SIAM J. OPTIM.
Vol. x, No. x, pp. xxx-xxx

CONVERGENCE OF MESH ADAPTIVE DIRECT SEARCH TO SECOND-ORDER STATIONARY POINTS *

MARK A. ABRAMSON † AND CHARLES AUDET ‡

Abstract. A previous analysis of second-order behavior of generalized pattern search algorithms for unconstrained and linearly constrained minimization is extended to the more general class of mesh adaptive direct search (MADS) algorithms for general constrained optimization. Because of the ability of MADS to generate an asymptotically dense set of search directions, we are able to establish reasonable conditions under which a subsequence of MADS iterates converges to a limit point satisfying second-order necessary or sufficient optimality conditions for general set-constrained optimization problems.

Key words. nonlinear programming, mesh adaptive direct search, derivative-free optimization, convergence analysis, second-order optimality conditions

1. Introduction. In this paper, we consider the class of derivative-free mesh adaptive direct search (MADS) algorithms applied to general constrained optimization problems of the form,

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

with $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ and $\Omega \subseteq \mathbb{R}^n$.

We treat the constraints by the "barrier" approach of applying the algorithm, not to f, but to the barrier objective function $f_\Omega = f + \psi_\Omega$, where ψ_Ω is the indicator function for Ω; *i.e.*, it is zero on Ω, and infinity elsewhere. If a point x is not in Ω, then we set $f_\Omega(x) = \infty$, and f is not evaluated. This is important in many practical engineering problems in which f is expensive to evaluate.

The class of MADS algorithms was introduced and analyzed in [4], as an extension of generalized pattern search (GPS) algorithms [3, 21] for solving nonlinearly constrained problems. Rather than applying a penalty function [18] or filter [5] approach to handle the nonlinear constraints, MADS defines an additional parameter that enables the algorithm to perform an exploration of the space of variables in an asymptotically dense set of directions. Under mild assumptions, the Clarke [9] calculus together with three types of tangent cones (hypertangent, Clarke tangent and contingent cones) are used to prove convergence of a subsequence of iterates to a point satisfying certain first-order conditions for optimality. An implementable instance of MADS is introduced in [4], in which positive spanning directions are chosen in a random fashion and almost sure convergence to a first-order stationary point is obtained. A similar first-order analysis is done in [15] for the DIRECT algorithm.

This paper extends the MADS analysis to show convergence to points satisfying certain second-order stationarity properties, in a manner similar to that of [1] for GPS. An important result of [1], is that the iterates produced by a GPS algorithm on a sufficiently smooth problem cannot converge in an infinite number of steps to a local maximizer. We show here that it may, unfortunately, converge in an infinite number of steps to a saddle point. The analysis in the present paper gives sufficient conditions under which a subsequence of the iterates produced by a MADS algorithm converges to a strict local minimizer. The necessary optimality condition is not based on any of the three tangent cones used in [4], but rather on the cone of feasible directions.

* Work of the second author was supported by FCAR grant NC72792 and NSERC grant 239436-05, AFOSR F49620-01-1-0013, and ExxonMobil Upstream Research Company.

† Air Force Institute of Technology, Department of Mathematics and Statistics, 2950 Hobson Way, Wright Patterson AFB, Ohio, 45433 USA (Mark.Abramson@afit.edu, http://en.afit.edu/ENC/Faculty/MAbramson/abramson.html)

‡ École Polytechnique de Montréal and GERAD, Département de Mathématiques et de Génie Industriel, C.P. 6079, Succ. Centre-ville, Montréal (Québec), H3C 3A7 Canada (Charles.Audet@gerad.ca, http://www.gerad.ca/Charles.Audet)

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2005		**00-00-2005 to 00-00-2005**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Convergence of Mesh Adaptive Direct Search to Second-Order Stationary Points	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Air Force Institute of Technology,Department of Mathematics and Statistics,2950 Hobson Way Building 640,Wright Patterson AFB,OH,45433-7765**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**13**	

The paper is outlined as follows. The MADS algorithm is briefly described in Section 2, with first-order properties restated in Section 3. Section 4 introduces the generalized Hessian [16] with some associated properties, followed by necessary and sufficient second-order optimality conditions and convergence results. Section 5 provides some examples to illustrate the strength of these results, and Section 6 offers some concluding remarks.

Notation. $\mathbb{R}$, $\mathbb{Z}$, and $\mathbb{N}$ denote the set of real numbers, integers, and nonnegative integers, respectively. For any set S, $\text{int}(S)$ denotes its interior, and $\text{cl}(S)$ its closure. For any matrix A, the notation $a \in A$ means that a is a column of A. For $x \in \mathbb{R}^n$ and $\varepsilon > 0$, we denote by $B_\varepsilon(x)$ the open ball $\{y \in \mathbb{R}^n : \|y-x\| < \varepsilon\}$. We say that f is $C^{1,1}$ near x if there exists an open set S containing x such that f is continuously differentiable with Lipschitz derivatives for every point in S. The reader is invited to consult [16] for a discussion and examples of $C^{1,1}$ functions.

2. Mesh Adaptive Direct Search. Like GPS methods, each iteration k of a MADS algorithm is characterized by two steps – an optional SEARCH step and a local POLL step, in which f_Ω is evaluated at specified points that lie on a mesh. The mesh is constructed from a finite fixed set of n_D directions $D \subset \mathbb{R}^n$ scaled by a mesh size parameter $\Delta_k^m > 0$. The directions form a positive spanning set [14] (*i.e.*, nonnegative linear combinations of D must span $\mathbb{R}^n$), and each direction $d \in D$, must be constructed as the product Gz, where $G \in \mathbb{R}^{n\times n}$ is a nonsingular generating matrix, and $z \in \mathbb{Z}^n$ is a vector of integers.

The following definition, taken from [4] and [5], precisely defines the current mesh so that all previously visited points lie on the current mesh.

DEFINITION 2.1. *At iteration k, the* current mesh *is defined to be the following union:*

$$M_k = \bigcup_{x \in S_k} \{x + \Delta_k^m Dz : z \in \mathbb{N}^{n_D}\},$$

where S_k is the finite set of points where the objective function f had been evaluated by the start of iteration k, and S_0 is a finite set of initial feasible points.

In both the SEARCH and POLL steps, the algorithm seeks to find an *improved mesh point*; *i.e.*, a point $y \in M_k$ for which $f_\Omega(y) < f_\Omega(x_k)$, where x_k is the current iterate or incumbent best iterate found thus far.

The SEARCH step allows evaluation of f_Ω at any finite set of mesh points. Any strategy may be used, including none. This is more restrictive than the frame methods of Coope and Price [12], but it helps to ensure convergence without a sufficient decrease condition or any other assumptions on mesh directions. The SEARCH step adds nothing to the convergence theory, but well-chosen SEARCH strategies can greatly improve algorithm performance (see [2, 6, 7, 19]).

