in T^H_{\mathcal{D}}(x^*)$.

THEOREM 3.3. *Let f be Lipschitz continuous on $\mathcal{D}$ and let x^* be any cluster point of $\mathcal{B}$. Then $f^o(x^*; v) \geq 0$ for all $v \in T^H_{\mathcal{D}}(x^*)$.*

Proof. The proof is a simple extension of the proof of Theorem 3.1. Assume that $f^o(x^*; v) < 0$ for some $v \in T^H_{\mathcal{D}}(x^*)$. Then, there exists $0 < \delta < \epsilon$ such that

$$y^* = x^* + \delta v \in \mathcal{D},$$

and $f(y^*) < f(x^*)$.

Define Δ and $\mathcal{N}$ by (3.4) and (3.5), and let $\mathcal{D}^C$ denote the complement of $\mathcal{D}$. We have

$$\left(\mathcal{N} \cap \mathcal{D}^C\right) \cap \mathcal{B}_k = \emptyset, \tag{3.9}$$

because $\mathcal{B}_k \subset \mathcal{D}$.

Since $\mathcal{S}$ is dense in Ω, it follows that there exists K and $\hat{x} = x^* + tw \in \mathcal{S}_K$, for some $0 < t \leq \delta$, and $w \in B_\epsilon(v)$. Since $v \in T^H_{\mathcal{D}}(x^*)$ it follows that $\hat{x} \in \mathcal{D}$. Furthermore, we choose $\hat{x}$ so that $\|w - v\|$ and $|t - \delta|$ are small enough to ensure that

$$\|\hat{x} - y^*\| = \|tw - \delta v\| \leq \Delta/2.$$

As already shown in (3.6) and (3.7), $f(\hat{x}) < f(x)$ for all $x \in \mathcal{N} \cap \mathcal{D}$, so

$$(\mathcal{N} \cap \mathcal{D}) \cap \mathcal{B}_k = \emptyset, \tag{3.10}$$

and thus (from (3.9) and (3.10)),

$$\mathcal{N} \cap \mathcal{B}_k = \emptyset,$$

for all $k \geq K$, which proves our assertion. □

Theorem 3.3 differs from Theorem 3.1 in that the set of directions for which (3.2) holds are not, in general, the same. In the case of simple bound constraints, $T^H_{\mathcal{D}}(x^*)$ is non-empty, and its closure is $T^{Cl}_{\mathcal{D}}(x)$, so if (3.2) holds for all $v \in T^H_{\mathcal{D}}(x^*)$, it holds by continuity for all $v \in T^{Cl}_{\mathcal{D}}(x)$. This is not so in the general case. To explore the new assumptions necessary to we follow [1], and use the more general definitions of the Clarke and contingent cones [9,18,23] for this purpose.

DEFINITION 3.4. *A vector $v \in \mathbb{R}^N$ is said to be a Clarke tangent vector to the set $\mathcal{D} \subset \mathbb{R}^N$ at the point x in the closure of $\mathcal{D}$ if for every sequence $\{y_k\}$ of elements of $\mathcal{D}$ that converges to x and for every sequence of positive real numbers $\{t_k\}$ converging to zero, there exists a sequence of vectors $\{w_k\}$ converging to v such that $y_k + t_k w_k \in \mathcal{D}$. The set $T^{Cl}_{\mathcal{D}}(x)$ of all Clarke tangent vectors to $\mathcal{D}$ at x is called the Clarke tangent cone to $\mathcal{D}$ at x.*

DEFINITION 3.5. *A vector $v \in \mathbb{R}^N$ is said to be a tangent vector to the set $\mathcal{D} \subset \mathbb{R}^N$ at the point x in the closure of $\mathcal{D}$ if there exists a sequence $\{y_k\}$ of elements of $\mathcal{D}$ that converges to x and a sequence of positive real numbers $\{\lambda_k\}$ for which $v = \lim_k \lambda_k(y_k - x)$. The set $T^{Co}_{\mathcal{D}}(x)$ of all tangent vectors to $\mathcal{D}$ at x is called the contingent cone (or sequential Bouligand tangent cone) to $\mathcal{D}$ at x.*

The three cones are nested [1],

$$T^H_{\mathcal{D}}(x) \subseteq T^{Cl}_{\mathcal{D}}(x) \subseteq T^{Co}_{\mathcal{D}}(x).$$

They also note that, if $T_{\mathcal{D}}^{Cl}(x) = T_{\mathcal{D}}^{Co}(x)$, the set $\mathcal{D}$ is said to be *regular* at x.

