

Mesh Adaptive Direct Search Algorithms for Mixed Variable Optimization

Mark A. Abramson · Charles Audet · James
W. Chrissis · Jennifer G. Walston

July 16, 2007

Abstract In previous work, the generalized pattern search (GPS) algorithm for linearly constrained (continuous) optimization was extended to mixed variable problems, in which some of the variables may be categorical. In another paper, the mesh adaptive direct search (MADS) algorithm was introduced as a generalization of GPS for problems with general nonlinear constraints. The convergence analyses of these methods rely on the Clarke calculus for nonsmooth functions. In the present paper, we generalize both of these approaches by proposing an algorithm for problems with both mixed variables and general nonlinear constraints, called *mixed variable-MADS* (MV-MADS). A new convergence analysis is presented which generalizes the existing GPS and MADS theory.

Keywords nonlinear programming, mesh adaptive direct search, mixed variables, derivative-free optimization, convergence analysis

1 Introduction

In this paper, we generalize the class of mesh adaptive direct search algorithms to mixed variable optimization problems and establish a unifying convergence theory, so that the existing theorems act as corollaries to the results presented here. This is done in a relatively straightforward manner, somewhat similar to the work of Audet and Dennis [6] in their extension of pattern search algorithms to bound constrained mixed variable problems.

Mark A. Abramson

Air Force Institute of Technology, Department of Mathematics and Statistics, 2950 Hobson Way, Wright Patterson AFB, Ohio, 45433 USA E-mail: Mark.Abramson@afit.edu

Charles Audet

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

James W. Chrissis and Jennifer G. Walston

Air Force Institute of Technology, Department of Operational Sciences, 2950 Hobson Way, Wright Patterson AFB, Ohio, 45433 USA E-mail: James.Chrissis@afit.edu and Jennifer.Walston@afit.edu

Mixed variable optimization problems are characterized by a combination of continuous and categorical variables, the latter being discrete variables that must take their values from a finite pre-defined list or set of categories. Categorical variables may be nonnumeric, such as color, shape, or type of material; thus, traditional approaches involving branch-and-bound for solving mixed integer nonlinear programming (MINLP) problems are not directly applicable. There are cases where a problem modeled using categorical variables may be reformulated using nonlinear programming. For example, see [1] for a non-trivial reformulation of the problem presented in [16]. Such reformulations are impossible in the cases where the functions are provided as black-boxes.

To be as general as possible, we allow for the case where changes in the categorical variable values can mean a change in problem dimensions. We denote the *maximum* dimensions of the continuous and discrete variables by n^c and n^d , respectively, and we partition each point $x = (x^c, x^d)$ into its continuous and categorical components, so that $x^c \in \Omega^c \subseteq \mathbb{R}^{n^c}$ and $x^d \in \Omega^d \subseteq \mathbb{Z}^{n^d}$. We adopt the convention of ignoring unused variables.

The problem under consideration is given by

$$\min_{x \in \Omega} f(x) \quad (1)$$

where $f : \Omega \rightarrow \mathbb{R} \cup \{\infty\}$, and the domain Ω (feasible region) is the union of continuous domains across possible discrete variable values; *i.e.*,

$$\Omega = \bigcup_{x^d \in \Omega^d} (\Omega^c(x^d) \times \{x^d\}),$$

with the convention that $x = x^c$ and $\Omega = \Omega^c$ if $n^d = 0$.

We treat the constraints by the extreme “barrier” approach of applying our class of algorithms, 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 does not belong to Ω , then we set $f_\Omega(x) = \infty$, and f is not evaluated. This is important in many practical engineering problems where f is expensive to evaluate. In fact, this approach is suitable for general set constraints, in which the only output of a constraint evaluation is a binary response to indicate feasibility or not.

The class of MADS algorithms was introduced and analyzed by Audet and Dennis [9], as an extension of generalized pattern search (GPS) algorithms [7, 17, 21] for solving nonlinearly constrained problems, in the absence of categorical variables. Rather than applying a penalty function [18] or filter [8] approach to handle the nonlinear constraints, MADS defines an additional parameter that enables the algorithm to perform an exploration of the variable space in an asymptotically dense set of directions. Under reasonable assumptions, this enables convergence to both first-order [9] and second-order [3] stationary points in the Clarke [12] sense, depending on assumptions made on the smoothness of the objective function.

