Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 5, No. 2, pp. 339–349, May 2009

# CONVERGENCE ANALYSIS OF SAMPLING METHODS FOR PERTURBED LIPSCHITZ FUNCTIONS*

D.E. Finkel and C.T. Kelley

**Abstract:** In this short note we observe that results of Dennis and Audet extend naturally to a wide variety of deterministic sampling methods. For bound-constrained problems, we show that any method based on coordinate search which includes a sufficiently rich set of directions, for example random directions at each state of the sampling, will, when applied to Lipschitz continuous problems, have cluster points that satisfy generalized necessary conditions for optimality. The results also apply to the case of more general constraints, including so-called "hidden" or "yes-no" constraints.

**Key words:** *sampling methods, Clarke derivative, Lipschitz functions*

**Mathematics Subject Classification:** *65K05, 65K10*

## 1 Introduction

In this short paper we make the observation that the analysis which supports the MADS [2] approach to generating search directions for direct search methods applies equally well to problems with noise in the objective function. This enables us to apply the theory to the paradigm which motivates sampling methods such as implicit filtering [24, 30].

Deterministic sampling methods such as Nelder-Mead [33], Hooke-Jeeves [26], the DFO method [13, 15], and the many variants of the direct search algorithm [2, 27, 31, 36] are most typically applied to problems which are nonsmooth, can be discontinuous, and are often corrupted by high-frequency, low-amplitude noise. By noise we mean something broader than statistical noise, and allow for any high-frequency and low-amplitude variation in the function. This noise can arise from truncation error in simulations which are internal to the objective function [16, 17, 22], stochastic methods used to compute the objective function [8, 9, 24], or termination of internal iterations [22]. The combination of nonsmoothness and noise can trap a conventional, gradient-based optimization algorithm in a local minimum.

Such a function may not even return a value. There are several reasons for this: an iteration within the simulator may not converge; a set of design variables generated by the optimization algorithm may not be acceptable to the simulator (*e. g.* a negative damping coefficient); or the simulator may exceed its own limits on internal iterations, cpu time, or storage. Failure of the function evaluation has been observed in practice, [4, 7, 8, 9, 11, 14, 21, 24]. When such a failure takes place we say that a "hidden constraint" has been

*This research was supported in part by National Science Foundation grants DMS-0070641, DMS-0112542, and DMS-0209695.

violated. We use the term "hidden constraint" because only an attempt to evaluate $f(x)$ can determine if $x$ is a feasible point. Such constraints are also called "virtual constraints" [14] or "yes/no" constraints [6]. In some cases [10, 11, 21], it is even difficult to find a point at which one can evaluate the function and begin an iteration. We discuss hidden constraints in § 2.3.

We will begin in § 1.1 with a description of the kinds of structured sampling methods we consider and how they have coordinate search as a core part of the algorithms. We will then describe our assumptions on the set of directions.

In § 2.2 we start with an analysis of the bound constrained problem

$$\min_{x\in\Omega} f(x) \tag{1.1}$$

where

$$\Omega = \left\{x \in R^N : l \leq x \leq u\right\}, \tag{1.2}$$

and the inequality in (1.2) is component-wise.

In § 2.3 we discuss more general constraints. In particular our analysis allows for so-called "hidden" constraints, for which infeasibility can only be detected when $f$ fails to return a value.

## 1.1 Sampling Methods and Directions

We consider a class of stencil-based sampling methods in this paper.

Sampling methods control the progress of the optimization by evaluating (sampling) the objective function at points in $\Omega$. Sampling methods include the many variants on the direct search paradigm [1, 2, 27, 31], the Hooke-Jeeves [26] method, implicit filtering [25, 30], DIRECT [29], and the Nelder-Mead [33] algorithm. With the exceptions of DIRECT and the Nelder-Mead method, all of the methods mentioned above are stencil-based. This means that some or all of the sampling is done on a stencil of points centered at the current best point and that the method changes the size or scale of the stencil if no new point is better than the current best point.

Sampling methods do not require gradient information, but may, as implicit filtering does, attempt to infer gradient and even Hessian information from the sampling.