Our next two results state the necessary assumptions to show that DIRECT finds Clarke and contingent stationary points.

THEOREM 3.6. *Let x^* be a cluster point of $\mathcal{B}$. If $T_{\mathcal{D}}^{H}(x^*) \neq \emptyset$, then $f^o(x^*; v) \geq 0$ for all $v \in T_{\mathcal{D}}^{Cl}(x^*)$.*

Proof. By Theorem 3.3, $f^o(x^*; w) \geq 0$ for all $w \in T_{\mathcal{D}}^{H}(x^*)$. If $T_{\mathcal{D}}^{H}(x^*)$ is non-empty, then [1],

$$f^o(\hat{x}; v) = \lim_{\substack{w \to v \\ w \in T_{\mathcal{D}}^{H}(\hat{x})}} f^o(\hat{x}; w) \geq 0,$$

as asserted. □

We conclude by extending our results to differentiable functions. We first state a simple observation (also made for MADS methods in [1]) that if the set $\mathcal{D}$ is regular (*i. e.* $T_{\mathcal{D}}^{Cl}(x) = T_{\mathcal{D}}^{Co}(x)$) at x^*, then cluster points of $\mathcal{B}$ are stationary with respect to the contingent cone.

COROLLARY 3.7. *Let x^* be a cluster point of $\mathcal{B}$. If $T_{\mathcal{D}}^{H}(x^*) \neq \emptyset$, and if $\mathcal{D}$ is regular at x^*, then $f^o(x^*; v) \geq 0$ for all $v \in T_{\mathcal{D}}^{Co}(x^*)$.*

Our final result extends Corollary 3.7. If we assume strict differentiability of f, then we can state the constraint qualifications needed to show that cluster points of $\mathcal{B}$ are KKT points.

THEOREM 3.8. *Let f be strictly differentiable, and let x^* be a cluster point of $\mathcal{B}$. If $T_{\mathcal{D}}^{H}(x^*) \neq \emptyset$, and if $\mathcal{D}$ is regular at x^*, then x^* is a contingent KKT stationary point of f over $\mathcal{D}$.*

Proof. As pointed out in [1,9], strict differentiability of f at x^* implies that $\nabla f(x^*)^T v = f^o(x^*; v)$ for all $v \in T_{\mathcal{D}}^{Co}(x^*)$. Thus, it follows from the previous corollary that $-\nabla f(x^*)^T v \leq 0$ for all v in the contingent cone. □

Acknowledgments. The authors are very grateful to John Dennis and Evin Kramer for their thoughtful comments on a preliminary version of this paper.

REFERENCES

[1] C. Audet and J. E. Dennis, *Mesh adaptive direct search algorithms for constrained optimization*, Tech. Rep. TR04-02, Department of Computational and Applied Mathematics, Rice Univeristy, 2004.

[2] C. Audet and J. D. Jr., *Analysis of generalized pattern searches*, SIAM J. Optim., 13 (2003), pp. 889–903.

[3] C. A. Baker, L. T. Watson, B. Grossman, R. T. Haftka, and W. H. Mason, *Parallel global aircraft configuration design space exploration*. preprint, 1999.

[4] C. A. Baker, L. T. Watson, B. Grossman, W. H. Mason, S. E. Cox, and R. T. Haftka, *Study of a global design space exploration method for aerospace vehicles*. preprint, 1999.