In the POLL step, f_Ω is evaluated at points adjacent to the current iterate in a subset of the mesh directions. Unlike GPS, the class of MADS algorithms has a second mesh parameter Δ_k^p, called the *poll size parameter*, which satisfies $\Delta_k^m \le \Delta_k^p$ for all k, and also

$$\lim_{k\in K} \Delta_k^m = 0 \Leftrightarrow \lim_{k\in K} \Delta_k^p = 0 \text{ for any infinite subset of indices } K. \tag{2.1}$$

Under this construction, GPS methods now become the specific MADS instance in which $\Delta_k = \Delta_k^p = \Delta_k^m$.

The set of points generated in the POLL step is called a *frame*, with x_k referred to as the *frame center*. These terms are now formally defined as follows:

DEFINITION 2.2. *At iteration k, the* MADS frame *is defined to be the set:*

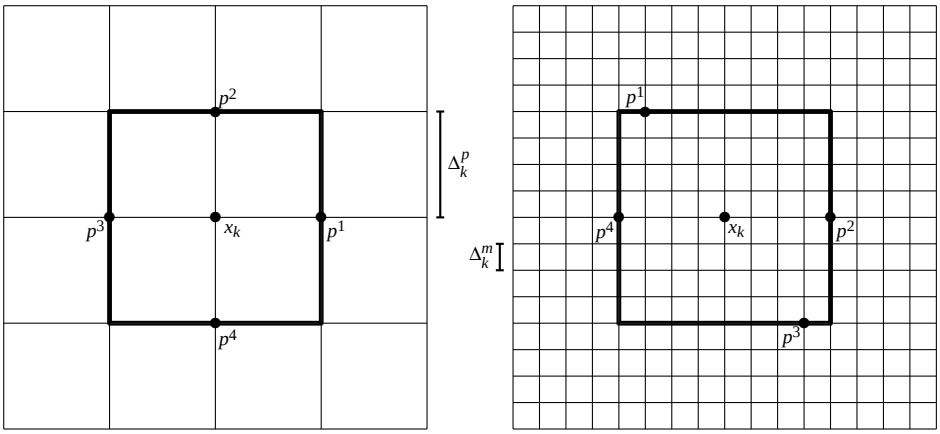
$$P_k = \{x_k + \Delta_k^m d : d \in D_k\} \subset M_k,$$

where D_k is a positive spanning set such that for each $d \in D_k$,

- $d \neq 0$ *can be written as a nonnegative integer combination of the directions in* D : $d = Du$ *for some vector* $u \in \mathbb{N}^{n_D}$ *that may depend on the iteration number* k
- *the distance from the frame center* x_k *to a poll point* $x_k + \Delta_k^m d$ *is bounded by a constant times the poll size parameter :* $\Delta_k^m \|d\| \leq \Delta_k^p \max\{\|d'\| : d' \in D\}$
- *limits (as defined in Coope and Price [11]) of the normalized sets* D_k *are positive spanning sets.*

In GPS, the set of directions D_k used to construct the frame is a subset of the finite set D. There is more flexibility in MADS. In [4], an instance of MADS is presented in which the closure of the cone generated by the set $\bigcup_{k=1}^{\infty} \left\{ \frac{d}{\|d\|} : d \in D_k \right\}$ equals $\mathbb{R}^n$. In this case, we say that the set of poll directions is *asymptotically dense* in $\mathbb{R}^n$.

Figure 2.1 illustrates typical GPS and MADS frames in $\mathbb{R}^2$ using the standard $2n$ coordinate directions. In each case, the mesh M_k is the set of points at the intersections of the horizontal and vertical lines. The thick lines delimit the points that are at a relative distance equal to the poll size parameter Δ_k^p from the frame center x_k. In MADS, the mesh size parameter Δ_k^m is much smaller than the poll size parameter; this allows many more possibilities in the frame construction.



A GPS frame A MADS frame

FIG. 2.1. *GPS and MADS frames* $P_k = \{p^1, p^2, p^3, p^4\}$ *around the frame center* x_k*, with the same poll size parameter* Δ_k^p.

If the POLL step fails to produce an improved mesh point, P_k is said to be a *minimal frame* with *minimal frame center* x_k. If either the SEARCH or POLL step is successful in finding an improved mesh point, the improved mesh point becomes the new current iterate $x_{k+1} \in \Omega$, and the mesh is either retained or coarsened. If neither step is successful, then the minimal frame center is retained as the current iterate (*i.e.*, $x_{k+1} = x_k$) and the mesh is refined.

Rules for refining and coarsening the mesh are as follows. Given a fixed rational number $\tau > 1$ and two integers $w^- \leq 1$ and $w^+ \geq 0$, the mesh size parameter Δ_k^m is updated according to the rule,

$$\Delta_{k+1}^m = \tau^{w_k} \Delta_k^m$$
$$\text{for some } w_k \in \begin{cases} \{0, 1, \ldots, w^+\} & \text{if an improved mesh point is found} \\ \{w^-, w^- + 1, \ldots, -1\} & \text{otherwise.} \end{cases} \tag{2.2}$$

The class of MADS algorithms is stated formally as follows:

A GENERAL MADS ALGORITHM

- INITIALIZATION: Let $x_0 \in \Omega$, set $\Delta_0^p \geq \Delta_0^m > 0$. Set the iteration counter k to 0.
- SEARCH AND POLL STEP: Perform the SEARCH and possibly the POLL steps until an improved mesh point x_{k+1} is found on the mesh M_k (see Definition 2.1).
  - OPTIONAL SEARCH: Evaluate f_Ω on a finite subset of trial points on the mesh M_k.
  - LOCAL POLL: Evaluate f_Ω on the frame P_k (see Definition 2.2).
- PARAMETER UPDATE: Update Δ_{k+1}^m according to (2.2), and Δ_{k+1}^p according to (2.1). Increase $k \leftarrow k+1$ and go back to the SEARCH and POLL step.

3. Existing First-Order Stationarity Results. Before presenting new results, we reproduce known convergence properties of MADS, originally published in [4]. All results are based on the following assumptions:

A1. A feasible initial point x_0 is provided.

A2. The initial objective function value $f(x_0)$ is finite.

A3. All iterates $\{x_k\}$ generated by MADS lie in a compact set.

Under these assumptions, Audet and Dennis [4] prove that

$$\liminf_{k\to+\infty} \Delta_k^p \;=\; \liminf_{k\to+\infty} \Delta_k^m \;=\; 0.$$

This ensures the existence of infinitely many minimal frame centers, since Δ_k^m only shrinks when a minimal frame is found. The following definition, taken from [4], is needed for later results.