## 1.2 Coordinate Search

We will use coordinate search as an example. The algorithm begins with a base point $x$ and a scale $h \leq 1/2$. The algorithm begins by evaluating $f$ evaluated at the $2N$ points on the **stencil**

$$S(x, h) = \{z \,|\, z = x \pm he_i\} \cap \Omega, \tag{1.3}$$

centered at $x$ and restricted to lie within the bounds. In (1.3) $e_i$ is the unit vector in the $i$th coordinate direction. In a non-opportunistic search we sample the entire stencil and replace $x$ with $z \in S(x, h)$, where

$$f(z) = \min_{w\in S(x,h)} f(w)$$

if $f(z) \geq f(x)$, a condition we will refer to as *stencil failure*, holds. When stencil failure happens we shrink the stencil by reducing $h$. The standard way to reduce $h$ is by a factor of 2.

This paper is not about termination of algorithms like coordinate search but rather about asymptotic behavior. However we will remark that sampling methods are unlike gradient-based methods [18, 20, 23, 30, 34] where one can test for satisfaction of necessary conditions for optimality. One common way to control the cost of a sampling method is to give the iteration a budget of calls to $f$, and to terminate the iteration when that budget is exhausted.

In the case of smooth (Lipschitz continuously differentiable) $f$, it is easy to show [5, 30] that if stencil failure happens, then the necessary conditions for optimality [3, 30] hold to first order *i. e.*

$$x - \mathcal{P}(x - \nabla f(x)) = O(h). \tag{1.4}$$

In (1.4) $\mathcal{P}$ is the $l^2$ projection on to $\Omega$,

$$\mathcal{P}(x) = \max(l, \min(x, u)),$$

where the max and min are taken component-wise.

The proof that stencil failure implies (1.4) is simple and we summarize it here. Let $\{x_n\}$ denote the sequence of of coordinate search iterations and $\{h_n\}$ the non-increasing sequence of scales. Let $x \in \{x_n\}$ be an iteration for which stencil failure happens and let $h = h_n$. To avoid confusion with the iteration index, we let $(x)_i$ denote the $i$th component of the vector $x$.

Let $1 \le i \le N$. Suppose first that $x \pm he_i \in \Omega$. In that case Lipschitz continuity and $f(x) \le f(x \pm he_i)$ implies

$$\begin{aligned} 0 \le f(x + he_i) - f(x) &= h\partial f(x)/\partial x_i + O(h^2) \\ 0 \le f(x - he_i) - f(x) &= -h\partial f(x)/\partial x_i + O(h^2) \end{aligned}$$

which implies that $|\partial f(x)/\partial x_i| = O(h)$. Since $h < 1/2$ at most one of $x \pm he_i \notin \Omega$. If, say $x - he_i \notin \Omega$ then we may only make conclusions from $f(x + he_i)$. So $\partial f(x)/\partial x_i \ge O(h)$. Since $x$ is a distance at most $h$ from the $x_i = l$ face of $\Omega$ and $\partial f(x)/\partial x_i \ge 0$, we have

$$x - \max(l, \min((x)_i - \partial f(x)/\partial x_i, u)) = O(h).$$

Hence (1.4) holds.

Convergence of the iteration will follow from (1.4). To see this note that the sampling evaluates points at a grid of size $h$ centered at $x_0$ and aligned with the coordinate directions. Since there are only finitely many grid points in $\Omega$, we must have stencil failure after finitely many evaluations of $f$. At that point (1.4) holds and we must reduce $h$. Hence any limit point of the iteration must satisfy the necessary conditions

$$x - \mathcal{P}(x - \nabla f(x)) = 0, \tag{1.5}$$

for optimality.

One can also prove convergence if the objective function is contaminated with a certain type of noise. If

$$f = f_s + \phi, \tag{1.6}$$

where $f_s$ is smooth and $\phi$ is low-amplitude noise. We measure the size of the noise at $x$ by taking the maximum of $|\phi|$ over the union of the center $x$ and the stencil $S(x, h)$ (1.3). We define

$$\|\phi\|_{S(x,h)} = \max_{z \in \{x\} \cup S(x,h)} |\phi(z)|.$$

If [30]

$$\lim_{n \to \infty} \frac{\|\phi\|_{S(x_n,h_n)}}{h_n} \to 0, \tag{1.7}$$

then every limit point of $\{x_n\}$ satisfies (1.5).