[5] R. Carter, J. Gablonsky, A. Patrick, C. Kelley, and O. Eslinger, *Algorithms for nosiy problems in gas transmission pipeline optimization*, Optimization and Engineering, 2 (2001), pp. 139–157.

[6] R. G. Carter, *Compressor station optimization - computational accuracy and speed*, 1996. Proceedings of the Pipeline Simulation Interest Group, Twenty-Eighth Annual Meeting, San Francisco CA, Paper number PSIG-9605.

[7] ——, *Pipeline optimization: dynamic programming after 30 years*, 1998. Proceedings of the Pipeline Simulation Interest Group, Denver Colorado, Paper number PSIG-9803.

[8] R. G. Carter, W. W. Schroeder, and T. D. Harbick, *Some causes and effect of discontinuities in modeling and optimizing gas transmission networks*, 1993. Proceedings of the Pipeline Simulation Interest Group, Pittsburg PA, Paper number PSIG-9308.

[9] F. Clarke, *Optimization and Nonsmooth Analysis*, Canadien Mathematical Society Series of Monographs and Advanced Texts, Wiley-Interscience, New York, 1983.

[10] J. E. Dennis and V. Torczon, *Direct search methods on parallel machines*, SIAM J. Optim., 1 (1991), pp. 448 – 474.

[11] R. Fletcher, *Practical Methods of Optimization*, John Wiley and Sons, New York, second ed., 1987.

[12] J. Gablonsky, *Modifications of the Direct Algorithm*, PhD thesis, North Carolina State University, 2001.

[13] J. Gablonsky and C. Kelley, *A locally-biased form of the direct algorithm*, Journal of Global Optimization, 21 (2001), pp. 27–37.

[14] J. M. Gablonsky, *DIRECT Version 2.0 User Guide*, Tech. Rep. CRSC-TR01-08, Center for Research in Scientific Computation, North Carolina State University, April 2001.

[15] P. Gilmore and C. T. Kelley, *An implicit filtering algorithm for optimization of functions with many local minima*, SIAM J. Optim., 5 (1995), pp. 269–285.

[16] J. He, L. Watson, N. Ramakrishnan, C. Shaffer, A. Verstak, J. Jian, K. Bae, K. Bae, and W. Tranter, *Dynamic data structures for a direct search algorithm*, Computational Optimization and Applications, 23 (2002), pp. 5–25.

[17] R. Hooke and T. Jeeves, *Direct search solution of numerical and statistical problems*, Journal of the Association for Computing Machinery, 8 (1961), pp. 212–229.

[18] J. Jahn, *Introduction to the Theory of Nonlinear Optimization*, Springer, Berlin, 1994.

[19] D. Jones, C. Perttunen, and B. Stuckman, *Lipschitzian optimization without the lipschitz constant*, Journal of Optimization Theory and Application, 79 (1993), pp. 157–181.

[20] C. Kelley, *Iterative Methods for Optimization*, Frontiers in Applied Mathematics, SIAM, Philadelphia, PA, first ed., 1999.

[21] T. Kolda, R. Lewis, and V. Torczon, *Optimization by direct search: New perspective on some classical and modern methods*, SIAM Review, 45 (2003), pp. 385–482.

[22] J. Nelder and R. Mead, *A simplex method for function minimization*, Comput. J., 7 (1965), pp. 308–313.

[23] R. Rockafellar, *Generalized directional derivatives and subgradients of nonconvex functions*, Canadien Journal of Mathematics, 32 (1980), pp. 157–180.

[24] C. P. Stephens and W. Baritompa, *Global optimization requires global information*, J. Optim. Theory Appl., 96 (1998), pp. 575–588.

[25] M. Wright, *Direct search methods: Once scorned, now respectable*, in Proceedings of the 1995 Dundee Bennial Conference in Numerical Analysis, D. Griffiths and G. Watson, eds., CRC Press, Boca Raton, FL, 1996, Pirman Res. Notes Math. Ser. 344, pp. 191–208.