DEFINITION 3.1. *A subsequence of the MADS iterates consisting of minimal frame centers, $\{x_k\}_{k\in K}$ for some subset of indices K, is said to be a* refining subsequence *if $\{\Delta_k^p\}_{k\in K}$ converges to zero.*

Let $\hat{x}$ be the limit of a convergent refining subsequence. If $\lim_{k\in L} \frac{d_k}{\|d_k\|}$ exists for some subset $L \subseteq K$ with poll direction $d_k \in D_k$, and if $x_k + \Delta_k^m d_k \in \Omega$ for infinitely many $k \in L$, then this limit is said to be a refining direction *for $\hat{x}$.*

Existence of refining subsequences for MADS was proved in [4]. The following four definitions [9, 17, 20] are needed in the main theorems.

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

$$y + tw \in \Omega \qquad \text{for all } y \in \Omega \cap B_\varepsilon(x), \quad w \in B_\varepsilon(v) \;\text{ and }\; 0 < t < \varepsilon. \tag{3.1}$$

The set of hypertangent vectors to Ω at x is called the hypertangent cone to Ω at x *and is denoted by $T_\Omega^H(x)$.*

DEFINITION 3.3. *A vector $v \in \mathbb{R}^n$ is said to be a* Clarke tangent vector *to the set $\Omega \subset \mathbb{R}^n$ at the point $x \in \mathrm{cl}(\Omega)$ if for every sequence $\{y_k\}$ of elements of Ω 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 $y_k + t_k w_k \in \Omega$. The set $T_\Omega^{Cl}(x)$ of all Clarke tangent vectors to Ω at x is called the* Clarke tangent cone *to Ω at x.*

DEFINITION 3.4. *A vector $v \in \mathbb{R}^n$ is said to be a* tangent vector *to the set $\Omega \subset \mathbb{R}^n$ at the point $x \in \mathrm{cl}(\Omega)$ if there exists a sequence $\{y_k\}$ of elements of Ω 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 $T_\Omega^{Co}(x)$ of all tangent vectors to Ω at x is called the* contingent cone *(or* sequential Bouligand tangent cone*) to Ω at x.*

DEFINITION 3.5. *The set Ω is said to be* regular *at x if $T_\Omega^{Cl}(x) = T_\Omega^{Co}(x)$.* In addition to these definitions, we add the following clarifying remarks, due to Clarke [9] unless otherwise noted:

- Any convex set is regular at each of its points.
- Both $T_\Omega^{Co}(x)$ and $T_\Omega^{Cl}(x)$ are closed, and both $T_\Omega^{Cl}(x)$ and $T_\Omega^{H}(x)$ are convex.
- $T_\Omega^{H}(x) \subseteq T_\Omega^{Cl}(x) \subseteq T_\Omega^{Co}(x)$.
- Rockafellar [20] showed that, if $T_\Omega^{H}(x)$ is nonempty, $T_\Omega^{H}(x) = \text{int}(T_\Omega^{Cl}(x))$, and therefore, $T_\Omega^{Cl}(x) = \text{cl}(T_\Omega^{H}(x))$.

In order to establish the results of this section, we apply a generalization of the Clarke [9] directional derivative, as presented in [17], in which function evaluations are restricted to points in the domain. Specifically, the Clarke generalized directional derivative of the locally Lipschitz function f at $x \in \Omega$ in the direction $v \in \mathbb{R}^n$ is defined by

$$f^\circ(x;v) := \limsup_{\substack{y \to x,\ y \in \Omega \\ t \downarrow 0,\ y+tv \in \Omega}} \frac{f(y+tv)-f(y)}{t}. \tag{3.2}$$

The next definition, also from [4], provides some nonsmooth terminology for stationarity.

DEFINITION 3.6. *Let f be Lipschitz near $\hat{x} \in \Omega$. Then, $\hat{x}$ is said to be a Clarke, or contingent stationary point of f over Ω, if $f^\circ(\hat{x};v) \geq 0$ for every direction v in the Clarke tangent cone, or contingent cone, to Ω at $\hat{x}$, respectively.*

In addition, $\hat{x}$ is said to be a Clarke, or contingent KKT stationary point of f over Ω, if $-\nabla f(\hat{x})$ exists and belongs to the polar of the Clarke tangent cone, or contingent cone, to Ω at $\hat{x}$, respectively.

If $\Omega = \mathbb{R}^n$ or $\hat{x}$ lies in the interior of Ω, then a stationary point as described by Definition 3.6 meets the condition that $f^\circ(\hat{x};v) \geq 0$ for all $v \in \mathbb{R}^n$. This is equivalent to $0 \in \partial f(\hat{x})$, the generalized gradient of f at $\hat{x}$ [9], which is defined by

$$\partial f(x) := \left\{ s \in \mathbb{R}^n : f^\circ(x;v) \geq v^T s \text{ for all } v \in \mathbb{R}^n \right\}.$$

The function f is said to be *strictly differentiable* at x if the generalized gradient of f at x is a singleton; *i.e.*, $\partial f(x) = \{\nabla f(x)\}$.

We now restate the main results from [4]. Theorem 3.7 is a directional result obtained under very mild assumptions, and Theorem 3.8, the main result of this section, is a restatement of four different theorems found in [4].

THEOREM 3.7. *Let f be Lipschitz near a limit $\hat{x} \in \Omega$ of a refining subsequence, and $v \in T_\Omega^H(\hat{x})$ be a refining direction for $\hat{x}$. Then the generalized directional derivative of f at $\hat{x}$ in the direction v is nonnegative,* i.e., *$f^\circ(\hat{x};v) \geq 0$.*

THEOREM 3.8. *Let $\hat{x} \in \Omega$ be the limit of a refining subsequence, and assume that $T_\Omega^H(\hat{x}) \neq \emptyset$ and the set of refining directions is dense in $T_\Omega^H(\hat{x})$.*

1. *If f is Lipschitz near $\hat{x}$, then $\hat{x}$ is a Clarke stationary point of f on Ω.*
2. *If f is strictly differentiable at $\hat{x}$, then $\hat{x}$ is a Clarke KKT stationary point of f on Ω.*

Furthermore, if Ω is regular at $\hat{x}$, then the following hold:

1. *If f is Lipschitz near $\hat{x}$, then $\hat{x}$ is a contingent stationary point of f on Ω.*
2. *If f is strictly differentiable at $\hat{x}$, then $\hat{x}$ is a contingent KKT stationary point of f on Ω.*

4. New Second-Order Stationarity Results. This section contains second-order convergence theory for MADS. In Section 4.1 we recall the definition of the generalized Hessian and identifies some useful properties. In Section 4.2 we present second-order necessary and sufficient conditions for optimality for set-constrained optimization problems. Finally, in Section 4.3, we establish conditions under which convergence of MADS iterates to a point satisfying second-order necessary and sufficient conditions is achieved.