The paper is divided as follows. Section 2 describes the mixed variable MADS (MV-MADS) algorithm in detail, Section 3 contains new theoretical convergence results for the MV-MADS algorithm, and Section 4 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 $x^c \in \mathbb{R}^{n^c}$ and $\varepsilon > 0$, we denote by $B_\varepsilon(x^c)$ the open ball $\{y \in \mathbb{R}^{n^c} : \|y - x^c\| < \varepsilon\}$.

2 Mixed Variable MADS (MV-MADS)

As in [2, 5, 6], local optimality is defined in terms of local neighborhoods. However, since there is no metric associated with categorical variables, the notion of a local neighborhood must be defined by the user. This is well-defined for continuous variables, but not for categorical variables; special knowledge of the underlying engineering process or physical problem may be the only guide.

We can define a general local neighborhood in terms of a set-valued function $\mathcal{N} : \Omega \rightarrow 2^\Omega$, where 2^Ω denotes the power set (or set of all possible subsets) of Ω . By convention, we assume that for all $x \in \Omega$, the user-defined set $\mathcal{N}(x)$ is finite, and $x \in \mathcal{N}(x)$. As an example, one common choice of neighborhood function for integer variables is the one defined by $\mathcal{N}(x) = \{y \in \Omega : y^c = x^c, \|y^d - x^d\|_1 \leq 1\}$. However, categorical variables may have no inherent metric, which would make this particular choice inapplicable.

With this construction, the classical definition of local optimality is extended to mixed variable domains by the following definition, which is similar to one found in [6].

Definition 1 *A point $x = (x^c, x^d) \in \Omega$ is said to be a local minimizer of f on Ω with respect to the set of neighbors $\mathcal{N}(x) \subset \Omega$ if there exists an $\varepsilon > 0$ such that $f(x) \leq f(v)$ for all v in the set*

$$\Omega \cap \left(\bigcup_{y \in \mathcal{N}(x)} \left(B_\varepsilon(y^c) \times \{y^d\} \right) \right).$$

Each iteration k of a MADS algorithm is characterized by an optional SEARCH step, a local POLL step, and an EXTENDED POLL step, in which f_Ω is evaluated at specified points that lie on a mesh M_k with the goal of finding a feasible *improved mesh point*; *i.e.*, a point $y \in M_k$ for which $f_\Omega(y) < f_\Omega(x_k)$, where $x_k \in \Omega$ is the current iterate or incumbent best iterate found thus far. The mesh is purely conceptual and is never explicitly constructed; instead, mesh points are generated as necessary in the algorithm. Consistent with [5], the mesh is described as follows. For each combination $i = 1, 2, \dots, i_{\max}$ of values that the discrete variables may possibly take, let $D^i = G^i Z^i$ be a set of positive spanning directions [14] (*i.e.*, nonnegative linear combinations of D must span \mathbb{R}^{n^c}), where $G^i \in \mathbb{R}^{n^c \times n^c}$ is a nonsingular generating matrix, and $Z^i \in \mathbb{Z}^{n^c \times |D^i|}$.

The mesh M_k at iteration k is formed as the direct product of Ω^d with the union of a finite number of lattices in Ω^c , *i.e.*,

$$M_k = \bigcup_{i=1}^{i_{\max}} M_k^i \times \Omega^d \quad \text{with} \quad M_k^i = \bigcup_{x \in V_k} \{x^c + \Delta_k^m D^i z : z \in \mathbb{N}^{|D^i|}\} \subset \mathbb{R}^{n^c}, \quad (2)$$

where $\Delta_k^m > 0$ is the mesh size parameter and V_k denotes all previously evaluated trial points at iteration k (V_0 is the set of initial trial points). Furthermore, the neighborhood function must be constructed so that all discrete neighbors lie on the current mesh; *i.e.*, $\mathcal{N}(x_k) \subseteq M_k$ for all k .

The SEARCH step allows evaluation of f_Ω at any finite set of mesh points. Any strategy may be used, including none. The SEARCH step adds nothing to the convergence theory, but well-chosen SEARCH strategies, such as those that make use of surrogates, can greatly improve algorithm performance (see [4, 10, 11, 19]).