The version of coordinate search we described above can be significantly improved with some very simple changes. One simple way to improve the performance is to do the function evaluations in parallel. Another way is to do an "opportunistic search", where the new point is the first point sampled in the stencil which is better than the current point. Aside from the sampling strategy, one can improve the performance of the method by building surrogate models of $f$ using the history of the optimization or by incorporating more points into the sample [15, 30, 31]. This is the "search-poll" paradigm defined in [19] where the exploratory phase (search) attempts to find a better point independently of the complete sampling of the stencil (poll) is applied to determine if the scale should be reduced. Any method that reduces the scale after stencil failure can be analyzed with the convergence theory for coordinate search. Examples include implicit filtering [24, 30] and the DFO method [13, 15] which sample the full stencil to build a model of $f$ and methods like the Hooke-Jeeves algorithm [26], APPS [27], and MADS which are opportunistic searches and only sample the full stencil after the search phase has failed.

Many sampling methods [2, 26, 31] restrict the search to a fixed grid for each value of $h$, and the analysis for coordinate search applies equally well to such methods provided the search directions form a positive spanning set. Methods such as implicit filtering [24, 30], for example, sample on a positive spanning set but may find a new point in the search phase that does not lie on the grid of size $h$ which includes the current point. For such algorithms one must assume that stencil failure occurs infinitely often and then [30] the convergence results will apply to any limit point of the sequence of points on which stencil failure occurs.

### 1.2.1 The need for more directions

The important component of the proof of (1.4) is the fact that if $f$ is smooth, then the size of $\nabla f$ can be determined by the directional derivatives in the coordinate directions or, more generally, the directional derivatives from any set of directions which positively span $R^N$ [15, 31, 37]. However, if $f$ is Lipschitz continuous but not smooth, no finite set of search directions will enable us to conclude (1.4) from stencil failure [2]. Therefore we must enrich the stencil to obtain convergence.

More general constraints can also raise a need for additional directions. Even linear constraints can confound a sampling method if the tangent directions for the constraints are not in the direction set [31, 32]. In the case of linear constraints one can use the constraint matrix to add directions as needed [32]. This approach becomes problematic for general nonlinear constraints and impossible for the so-called "hidden constraints", for which infeasibility is only determined when $f$ fails to return a value.

The results in [2] show that a grid-based method can be designed which has a sufficiently rich set of directions to resolve many of these problems for Lipschitz continuous $f$ and sufficiently regular feasible sets. In this paper we extend those results to the case where $f$ is contaminated with noise.

## 2 Convergence Results

In this section we present two convergence results. In § 2.2 we consider only bound constraints and assume that $f$ is defined for all points in $\Omega$. In § 2.3 we assume that $f$ can fail to return a value. This failure can be a result of an internal failure in the call to $f$ or a failure to satisfy a nonlinear constraint. In the latter case one would simply assign $f$ an artificial value (which could include $\infty$ or NaN) as a way to flag the failure without calling

$f$ itself. The two results differ only in that the notion of stationarity differs between them, not in the details of the analysis. Hence § 2.3 is brief and focuses only on the appropriate generalization of stationarity. In the constrained case we will denote the feasible set by $\mathcal{D} \subset \Omega$.

## 2.1 Iterates, Directions, and Necessary Conditions

As in [2] we will need a richer set of directions than a stencil which does not change as the iteration progresses. We let $\mathcal{I} = \{\xi_n\}$ be the sequence of iterations and $\mathcal{V} = \{V_n\}$ be the sequence of search directions at iteration $n$. By this we mean that

$$V_n = \{v_{n_k}\}_{k=1}^{N_n}$$

with $v_{n_k}$ a sequence of unit vectors. In this case the stencil at iteration $n$ would be

$$S(\xi_n, h_n) = \{z \,|\, z = \xi_n + h_n v_{n_k},\ 1 \le k \le N_n\} \cap \Omega. \tag{2.1}$$

In the standard case where $V_n = V$ is independent of $n$, we would require that $V$ contain a positive spanning set. In this paper, however, we assume that the sequence $\{V_n\}$ is rich in directions, but not necessarily that any single $V_n$ is. In fact $V_n$ could have only one element for each $n$ and the theory would hold.