4.1. Generalized Second-Order Derivatives. Before proving convergence to second-order points, we present nonsmooth notions of second derivatives and introduce second-order optimality conditions. Generalized second-order directional derivatives are developed in [10] and [16], consistent with the Clarke [9] calculus for first order derivatives. In this paper, we follow the Hirriart-Urruty *et al.* [16] definition of a generalized Hessian, given as follows.

DEFINITION 4.1. *Let $g : \mathbb{R}^n \to \mathbb{R}$ be $C^{1,1}$ near $x \in \Omega \subseteq \mathbb{R}^n$. The* generalized Hessian *of g at x, denoted by $\partial^2 g(x)$, is the set of matrices defined as the convex hull of the set*

$$\{A \in \mathbb{R}^{n\times n} : \text{ there exists } x_k \to x \text{ with } g \text{ twice differentiable at } x_k \text{ and } \nabla^2 g(x_k) \to A\}.$$

By construction, $\partial^2 g(x)$ is a nonempty, compact, and convex set of symmetric matrices [16]. The function g is said to be *twice strictly differentiable* at x if the generalized Hessian is a singleton; *i.e.*, $\partial^2 g(x) = \{\nabla^2 g(x)\}$. Furthermore, as a set-valued mapping, $x \rightrightarrows \partial^2 g(x)$ has two key properties, also identified in [16], which are necessary to establish optimality conditions in the next section.

- $\partial^2 g(x)$ is a *locally bounded* set-valued mapping:
  Given a matrix norm $\|\cdot\|$, there exists an $\varepsilon > 0$ and $\eta \in \mathbb{R}$ such that
  $$\sup\{\|A\| : A \in \partial^2 g(y), y \in B_\varepsilon(x)\} \ \leq \ \eta;$$
- $\partial^2 g(x)$ is a *closed* set-valued mapping:
  If $x_k \to x$ and $A_k \to A$ with $A_k \in \partial^2 g(x_k)$ for all k, then $A \in \partial^2 g(x)$.

The following second-order Taylor series result also comes from [16].

THEOREM 4.2. *Let $g : \mathbb{R}^n \to \mathbb{R}$ be $C^{1,1}$ in a open set $U \subset \mathbb{R}^n$, and let $[a,b] \subset U$ be a line segment. Then there exist an $x \in]a,b[$ and a matrix $A_x \in \partial^2 f(x)$ such that*

$$g(b) \ = \ g(a) + (b-a)^T \nabla f(a) + \frac{1}{2}(b-a)^T A_x (b-a).$$

In the next section, we apply this result to feasible points that may lie on the boundary of Ω. We are able to do this because our assumptions on the local smoothness of f are independent of Ω.

4.2. Second-Order Optimality Conditions. Second-order necessary and sufficient optimality conditions for constrained problems are traditionally expressed in terms of the Lagrangian function. However, our use of the barrier approach in handling constraints provides no useful information about the constraint gradients, and thus prevents us from proving anything with respect to traditional optimality conditions. Therefore, instead of dealing with the Lagrangian function, we extend optimality conditions for set-constrained problems (see [8] for further discussions).

We now establish Clarke-based second-order necessary and sufficient conditions for set-constrained optimality. The proof for the former is very similar to one found in [16] for unconstrained problems, the only difference being the first-order condition satisfied by the local minimizer. It is expressed in terms of feasible directions, formally given in Definition 4.3.

DEFINITION 4.3. *The direction $v \in \mathbb{R}^n$ is said to be feasible for $\Omega \subset \mathbb{R}^n$ at $x \in \Omega$ if there exists an $\varepsilon > 0$ for which $x + tv \in \Omega$ for all $0 \leq t < \varepsilon$. The set of feasible directions for Ω at $x \in \Omega$ is a cone and is denoted by $T_\Omega^F(x)$.*

It follows immediately that $T_\Omega^H(x) \subseteq T_\Omega^F(x) \subseteq T_\Omega^{Co}(x)$ for any $x \in \Omega$. Moreover, if $T_\Omega^H(x) \neq \emptyset$ for some $x \in \Omega$, and if Ω is regular at x, then $\mathrm{cl}(T_\Omega^H(x)) = \mathrm{cl}(T_\Omega^F(x)) = T_\Omega^{Cl}(x) = T_\Omega^{Co}(x)$. However, without regularity it is possible that either of the following holds:

- $T_\Omega^{Cl}(x) \subset \text{int}(T_\Omega^F(x))$: e.g., if $\Omega = \{(a,b) \in \mathbb{R}^2 : a \geq 0 \text{ or } b \geq 0\}$, then $T_\Omega^{Cl}(0,0) = \mathbb{R}^2_+$ and $T_\Omega^F(0,0) = \Omega$,
- $\text{cl}(T_\Omega^F(x)) \subset \text{int}(T_\Omega^{Cl}(x))$: e.g., if $\Omega = \mathbb{R}^2 \setminus \{(-\frac{1}{k}, b) \in \mathbb{R}^2 : b \in \mathbb{R}, k = 1,2,\ldots\}$, then $T_\Omega^F(0,0) = \{(a,b) \in \mathbb{R}^2 : a \geq 0\}$ and $T_\Omega^{Cl}(0,0) = \mathbb{R}^2$.

THEOREM 4.4 (Second-order necessary condition for set-constrained optimality.). *Let $x^* \in \Omega$ be a local solution of (1.1). If f is $C^{1,1}$ near x^*, then any feasible direction $v \in T_\Omega^F(x^*)$ for which $v^T \nabla f(x^*) = 0$ satisfies $v^T A v \geq 0$ for some $A \in \partial^2 f(x^*)$.*

Proof. Let $v \in \mathbb{R}^n$ be a feasible direction that satisfies $v^T \nabla f(x^*) = 0$, and consider the sequence $\{x_k\}$, where $x_k = x^* + \frac{1}{k} v$. It follows that $x_k \in \Omega$ when k is sufficiently large. Then by second-order Taylor series in a neighborhood of the local minimizer x^*, we have for each k sufficiently large

$$0 \;\leq\; f(x_k) - f(x^*) \;=\; \frac{1}{k} \nabla f(x^*)^T v + \frac{1}{2k^2} v^T A_k v \;=\; \frac{1}{2k^2} v^T A_k v, \tag{4.1}$$

where $A_k \in \partial^2 f(\bar{x_k})$ for some $\bar{x_k} \in]x^*, x_k[$.

Since $\partial^2 f$ is locally bounded and $\bar{x_k} \to x^*$, the sequence $\{A_k\}$ is locally bounded and thus possesses an accumulation point A. Furthermore, since $\partial^2 f$ is a closed set-valued mapping, we have $A \in \partial^2 f(x^*)$. Taking limits in (4.1) leads to $v^T A v \geq 0$. ∎

Theorem 4.4 applies to the set of hypertangent vectors as well as feasible directions since the set of feasible directions contains the hypertangent cone. However, this necessary condition does not necessarily hold for directions in the Clarke tangent or contingent cone, as the following example shows.