The POLL step includes both a traditional MADS [9] POLL with respect to the continuous variables, and an evaluation of the discrete neighbors. In the MADS POLL step, a second mesh parameter $\Delta_k^p > 0$, called the *poll size parameter*, is introduced, which satisfies $\Delta_k^m \leq \Delta_k^p$ for all k , such that

$$\lim_{k \in K} \Delta_k^m = 0 \Leftrightarrow \lim_{k \in K} \Delta_k^p = 0 \text{ for any infinite subset of indices } K. \quad (3)$$

Thus, GPS becomes the specific MADS instance in which $\Delta_k = \Delta_k^p = \Delta_k^m$, where Δ_k is the mesh size parameter using the notation from [7].

At iteration k , let $D_k(x) \subseteq D^{i_0} \subseteq D$ denote the positive spanning set of poll directions for some $x \in V_k$ corresponding to the i_0 -th set of discrete variable values. The set of points generated when polling about x with respect to continuous variables is called a *frame*, and x is called the *frame center*. The formal definition given below (generalized slightly from [9]) allows more flexibility than GPS, which requires D_k to be a subset of the finite set D_k^i .

Definition 2 *At iteration k , the MADS frame is defined to be the set:*

$$P_k(x) = \{(x^c + \Delta_k^m d, x^d) : d \in D_k(x)\} \subset M_k,$$

where $D_k(x)$ 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 to a poll point $x + \Delta_k^m d$ 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 [13]) of the normalized sets $D_k(x)$ are positive spanning sets.

In [9], an instance of MADS, called LTMADS, is presented in which the closure of the cone generated by the set of normalized directions $\bigcup_{k=1}^{\infty} \left\{ \frac{d}{\|d\|} : d \in D_k \right\}$ equals \mathbb{R}^n with probability one. In this case, we say that the set of poll directions is *asymptotically dense* in \mathbb{R}^n with probability one.

Thus the POLL step consists of evaluating points in $P_k(x_k) \cup \mathcal{N}(x_k)$. If the POLL step fails to generate a lower objective function value, then the EXTENDED POLL step is performed around each promising point in $\mathcal{N}(x_k)$ whose objective function value is sufficiently close to the incumbent value. That is, for a fixed positive scalar ξ , if $y \in \mathcal{N}(x_k)$ satisfies $f_{\Omega}(x_k) \leq f_{\Omega}(y) < f_{\Omega}(x_k) + \xi_k$ for some user-specified tolerance value $\xi_k \geq \xi$ (called the *extended poll trigger*), then a finite sequence of POLL steps about the points $\{y_k^j\}_{j=1}^{J_k}$ is performed, beginning with $y_k^0 = y_k \in \mathcal{N}(x_k)$ and ending with $z_k = y_k^{J_k}$. The *extended poll endpoint* z_k occurs when either $f_{\Omega}(z_k^c + \Delta_k^m d, z_k^d) < f_{\Omega}(x_k)$ for some $d \in D_k(z_k)$, or when $f_{\Omega}(x_k) \leq f_{\Omega}(z_k^c + \Delta_k^m d, z_k^d)$ for all $d \in D_k(z_k)$. In Section 3, we make the common assumption that all iterates lie in a compact set, which ensures that J_k is always finite, *i.e.*, that any EXTENDED POLL step generates a finite number of trial points.

The set of EXTENDED POLL points can be expressed as

$$\mathcal{X}_k(\xi_k) = \bigcup_{y_k \in \mathcal{N}_k^{\xi_k}} \bigcup_{j=1}^{J_k} P_k(y_k^j), \quad (4)$$

where $\mathcal{N}_k^{\xi_k} := \{y \in \mathcal{N}(x_k) : f_\Omega(x_k) \leq f_\Omega(y) \leq f_\Omega(x_k) + \xi_k\}$. In practice, the parameter ξ_k is typically set as a percentage of the objective function value (but bounded away from zero), such as $\xi_k = \max\{\xi, 0.05|f(x_k)|\}$. Higher values of ξ_k generate more extended polling, which is more costly, but which may lead to a better local solution since more of the design space is searched.

As soon as any of the three steps is successful in finding an improved mesh point, the iteration ends, the improved mesh point becomes the new current iterate $x_{k+1} \in \Omega$, and the mesh is either retained or coarsened. If no improved mesh point is found, then $P_k(x_k)$ is said to be a *minimal frame* with *minimal frame center* x_k , 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 the same as in [9] and [6]. 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,

$$\Delta_{k+1}^m = \tau^{w_k} \Delta_k^m$$

$$\text{for some } w_k \in \begin{cases} \{0, 1, \dots, w^+\} & \text{if an improved mesh} \\ & \text{point is found} \\ \{w^-, w^- + 1, \dots, -1\} & \text{otherwise.} \end{cases} \quad (5)$$

The class of MV-MADS algorithms is stated formally in Figure 1.