**Assumption 2.1.** Let $\mathcal{W} = \{W_n\}$ be any infinite subsequence of $\mathcal{V}$ and let $v$ be any unit vector in $R^N$. Then

$$\liminf_{n\to\infty} \min_{1\le k\le N_n} \|w_{n_k} - v\| = 0. \tag{2.2}$$

Examples of sets satisfying Assumption 2.1 (at least with probability one) include the cases where $V_n$ is a random rotation of the centered difference stencil, a positive spanning set which is augmented with one or more random vectors, or the MADS basis from [2]. The important property of $\mathcal{V}$, as was the case in [2], is the assumption that any direction can be approximated arbitrarily by vectors taken from any subsequence of $\mathcal{V}$. We make this precise in Assumption 2.1

We will relax the smoothness assumptions in (1.6) by assuming that

$$f = f_l + \phi \tag{2.3}$$

where $f_l$ is Lipschitz continuous.

Now let $\mathcal{I}^{sf} = \{\xi_n^{sf}\}$ be the subset of $\mathcal{I}$ of iterations for which stencil failure holds. We consider limit points of $\mathcal{S}$ at which $\phi \to 0$. We make this precise.

**Definition 2.2.** Let $\mathcal{S} = \{\xi_n\}$ be a convergent subsequence of $\mathcal{I}^{sf}$ with limit $x^*$, corresponding directions $\{W_n\}$, and scales $\{h_n\}$. We say that $x^*$ is a *smooth limit point* of $\mathcal{I}^{sf}$ if

$$\lim_{n\to\infty} \frac{\|\phi\|_{S(\xi_n,h_n)}}{h_n} = 0. \tag{2.4}$$

$\mathcal{I}^{sf}$ may be empty for methods which are not bound to a grid. Therefore we will assume that $\mathcal{I}^{sf}$ is nonempty for the remainder of this paper.

We now review the tools from nonsmooth analysis [12] that we will need to state and prove the result for simple bound constraints. We will assume that $f_l$ is a Lipschitz continuous real-valued function on $X \subset R^N$. In the context of this paper, $X = \Omega$ if there are no constraints other than simple bounds and $X \subset \Omega$ is the feasible set $\mathcal{D}$ otherwise.

Following [12, 28], we define the *generalized directional derivative* of a Lipschitz continuous function $g$ at $x \in X$ in the direction $v$ as

$$g^\circ(x;v) = \limsup_{\substack{y \to x,\ y \in X \\ t \downarrow 0,\ y+tv \in X}} \frac{g(y+tv) - g(y)}{t}. \tag{2.5}$$

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

$$f_l^o(x^*;v) \geq 0 \tag{2.6}$$

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

## 2.2 Bound Constraints

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

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

is the same set as the Hypertangent cone we will define in § 2.3. In this case $T_\Omega^{Cl}(x)$ is the closure of its interior for all $x \in \Omega$ (which is not the case for the more general situation we discuss in § 2.3).

The convergence result is

**Theorem 2.3.** *Assume that $f_l$ is Lipschitz continuous on $\Omega$ and that Assumption 2.1 holds. Let $\mathcal{S} = \{\xi_n\}$ be a convergent subsequence of $\mathcal{I}^{sf}$ with limit $x^*$, corresponding directions $\{W_n\}$, and scales $\{h_n\}$. Assume that $x^*$ is a smooth limit point of $\mathcal{I}^{sf}$. Then* (2.6) *holds.*

*Proof.* By Assumption 2.1 and taking subsequences as needed there are directions $w_n \in W_n$ such that

$$w_n \to u.$$

The definition of $f^o$ and the convergence of $\xi_n$ to $x^*$ then imply that

$$f_l^o(x^*;u) \geq \lim_{n\to\infty} \frac{f_l(\xi_n + h_n u) - f_l(\xi_n)}{h_n}. \tag{2.7}$$

Since

$$f_l(\xi_n + h_n u) \leq f_l(\xi_n + h_n w_n) + \lambda h_n \|u - w_n\|,$$

where $\lambda$ is the Lipschitz constant of $f_l$, we obtain

$$\begin{aligned} f_l(\xi_n + h_n u) - f_l(\xi_n) &\geq f_l(\xi_n + h_n w_n) - f_l(\xi_n) - \lambda h_n \|w_n - u\| \\ &= f_l(\xi_n + h_n w_n) - f_l(\xi_n) + o(h_n). \end{aligned} \tag{2.8}$$