EXAMPLE 4.5. *Consider the quadratic optimization problem, in which $f : \mathbb{R}^2 \to \mathbb{R}$ is defined by $f(a,b) = -(a^2 + b^2)$, and $\Omega = \{(a,b) \in \mathbb{R}^2 : a^2 + (b-1)^2 \leq 1\}$. The optimal solution is at $(0,2)$, where*

$$T_\Omega^H(0,2) = T_\Omega^F(0,2) = \{(v_1,v_2) : v_2 < 0\} \;\; \textit{and} \;\; T_\Omega^{Cl}(0,2) = T_\Omega^{Co}(0,2) = \{(v_1,v_2) : v_2 \leq 0\}.$$

The direction $v = (1,0)^T \in T_\Omega^{Cl}(0,2) = T_\Omega^{Co}(0,2)$, is not a feasible direction and makes a zero inner product with $\nabla f(0,2) = (0,-4)^T$, but the Hessian matrix is given by

$$\nabla^2 f(0,2) \;=\; \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix},$$

which yields $v^T \nabla^2 f(0,2) v = -2 < 0$.

THEOREM 4.6 (Second-order sufficient condition for set-constrained optimality.). *Let $x^* \in \Omega$ be a contingent stationary point for the optimization problem defined in (1.1), and suppose that $T_\Omega^H(x^*) \neq \emptyset$ and that Ω is convex near x^*. If f is $C^{1,1}$ near x^*, and if $v^T A v > 0$ for all matrices $A \in \partial^2 f(x^*)$ and all nonzero tangent directions $v \in T_\Omega^{Co}(x)$ that satisfy $v^T \nabla f(x^*) = 0$, then x^* is a strict local solution of (1.1).*

Proof. The proof is by contraposition. Suppose that x^* is not a strict local minimizer. Then there exists a sequence $\{y_k\} \subset \Omega$ (with $y_k \neq x^*$) converging to x^* satisfying $f(y_k) \leq f(x^*)$ for all k. By taking subsequences if necessary, we can assume that the sequence $\{w_k\}$ with $w_k = \frac{y_k - x^*}{\|y_k - x^*\|}$ converges to some vector $v \in \mathbb{R}^n$.

Local convexity of Ω near x^* implies that v and w_k are contingent directions for all $k \geq \ell$, for some integer $\ell \geq 0$. Moreover, since x^* is assumed to be a contingent stationary point, and since f is continuously differentiable, then $v^T \nabla f(x^*) \geq 0$ and $w_k^T \nabla f(x^*) \geq 0$ for all $k \geq \ell$. However, since $f(y_k) \leq f(x^*)$ for all k, then $v^T \nabla f(x^*) = 0$.

Theorem 4.2 on Taylor series ensures that for each $k \geq \ell$, there exists some matrix $A_k \in \partial^2 f(\bar{x_k})$ with

$\bar{x_k} \in]x^*, y_k[$ such that

$$0 \;\geq\; f(y_k) - f(x^*) = (y_k - x^*)^T \nabla f(x^*) + \frac{1}{2}(y_k - x^*)^T A_k (y_k - x^*)$$
$$\geq \frac{1}{2}(y_k - x^*)^T A_k (y_k - x^*). \tag{4.2}$$

Now, since $\bar{x_k} \to x^*$, and since $\partial^2 f(x^*)$ is a closed locally bounded set-valued mapping, then there exists an accumulation point $A \in \partial^2 f(x^*)$ of the sequence $\{A_k\}$. Dividing (4.2) by $\|y_k - x^*\|^2$ and taking limits leads to $0 \geq \frac{1}{2} v^T A v$, where $v \neq 0$ belongs to $T_\Omega^{Co}(x^*)$ and satisfies $v^T \nabla f(x^*) = 0$. ∎

The previous theorem requires as an assumption that the set Ω is locally convex. The following example shows that regularity of Ω is not sufficient to guarantee a local minimizer.

EXAMPLE 4.7. *Consider the quadratic optimization problem, in which $f : \mathbb{R}^2 \to \mathbb{R}$ is defined by $f(a,b) = a^2 + 2b$, and $\Omega = \{(a,b) \in \mathbb{R}^2 : 2a^2 + b \geq 0\}$. The solution $x^T = (0,0)$ is a contingent stationary point, and Ω is regular at x since*

$$T_\Omega^{Cl}(x) = T_\Omega^{Co}(x) = \{(v_1, v_2) : v_2 \geq 0\}.$$

The vector, $v = (1,0)^T \in T_\Omega^{Co}(x)$ satisfies $v^T \nabla f(x) = 0$ and $v^T \nabla^2 f(x) v = 2 > 0$. However, $(\varepsilon, -\varepsilon^2)$ belongs to the strict interior of Ω for all $\varepsilon \neq 0$, and $f(\varepsilon, -\varepsilon^2) = -\varepsilon^2 < 0 = f(x)$.

4.3. Second-Order Stationarity Results for MADS. The next two results are the main contributions of this paper. The first theorem establishes convergence of a subsequence of MADS iterates to a point satisfying the second-order necessary condition identified in Theorem 4.4, and the second establishes the sufficiency conditions of Theorem 4.6.

THEOREM 4.8. *Let f be $C^{1,1}$ near a limit $\hat{x}$ of a refining subsequence, and assume that $T_\Omega^H(\hat{x}) \neq \emptyset$ and that Ω is regular near $\hat{x}$. If the set of refining directions is dense in $T_\Omega^H(\hat{x})$, then $\hat{x}$ satisfies the second-order necessary condition for set-constrained optimality.*

Proof. Let $v \in \mathbb{R}^n$ be any nonzero feasible direction that satisfies $v^T \nabla f(\hat{x}) = 0$, and suppose, by way of contradiction, that $v^T \hat{A} v < 0$ for all matrices $\hat{A} \in \partial^2 f(\hat{x})$. Since $\partial^2 f(\hat{x})$ is nonempty and compact, and $\partial^2 f$ is a closed set-valued mapping, there exists some $\varepsilon > 0$ such that $v^T A v < 0$ for all $A \in \partial^2 f(x)$ and for all $x \in B_\varepsilon(\hat{x})$.