0. INITIALIZATION: Set $\xi > 0$ and $\xi_0 \geq \xi$. Let $x_0 \in \Omega$ such that $f_\Omega(x_0) < \infty$, set $\Delta_0^p \geq \Delta_0^m > 0$. Set iteration counter to $k = 0$.
1. SEARCH step: Evaluate f_Ω on a finite subset of trial points on the mesh M_k (see (2)). If an improved mesh point is found, then the SEARCH may terminate, skip the next POLL step and go directly to Step 4.
2. POLL step: Evaluate f_Ω on the set $P_k(x_k) \cup \mathcal{N}(x_k) \subset M_k$ until an improved mesh point is found or until all points have been exhausted. If an improved mesh point is found, go to Step 4.
3. EXTENDED POLL step: Evaluate f_Ω on the set $\mathcal{X}(\xi_k)$ until an improved mesh point is found or until all points have been exhausted (see (4)).
4. PARAMETER UPDATE: Update Δ_{k+1}^m and Δ_{k+1}^p according to (5) and (3). Update $\xi_k \geq \xi$, increment $k \leftarrow k + 1$, and go to Step 1.

Fig. 1 A general MV-MADS algorithm

3 Convergence Results

In this section, we establish convergence results for the new MV-MADS algorithm. Many of these results will appear very similar to results from either the MADS algorithm [9] or mixed variable pattern search [6]. Before we present our main results, we review some definitions, assumptions, preliminary results, followed by a subsection that extends some Clarke calculus ideas to mixed variables.

3.1 Preliminaries

The convergence analysis relies on the following standard assumptions.

- A1. An initial point x_0 with $f_\Omega(x_0) < \infty$ is available.
A2. All iterates $\{x_k\}$ generated by MV-MADS lie in a compact set.
A3. The set of discrete neighbors $\mathcal{N}(x_k)$ lies on the mesh M_k .

Under these assumptions, the following results are obtained by proofs that are identical to those found in [2, 6] for mixed variable GPS:

- $\liminf_{k \rightarrow +\infty} \Delta_k^p = \liminf_{k \rightarrow +\infty} \Delta_k^m = 0$;
- there exists a *refining subsequence* $\{x_k\}_{k \in K}$ of minimal frame centers for which there are limit points $\hat{x} = \lim_{k \in K} x_k$, $\hat{y} = \lim_{k \in K} y_k$, and $\hat{z} = (\hat{z}^c, \hat{y}^d) = \lim_{k \in K} z_k$, where each $z_k \in \Omega$ is the endpoint of the EXTENDED POLL step initiated at $y_k \in \mathcal{N}(x_k)$, and $\lim_{k \in K} \Delta_k^p = 0$.

The notation used in identifying these limit points will be retained and used throughout the remainder of this paper. Some of the results that follow require the additional assumption that $\hat{y} \in \mathcal{N}(\hat{x})$.

For the main results, the following four definitions [12, 15, 20] are needed. They have been adapted to our context, where only a subset of the variables are continuous. The standard definitions follow when all variables are continuous, *i.e.*, when $n^d = 0$ and $x = x^c$.

Definition 3 A vector $v \in \mathbb{R}^{n^c}$ is said to be a *hypertangent vector to the continuous variables of the set Ω at the point $x = (x^c, x^d) \in \Omega$* if there exists a scalar $\varepsilon > 0$ such that

$$(y + tw, x^d) \in \Omega \text{ for all } y \in B_\varepsilon(x^c) \text{ with } (y, x^d) \in \Omega, w \in B_\varepsilon(v) \text{ and } 0 < t < \varepsilon.$$

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

Definition 4 A vector $v \in \mathbb{R}^{n^c}$ is said to be a *Clarke tangent vector to the continuous variables of the set Ω at the point $x = (x^c, x^d) \in \text{cl}(\Omega)$* if for every sequence $\{y_k\}$ that converges to x^c with $(y_k, x^d) \in \Omega$ 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 $(y_k + t_k w_k, x^d) \in \Omega$. 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 5 A vector $v \in \mathbb{R}^{n^c}$ is said to be a *tangent vector to the continuous variables of the set Ω at the point $x = (x^c, x^d) \in \text{cl}(\Omega)$* if there exists a sequence $\{y_k\}$ that converges to x^c with $(y_k, x^d) \in \Omega$ and a sequence of positive real numbers $\{\lambda_k\}$ for which $v = \lim_k \lambda_k (y_k - x^c)$. The set $T_\Omega^{Co}(x)$ of all *tangent vectors to Ω at x* is called the *contingent cone to Ω at x* .