Since $x^*$ is a smooth limit point, (2.8) implies that

$$\begin{aligned} f_l(\xi_n + h_n w_n) - f_l(\xi_n) &\geq f(\xi_n + h_n w_n) - f(\xi_n) - 2\|\phi\|_{S(\xi_n,h_n)} \\ &= f(\xi_n + h_n w_n) - f(\xi_n) + o(h_n). \end{aligned} \tag{2.9}$$

Since stencil failure occurs at each $\xi_n$, we can combine (2.8) and (2.9) to obtain

$$\frac{f_l(\xi_n + h_n u) - f_l(\xi_n)}{h_n} \geq o(1), \tag{2.10}$$

and hence $f^o(x^*;u) \geq 0$. □

## 2.3 General Constraints

In this section we assume that $f$ does not return a value outside of a set $\mathcal{D} \subset \Omega$. Sampling methods often handle constraints, even simple bounds and linear constraints, by simply returning $\infty$, $NaN$, or an artificial value [2, 24, 30, 31]. The results in this section do not depend on the how infeasible points are processed, only that such points are flagged in a way that eliminates them as candidates for the new point and declares stencil failure if the current point is the best of all feasible points in the current stencil.

The geometry of $\mathcal{D}$ determines the set of admissible directions from $x^*$ and limits the strength of the necessary conditions we will be able to prove. We do not give proofs of the results in this section because the follow from Theorem 2.3 and the logic from [2].

We begin with the direct extension of Theorem 2.3. We define the hypertangent cone.

**Definition 2.4.** A vector $v \in R^N$ is said to be a hypertangent vector to the set $\mathcal{D} \subset R^N$ at the point $x \in \mathcal{D}$ if there exists a scalar $\epsilon > 0$ such that

$$y + tw \in \mathcal{D} \quad \text{for all } y \in \mathcal{D} \cap B_\epsilon(x), \quad w \in B_\epsilon(v), \text{ and } 0 < t < \epsilon, \tag{2.11}$$

where $B_\epsilon(x)$ is the ball of radius $\epsilon$ centered at $x$. The set of hypertangent vectors to $\mathcal{D}$ at $x$ is called the hypertangent cone to $\mathcal{D}$ at $x$ and is denoted by $T^H_{\mathcal{D}}(x)$.

If $x^*$ is a smooth limit point of $\mathcal{I}^{sf}$ we can show that $f^o(x^*; v) \geq 0$ for all $v \in T^H_{\mathcal{D}}(x^*)$ using the same arguments as in the proof of Theorem 2.3.

**Theorem 2.5.** *Assume that $f_l$ is Lipschitz continuous on $\mathcal{D}$ and that Assumption 2.1 holds. Let $\mathcal{S} = \{\xi_n\}$ be a convergent subsequence of $\mathcal{I}^{sf}$ with limit $x^*$, corresponding directions $\{W_n\}$, and scales $\{h_n\}$. Assume that that $x^*$ is a smooth limit point of $\mathcal{I}^{sf}$. Then* (2.6) *holds for all $v \in T^H_{\mathcal{D}}(x^*)$.*

For general constraints it is not the case that the closure of $T^H_{\mathcal{D}}(x)$ is $T^{Cl}_{\mathcal{D}}(x)$, and hence the conclusion of Theorem 2.5 is weaker that the full Clarke stationarity conditions. 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 (2.6) 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 we follow [2], and use the more general definitions of the Clarke and contingent cones [28, 12, 35] for this purpose.

We observe that if $x^*$ is in the interior of $\mathcal{D}$, then $T^H_{\mathcal{D}}(x^*) = R^N$ and the results of § 2.2 can be applied. So the differences arise only when $x^*$ is on the boundary of $\mathcal{D}$.