Let K denote the set of indices of unsuccessful iterations. Regularity of Ω, together with the assumption that $T_\Omega^H(\hat{x}) \neq \emptyset$ guarantee that $\mathrm{cl}(T_\Omega^H(\hat{x})) = T_\Omega^{Co}(\hat{x}) = \mathrm{cl}(T_\Omega^F(\hat{x}))$. Therefore, the denseness of the set of refining directions in $T_\Omega^H(\hat{x})$ ensures the existence of $\{w_k\}_{k \in K}$ converging to v with $w_k = \frac{d_k}{\|d_k\|}$, $d_k \in D_k$ for each $k \in K$. Taylor series yields

$$f(x_k + \Delta_k^p d_k) - f(x_k) = \;\; \Delta_k^p d_k^T \nabla f(x_k) + \frac{1}{2}(\Delta_k^p)^2 d_k^T A_k^+ d_k, \tag{4.3}$$

$$f(x_k - \Delta_k^p d_k) - f(x_k) = -\Delta_k^p d_k^T \nabla f(x_k) + \frac{1}{2}(\Delta_k^p)^2 d_k^T A_k^- d_k \tag{4.4}$$

where $A_k^+ \in \partial^2 f(x^+)$ for some $x^+ \in]x_k, x_k + \Delta_k^p d_k[$ and $A_k^- \in \partial^2 f(x^-)$ for some $x^- \in]x_k, x_k - \Delta_k^p d_k[$. Since $\Delta_k \to 0^+$ and $x_k \to \hat{x}$, there is a subsequence for which A_k^+ converges to some $A^+ \in \partial^2 f(\hat{x})$, and A_k^- converges to some $A^- \in \partial^2 f(\hat{x})$. Moreover, since $\partial^2 f(\hat{x})$ is a convex set, then $A = \frac{1}{2}(A^+ + A^-) \in \partial^2 f(\hat{x})$.

Adding (4.3) and (4.4) and substituting $d_k = \|d_k\| w_k$ yields

$$\frac{1}{\Delta_k^p \|d_k\|} \left[\frac{f(x_k + \Delta_k^p \|d_k\| w_k) - f(x_k)}{\Delta_k^p \|d_k\|} + \frac{f(x_k - \Delta_k^p \|d_k\| w_k) - f(x_k)}{\Delta_k^p \|d_k\|} \right] = w_k^T A_k w_k, \tag{4.5}$$

where $A_k = \frac{1}{2}(A_k^+ + A_k^-)$. Furthermore, since $w_k \to v$ and $v^T A v < 0$, there exists $\gamma < 0$ such that $w_k^T A_k w_k \leq \gamma < 0$ for all sufficiently large $k \in K$, which forces the left-hand side of (4.5) to also be negative and bounded away from zero. But since $d_k \in D_k$ for all sufficiently large $k \in K$, we have that $f(x_k) \leq f(x_k + \Delta_k^p d_k)$, which makes nonnegative the first term of the left-hand side of (4.5) (for all sufficiently large $k \in K$). Thus it must be the case that

$$\frac{f(x_k - \Delta_k^p \|d_k\| w_k) - f(x_k)}{\Delta_k^p \|d_k\|} \;\leq\; \gamma \;<\; 0 \tag{4.6}$$

for all sufficiently large $k \in K$. Taking the limit of (4.6) as $k \to \infty$ in K yields $\nabla f(\hat{x})^T(-v) < 0$, or $\nabla f(\hat{x})^T v > 0$, which contradicts the assumption that $\nabla f(\hat{x})^T v = 0$. ∎

The following result shows that the sufficient conditions of Theorem 4.6 can be satisfied by a subsequence of MADS iterates, given stronger hypotheses than those of Theorem 4.8.

THEOREM 4.9. *Let f be twice strictly differentiable at a limit $\hat{x}$ of a refining subsequence, and assume that $T_\Omega^H(\hat{x}) \neq \emptyset$, Ω is convex near $\hat{x}$, and $\nabla^2 f(\hat{x})$ is nonsingular. If the set of refining directions is dense in $T_\Omega^H(\hat{x})$, then $\hat{x}$ is a strict local minimizer of f on Ω.*

Proof. Since f is twice strictly differentiable at $\hat{x}$, $\partial^2 f(\hat{x}) = \{\nabla^2 f(\hat{x})\}$. Thus, it follows from Theorem 4.8 that $v^T \nabla^2 f(\hat{x}) v \geq 0$ for all feasible directions $v \in T_\Omega^F(\hat{x})$ satisfying $\nabla f(\hat{x})^T v = 0$. But since $\nabla^2 f(\hat{x})$ is assumed to be nonsingular, this inequality is strict. Furthermore, by Theorem 3.8 and the smoothness of f near $\hat{x}$, $\hat{x}$ is a first-order contingent stationary point. Thus the hypotheses of Theorem 4.6 are satisfied, and the result is proved. ∎

Clearly, these are strong results for a direct search method. However, in practice, achieving denseness of the refining directions in the hypertangent cone (a key assumption) requires increasingly more poll directions per iteration. To overcome this problem, an implementable instance of MADS is introduced in [4], called LTMADS, in which the positive spanning directions used at each iteration are limited in number, but chosen randomly from among the increasing number of possible poll directions. While this is not difficult to implement, the drawback is that denseness of the refining directions is only achieved *almost surely* (*i.e.*, with probability one). Thus, in practice, the convergence results proved both here and in [4] are only attained *almost surely*. This is a weaker measure of convergence, but it works well in practice [4]. We apply LTMADS to one of the numerical examples in the next section.

5. Examples. Second order results for GPS are presented in [1]. They are not as strong as those presented here for MADS. In this section, we illustrate this difference through a series of three bound constrained or unconstrained quadratic examples in $\mathbb{R}^2$. The first [1] illustrates how GPS, but not MADS, can converge to a local maximizer. The second shows how GPS, but not MADS, converges in an infinite number of iterations to a saddle point. Finally, since the second example requires an uncommon set of parameter choices, we recall a simpler numerical example [1] with more realistic parameter choices to show how MADS avoids a saddle point having narrow cones of descent, which GPS with reasonable parameter choices does not.

5.1. An example where GPS stalls at a global maximizer for a minimization problem. Consider the unconstrained problem of minimizing the function $f(a,b) = -(ab)^2$, subject to $-2 \leq a,b \leq 2$, with a starting point at $(0,0)$, the global maximizer. Using standard coordinate directions and their negatives as poll directions, GPS will stall at the starting point without moving. This is consistent with the result proved in [1], that convergence of GPS to a maximizer may only occur in a finite number of iterations, and when f is locally constant along all feasible poll directions. On the other hand, any reasonable implementation of MADS will generate more directions (a dense set in the limit), which is all that is needed to avoid stalling at

the maximizer. This is consistent with Theorem 4.8, which ensures that stalling at a local maximizer is not possible for MADS, provided the hypotheses are met.

5.2. An example where GPS converges in an infinite number of iterations to a saddle point. Consider the unconstrained quadratic optimization problem in which the polynomial objective function in $\mathbb{R}^2$ is $f(a,b) = a^2 + 3ab + b^2$. The point (0,0) is a saddle point, at which the descent directions lie in the cone generated by $a = \frac{1}{2}b(-3 \pm \sqrt{5})$.