Definition 6 The set Ω is said to be *regular at x* if $T_\Omega^{Cl}(x) = T_\Omega^{Co}(x)$.

3.2 Extension of the Clarke Calculus to Mixed Variables

For the results of this section, we make use of a generalization [15] of the Clarke [12] directional derivative, in which function evaluations are restricted to points in the domain. Furthermore, we restrict the notions of generalized directional derivatives and gradient to the subspace of continuous variables. The generalized directional derivative

of a locally Lipschitz function f at $x = (x^c, x^d) \in \Omega$ in the direction $v \in \mathbb{R}^{n^c}$ is defined by

$$f^\circ(x; v) := \limsup_{\substack{y \rightarrow x^c, (y, x^d) \in \Omega \\ t \downarrow 0, (y + tv, x^d) \in \Omega}} \frac{f(y + tv, x^d) - f(y, x^d)}{t}. \quad (6)$$

Furthermore, it is shown in [9] that if $T_\Omega^H(x)$ is not empty and $v \in T_\Omega^{Cl}(x)$, then

$$f^\circ(x; v) = \lim_{\substack{u \rightarrow v, \\ u \in T_\Omega^H(x)}} f^\circ(x; u). \quad (7)$$

We similarly generalize other derivative ideas. We denote by $\nabla f(x) \in \mathbb{R}^{n^c}$ and $\partial f(x) \subseteq \mathbb{R}^{n^c}$, respectively, the gradient and generalized gradient of the function f at $x = (x^c, x^d) \in \Omega$ with respect to the continuous variables x^c while holding the categorical variables x^d constant. In particular, the generalized gradient of f at x^* [12] with respect to the continuous variables is defined by

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

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

The final definition, which is adapted from [9] for the mixed variable case, provides some nonsmooth terminology for stationarity.

Definition 7 *Let f be Lipschitz near $x^* \in \Omega$. Then x^* is said to be a Clarke, or contingent, stationary point of f over Ω with respect to the continuous variables if $f^\circ(x^*; v) \geq 0$ for every direction v in the Clarke tangent cone, or contingent cone, to Ω at x^* , respectively.*

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

If $\Omega^c(x^d) = \mathbb{R}^{n^c}$ or x^{*c} lies in the relative interior of $\Omega^c(x^d)$, then a stationary point as described by Definition 7 meets the condition that $f^\circ(x^*; v) \geq 0$ for all $v \in \mathbb{R}^{n^c}$. This is equivalent to $0 \in \partial f(x^*)$.

3.3 Main Results

Our main convergence results consist of four theorems, all of which are generalizations of similar results from MADS [9] or mixed variable pattern search [2, 5, 6]. The first result establishes a notion of directional stationarity at certain limit points, and the second ensures local optimality with respect to the set of discrete neighbors. The remaining two results establish Clarke-based stationarity in a mixed variable sense.

Theorem 8 *Let \hat{w} be the limit point of a refining subsequence or the associated subsequence of EXTENDED POLL endpoints, and let v be a refining direction in the hypertangent cone $T_\Omega^H(\hat{w})$. If f is Lipschitz at \hat{w} with respect to the continuous variables, then $f^\circ(\hat{w}; v) \geq 0$.*

Proof. Let $\{w_k\}_{k \in K}$ be a refining subsequence converging to $\hat{w} = (\hat{w}^c, \hat{w}^d)$. Without any loss of generality, we may assume that $w_k = (w_k^c, \hat{w}^d)$ for all $k \in K$. In accordance with Definition 3.2 in [9], let $v = \lim_{k \in L} \frac{d_k}{\|d_k\|} \in T_{\Omega}^H(\hat{w})$ be a refining direction for \hat{w} , where $d_k \in D_k$ for all $k \in L$ and L is some subset of K .