**Definition 2.6.** A vector $v \in R^N$ is said to be a Clarke tangent vector to the set $\mathcal{D} \subset R^N$ 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 2.7.** A vector $v \in R^N$ is said to be a tangent vector to the set $\mathcal{D} \subset R^N$ 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 [2],

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

We conclude by extending our results to differentiable functions. We first state a simple observation (also made for MADS methods in [2]) that if the set $\mathcal{D}$ is regular (*i. e.* $T_{\mathcal{D}}^{Cl}(x) = T_{\mathcal{D}}^{Co}(x)$) at $x^*$, then smooth limit points of $\mathcal{I}^{sf}$ are stationary with respect to the contingent cone.

**Corollary 2.8.** *Assume that $f_l$ is Lipschitz continuous on $\mathcal{D}$ and that Assumption 2.1 holds. Let $\mathcal{S} = \{\xi_n\}$ be a convergent subsequence of $\mathcal{I}^{sf}$ with limit $x^*$, corresponding directions $\{W_n\}$, and scales $\{h_n\}$. Assume that that $x^*$ is a smooth limit point of $\mathcal{I}^{sf}$.*

*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 2.8 to the case of smoother $f_l$. We say that $f_l$ is strictly differentiable at $x$ if the generalized gradient at $x$

$$\partial f(x) = \{s \in R^N \mid f^o(x; v) \geq v^T s \text{ for all } v \in R^N \}$$

is a single point, which we denote by $\nabla f(x)$. In this case

$$f^o(x; v) = \nabla f(x)^T v,$$

for all $v \in R^N$.

$x$ is a contingent KKT stationary point of $f$ on $\mathcal{D}$ if $f$ is strictly differentiable at $x$ and $-\nabla f(x)^T v \leq 0$ for all $v \in T_{\mathcal{D}}^{Co}(x^*)$.

We state the constraint qualifications needed to show that smooth limit points of $\mathcal{I}^{sf}$ are contingent KKT stationary points.

**Theorem 2.9.** *Assume that $f_l$ is Lipschitz continuous on $\mathcal{D}$, strictly differentiable at $x^*$, and that Assumption 2.1 holds. Let $\mathcal{S} = \{\xi_n\}$ be a convergent subsequence of $\mathcal{I}^{sf}$ with limit $x^*$, corresponding directions $\{W_n\}$, and scales $\{h_n\}$. Assume that $x^*$ is a smooth limit point of $\mathcal{I}^{sf}$. 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 [12, 2], 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 Cramer for their thoughtful comments on a preliminary version of this paper.

## References

[1] C. Audet and J.E. Dennis, Analysis of generalized pattern searches, *SIAM J. Optim.* 13 (2003) 889–903.

[2] 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.

[3] D.P. Bertsekas, On the Goldstein-Levitin-Polyak gradient projection method, *IEEE Trans. Autom. Control* 21 (1976) 174–184.

[4] A.J. Booker, J.E. Dennis, P.D. Frank, D.B. Serafini, V. Torczon and M.W. Trosset, A rigorous framework for optimization of expensive function by surrogates, *Structural Optimization* 17 (1999) 1–13.

[5] D.M. Bortz and C.T. Kelley, The simplex gradient and noisy optimization problems, in *Computational Methods in Optimal Design and Control*, J.T. Borggaard, J. Burns, E. Cliff and S. Schreck (eds.), vol. 24 of Progress in Systems and Control Theory, Birkhäuser, Boston, 1998, pp. 77–90.

[6] R.G. Carter, Nonsequential dynamic programming for optimizing pipelines, Presentation at the 1999 SIAM Conference on Optimization.

[7] R.G. Carter, Numerical experience with a class of algorithms for nonlinear optimization using inexact function and gradient information, *SIAM J. Sci. Comput.* 14 (1993) 368–388.

[8] G.W. Characklis, B.R. Kirsch, K.E.M. Dillard and C.T. Kelley, More efficient optimization of long-term water supply portfolios. Submitted for publication, 2008.

[9] G.W. Characklis, B.R. Kirsch, J. Ramsey, K.E.M. Dillard and C.T. Kelley, Developing portfolios of water supply transfers, Water *Resources Research* 42 (2006) W05403–1 – W05403–14.

[10] T.D. Choi, *Bound Constrained Optimization*, PhD thesis, North Carolina State University, Raleigh, North Carolina, 1999.