We apply an instance of GPS where $D_k = D = \{e_1, e_2, -e_1, -e_2\}$ is constant throughout all iterations. On iterations that fail to improve the incumbent, the mesh size parameter is divided by 16. On successful iterations that follow an unsuccessful one, the mesh size is kept constant, and on other successful iterations, the mesh size parameter is multiplied by 8. Thus, the GPS parameters are $G = I$ (the identity matrix), $Z = D = [I;\ -I]$, $\tau = 2$, $w^- = -4$ and $w^+ = 3$.

Furthermore, we use an empty SEARCH and an opportunist POLL, i.e., an iteration terminates as soon as an improved mesh point is generated. Moreover, when the iteration number k modulo 3 is 1, the POLL step first evaluates $x_k - \Delta_k e_2$, and otherwise, the POLL step first evaluates $x_k - \Delta_k e_1$. The order in which the other poll points are explored is irrelevant to this example.

The initial parameters are $x_0^T = (1,1)$ with $f(x_0) = 5$ and $\Delta_0 = 8$. Figure 5.1 displays the first iterates generated by the algorithm. The figure also displays some level sets of f.

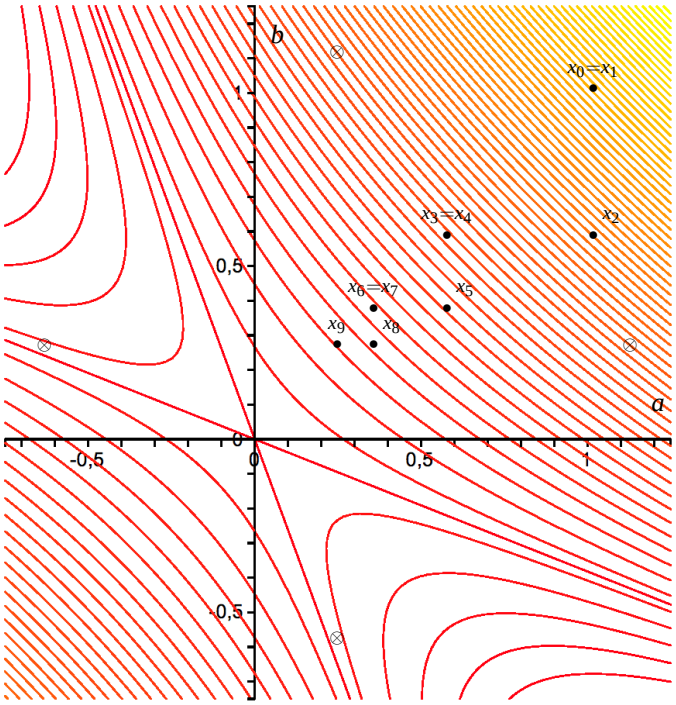


FIG. 5.1. *Initial GPS iterates*

We next show that the entire sequence of iterates converges to the origin. This happens because this instance of GPS never generates any trial points in the cone where f is negative. It either jumps over the cone, which results in an unsuccessful iteration, or take a small step which falls short of reaching the cone. For example, at iteration $k = 9$, the trial poll points are $(\frac{9}{8}, 1), (1, \frac{9}{8}), (-\frac{7}{8}, 1)$ and $(1, -\frac{7}{8})$. These four trial

points are represented by the symbol $\otimes$ in the figure.

PROPOSITION 5.1. *For any integer $\ell \geq 0$, the GPS iterates are such that $x_{3\ell} = x_{3\ell+1} = \left(2^{-\ell}, 2^{-\ell}\right)$, $x_{3\ell+2} = \left(2^{-\ell}, 2^{-\ell-1}\right)$, and $\Delta_{3\ell} = 2^{3-\ell}$, $\Delta_{3\ell+1} = \Delta_{3\ell+2} = 2^{-\ell-1}$.*

Proof. The proof is done by induction. The result is true for the initial iteration $k = 0$. Suppose that iteration $k = 3\ell$ is initiated with $\Delta_k = 2^{3-\ell}$ and $x_k = \left(2^{-\ell}, 2^{-\ell}\right)$. The current objective function value is $f(x_k) = 5 \times 4^{-\ell}$. Table 5.1 details the objective function values at the poll points for iterations $k = 3\ell, 3\ell+1$ and $3\ell+2$. Trial points that improve the incumbent appear in shaded boxes. This table shows that the iterate

k	x_k	Δ_k	Trial poll points at $x_k = (a_k, b_k)$
3ℓ	$(2^{-\ell}, 2^{-\ell})$	$2^{3-\ell}$	$f(a_k+\Delta_k, b_k) = f(9\times 2^{-\ell}, 2^{-\ell}) = 109 \times 4^{-\ell} = f(2^{-\ell}, 9\times 2^{-\ell})) = f(a_k, b_k+\Delta_k)$
			$f(a_k-\Delta_k, b_k) = f(-7\times 2^{-\ell}, 2^{-\ell}) = 29 \times 4^{-\ell} = f(2^{-\ell}, -7\times 2^{-\ell})) = f(a_k, b_k+\Delta_k)$
$3\ell+1$	$(2^{-\ell}, 2^{-\ell})$	$2^{-\ell-1}$	$f(a_k, b_k-\Delta_k) = f(2^{-\ell}, 2^{-\ell-1}) = 11 \times 4^{-\ell-1}$
$3\ell+2$	$(2^{-\ell}, 2^{-\ell-1})$	$2^{-\ell-1}$	$f(a_k-\Delta_k, b_k) = f(2^{-\ell-1}, 2^{-\ell-1}) = 5 \times 4^{-\ell-1}$
$3(\ell+1)$	$(2^{-\ell-1}, 2^{-\ell-1})$	$2^{2-\ell}$	...

TABLE 5.1
Iterations $k = 3\ell$ to $k = 3(\ell+1)$

for $k = 3(\ell+1)$ is $(2^{-\ell-1}, 2^{-\ell-1})$ and that the corresponding mesh size parameter is $2^{2-\ell}$. This concludes the proof. ∎

The previous proposition shows that the entire sequence of iterates generated by GPS converges to the saddle point $(0,0)$, which is not a local minimizer. Theorem 4.8 ensures that any MADS instance with an asymptotically dense set of refining directions will not converge to that saddle point, since the necessary optimality condition is not satisfied: $v^T = (-1,1)$ is a feasible direction for which $v^T \nabla f(0,0) = 0$ but

$$v^T \nabla^2 f(0,0) v \;=\; v^T \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix} v^T \;=\; -2$$

is negative.

5.3. An example where GPS reaches and stalls at a saddle point. Consider the bound constrained problem,

$$\min_{-2 \leq a,b \leq 2} \; f(a,b) \;=\; 99a^2 - 20ab + b^2 \;=\; (9a-b)(11a-b).$$

At the saddle point $(0,0)$, directions of descent lie only in the narrow cone formed by the lines $b = 9a$ and $b = 11a$. Thus to avoid stalling at the saddle point, GPS or MADS would have to generate a feasible iterate that lies inside this cone (see Figure 5.2). In this example, the SEARCH step is empty and the initial point is chosen to be $(1.01,\ 0.93)$. This starting point is chosen to be non-integral to make it more difficult for GPS to reach the integral point $(0,0)$. Both GPS and MADS were run using the NOMAD software package [13] with primarily default settings.

GPS (using the standard coordinate directions and their negatives as poll directions) reaches the saddle point at the 358th function evaluation with a poll size parameter of 10^{-17}. This implies that, regardless of the termination tolerance chosen, it stalls there because none of the poll directions are directions of descent. On the other hand, NOMAD's implementation of LTMADS successfully moved off of the saddle point to reach a local minimizer in 100 of 100 runs. This is again consistent with Theorem 4.8, since $v^T = (1,10)$ is a feasible direction for which $v^T \nabla f(0,0) = 0$ but

$$v^T \nabla^2 f(0,0) v \;=\; v^T \begin{bmatrix} 198 & -20 \\ -20 & 2 \end{bmatrix} v^T \;=\; -2$$

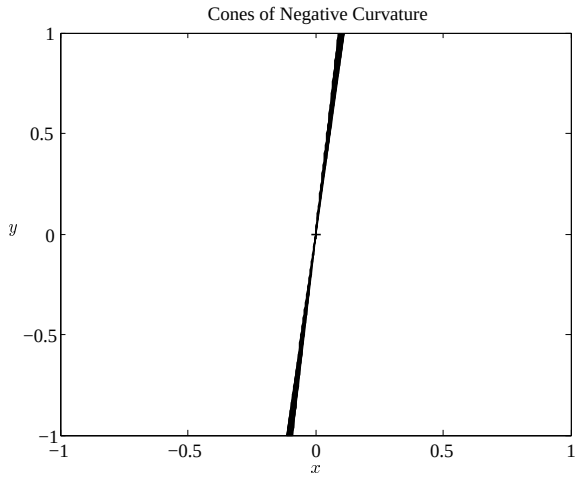


FIG. 5.2. *For $f(a,b) = (9a-b)(11a-b)$, the cones of descent at the saddle point (0,0) are shown in the shaded area between the lines $b = 9a$ and $b = 11a$.*

is negative.

6. Concluding Remarks. The theoretical results presented here establish strong convergence results for MADS. In spite of MADS being a derivative-free method, we have shown convergence of a subsequence of MADS iterates to a second-order stationary point under conditions weaker than standard Newton assumptions; namely, that f is continuously differentiable with Lipschitz derivatives near the limit point. Moreover, if Ω is locally convex and f twice strictly differentiable near the limit point, then the limit point is a local minimizer for (1.1).

In Section 5, we provided examples to illustrate the superior convergence properties of MADS over GPS. However, since our implementation involves random selection of positive spanning directions, the convergence properties established in Section 4.3 are achieved, in practice, with probability one. We envision a future area of research being the clever enumeration of these directions so that the stronger type of convergence is retained by an implementable instance of the algorithm. Specifically, we would like to deterministically generate an asymptotically dense set of directions in such a way that, after any finite number of iterations, the directions used by the algorithm are uniformly spaced (or as close to it as possible).

REFERENCES

[1] M. A. Abramson. Second-order behavior of pattern search. *SIAM J. Optim.*, 16(2):315–330, 2005.
[2] M. A. Abramson, C. Audet, and J. E. Dennis, Jr. Generalized pattern searches with derivative information. *Math. Programming*, Series B, 100:3–25, 2004.
[3] C. Audet and J. E. Dennis, Jr. Analysis of generalized pattern searches. *SIAM J. Optim.*, 13(3):889–903, 2003.
[4] C. Audet and J. E. Dennis, Jr. Mesh adaptive direct search algorithms for constrained optimization. *SIAM J. Optim.*, To appear, 2004.
[5] C. Audet and J. E. Dennis, Jr. A pattern search filter method for nonlinear programming without derivatives. *SIAM J. Optim.*, 14(4):980–1010, 2004.
[6] C. Audet and D. Orban. Finding optimal algorithmic parameters using the mesh adaptive direct search algorithm. Technical Report G-2004-96, Les Cahiers du GERAD, 2004. submitted to *SIAM J. Optim.*
[7] A. J. Booker, J. E. Dennis, Jr., P. D. Frank, D. B. Serafini, V. Torczon, and M. W. Trosset. A rigorous framework for optimization of expensive functions by surrogates. *Structural Optim.*, 17(1):1–13, February 1999.

[8] K. P. E. Chong and S. H. Zak. *Linear and Nonlinear Programming*. John Wiley and Sons, Inc., New York, 2001.
[9] F. H. Clarke. *Optimization and Nonsmooth Analysis*. Wiley, New York, 1983. Reissued in 1990 by SIAM Publications, Philadelphia, as Vol. 5 in the series Classics in Applied Mathematics.
[10] R. Cominetti and R. Correa. A generalized second-order derivative in nonsmooth optimization. *SIAM J. Control Optim.*, 28(4):789–809, 1990.
[11] I. D. Coope and C. J. Price. Frame-based methods for unconstrained optimization. *J. Optim. Theory Appl.*, 107(2):261–274, 2000.
[12] I. D. Coope and C. J. Price. On the convergence of grid-based methods for unconstrained optimization. *SIAM J. Optim.*, 11(4):859–869, 2001.
[13] G. Couture, C. Audet, J. E. Dennis, Jr., and M. A. Abramson. The NOMAD project. http://www.gerad.ca/NOMAD/.
[14] C. Davis. Theory of positive linear dependence. *Amer. J. Math.*, 76:733–746, 1954.
[15] D.E. Finkel and C.T. Kelley. Convergence analysis of the DIRECT algorithm. Technical report, NCSU Mathematics Department, 2004.
[16] J.-B. Hiriart-Urruty, J.-J. Strodiot, and V. H. Nguyen. Generalized Hessian matrix and second-order optimality conditions for problems with $C^{1,1}$ data. *Appl. Math. Optim.*, 11(1):43–56, 1984.
[17] J. Jahn. *Introduction to the Theory of Nonlinear Optimization*. Springer, Berlin, 1994.
[18] R. M. Lewis and V. Torczon. A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds. *SIAM J. Optim.*, 12(4):1075–1089, 2002.
[19] M. D. McKay, W. J. Conover, and R. J. Beckman. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.
[20] R. T. Rockafellar. Generalized directional derivatives and subgradients of nonconvex functions. *Canad. J. Math.*, 32(2):257–280, 1980.
[21] V. Torczon. On the convergence of pattern search algorithms. *SIAM J. Optim.*, 7(1):1–25, February 1997.