[11] T.D. Choi, O.J. Eslinger, C.T. Kelley, J.W. David and M. Etheridge, Optimization of automotive valve train components with implicit filtering, *Optimization and Engineering* 1 (2000) 9–28.

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

[13] A.R. Conn, K. Scheinberg, and P.L. Toint, Recent progress in unconstrained optimization without derivatives, *Math. Prog. Ser. B* 79 (1997) 397–414.

[14] A.R. Conn, K. Scheinberg, and P.L. Toint, A derivative free optimization algorithm in practice, in *Proceeedings of 7-th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, St Louis, MO. Sept 2-4, 1998., 1998.

[15] A.R. Conn, K. Scheinberg and L.N. Vicente, *Introduction to derivative-free optimization*, unpublished manuscript, 2007.

[16] J.W. David, C.Y. Cheng, T.D. Choi, C.T. Kelley and J. Gablonsky, *Optimal design of high speed mechanical systems*, Tech. Rep. CRSC-TR97-18, North Carolina State University, Center for Research in Scientific Computation, July 1997, Mathematical Modeling and Scientific Computing, to appear in Vol 9.

[17] J.W. David, C.T. Kelley and C.Y. Cheng, *Use of an implicit filtering algorithm for mechanical system parameter identification*, 1996, SAE Paper 960358, 1996 SAE International Congress and Exposition Conference Proceedings, Modeling of CI and SI Engines, pp. 189–194, Society of Automotive Engineers, Washington, DC.

[18] J.E. Dennis and R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, no. 16 in Classics in Applied Mathematics, SIAM, Philadelphia, 1996.

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

[20] R. Fletcher, *Practical methods of optimization*, John Wiley and Sons, New York, 1987.

[21] K.R. Fowler, C.T. Kelley, C.E. Kees and C.T. Miller, A hydraulic capture application for optimal remeidation design, in *Proceedings of Computational Methods in Water Resources XV*, C.T. Miller, M.W. Farthing, W.G. Gray and G.F. Pinter (eds.), Amsterdam, 2004, Elsevier, pp. 1149–1158.

[22] K.R. Fowler, J.P. Reese, C.E. Kees, J.E. Dennis, C. T. Kelley, C.T. Miller, C. Audet, A. J. Booker, G. Couture, R.W. Darwin, M.W. Farthing, D.E. Finkel, J.M. Gablonsky, G. Gray and T.G. Kolda, A comparison of derivative-free optimization methods for groundwater supply and hydraulic capture problems, *Advances in Water Resources* 31 (2008) 743–757.

[23] P.E. Gill, W. Murray and M.H. Wright, *Practical Optimization*, Academic Press, London, 1981.

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

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

[26] R. Hooke and T.A. Jeeves, 'Direct search' solution of numerical and statistical problems, *Journal of the Association for Computing Machinery* 8 (1961) 212–229.

[27] P.D. Hough, T.G. Kolda and V.J. Torczon, Asynchronous parallel pattern search for nonlinear optimization, *SIAM J. Sci. Comput.* 23 (2001) 134–156.

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

[29] D. Jones, C. Perttunen and B. Stuckman, Lipschitzian optimization without the Lipschitz constant, *J. Optim. Theory Appl.* 79 (1993) 157–181.

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

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

[32] R.M. Lewis and V. Torczon, Pattern search algorithms for linearly constrained minimization, *SIAM J. Optim.* 10 (2000) 917–941.

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

[34] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer, New York, 1999.

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

[36] V. Torczon, On the convergence of the multidimensional direct search, *SIAM J. Optim.* 1 (1991) 123–145.

[37] W. Yu, Positive basis and a class of direct search techniques, *Scientia Sinica, Speical Issue of Mathematics* 1 (1979) 53–67.

---

*Manuscript received 15 January 2008*
*revised 29 April 2008*
*accepted for publication 9 May 2008*

D.E. Finkel
Advanced Concepts and Technologies, MIT Lincoln Laboratory
244 Wood St, Lexington, MA 02420-9108, USA
E-mail address: `dfinkel@ll.mit.edu`

C.T. Kelley
North Carolina State University, Department of Mathematics
Box 8205, Raleigh, N. C. 27695-8205, USA
E-mail address: `Tim_Kelley@ncsu.edu`