Since Δ_p^k converges to zero and Definition 2 ensures that $\{\Delta_k^m \|d_k\|\}_{k \in L}$ is bounded above, it follows from (3) that $\{\Delta_k^m \|d_k\|\}_{k \in L}$ must also converge to zero. Thus, it follows from (6) and (7) that

$$\begin{aligned} f^\circ(\hat{w}; v) &= \limsup_{\substack{y \rightarrow \hat{w}^c, (y, w^d) \in \Omega \\ t \downarrow 0, (y + tu, w^d) \in \Omega \\ u \rightarrow v, u \in T_{\Omega}^H(\hat{w})}} \frac{f(y + tu, \hat{w}^d) - f(y, \hat{w}^d)}{t} \\ &\geq \limsup_{k \in L} \frac{f(w_k^c + \Delta_k^m \|d_k\| \frac{d_k}{\|d_k\|}, \hat{w}^d) - f(w_k^c, \hat{w}^d)}{\Delta_k^m \|d_k\|} \\ &= \limsup_{k \in L} \frac{f(w_k + \Delta_k^m d_k, \hat{w}^d) - f(w_k)}{\Delta_k^m \|d_k\|} \geq 0. \end{aligned}$$

The last inequality holds because $(w_k + \Delta_k^m d_k, \hat{w}^d) \in \Omega$ and $f(w_k + \Delta_k^m d_k, \hat{w}^d) \geq f(w_k)$ (since w_k is a minimal frame center) for all sufficiently large $k \in L$. ■

The next result gives sufficient conditions under which \hat{x} is a local minimizer with respect to its discrete neighbors.

Theorem 9 *If f is lower semi-continuous at \hat{x} and upper semi-continuous at $\hat{y} \in \mathcal{N}(\hat{x})$ with respect to the continuous variables, then $f(\hat{x}) \leq f(\hat{y})$.*

Proof. Since $k \in K$ ensures that $\{x_k\}_{k \in K}$ are minimal frame centers, we have $f(x_k) \leq f(y_k)$ for all $k \in K$. By the assumptions of lower and upper semi-continuity on f and the definitions of \hat{x} and \hat{y} , we have $f(\hat{x}) \leq \lim_{k \in K} f(x_k) \leq \lim_{k \in K} f(y_k) = f(\hat{y})$. ■

The next theorem lists conditions that ensure that \hat{x} satisfies certain stationary conditions, under various smoothness requirements.

Theorem 10 *Assume that $T_{\Omega}^H(\hat{x}) \neq \emptyset$ and the set of refining directions is asymptotically dense in $T_{\Omega}^H(\hat{x})$.*

1. *If f is Lipschitz near \hat{x} with respect to the continuous variables, then \hat{x} is a Clarke stationary point of f on Ω with respect to the continuous variables.*
2. *If f is strictly differentiable at \hat{x} with respect to the continuous variables, then \hat{x} is a Clarke KKT stationary point of f on Ω with respect to the continuous variables.*

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

1. *If f is Lipschitz near \hat{x} with respect to the continuous variables, then \hat{x} is a contingent stationary point of f on Ω with respect to the continuous variables.*
2. *If f is strictly differentiable at \hat{x} with respect to the continuous variables, then \hat{x} is a contingent KKT stationary point of f on Ω .*

Proof. First, Rockafellar [20] showed that if the hypertangent cone is not empty at \hat{x} , then $T_{\Omega}^{Cl}(\hat{x}) = \text{cl}(T_{\Omega}^H(\hat{x}))$. Since the set S of refining directions for f at \hat{x} is a dense subset of $T_{\Omega}^H(\hat{x})$, S is also a dense subset of $T_{\Omega}^{Cl}(\hat{x})$. Thus, any vector $v \in T_{\Omega}^{Cl}(\hat{x})$ can be expressed as the limit of directions in S , and the first result follows directly from (7) and Theorem 8.

Strict differentiability ensures the existence of $\nabla f(\hat{x})$ and that $\nabla f(\hat{x})^T v = f^{\circ}(\hat{x}; v)$ for all $v \in T_{\Omega}^{Cl}(\hat{x})$. Since $f^{\circ}(\hat{x}; v) \geq 0$ for all $v \in T_{\Omega}^{Cl}(\hat{x})$, we have $(-\nabla f(\hat{x}))^T v \leq 0$, and the second result follows from Definition 7.

Furthermore, if Ω is regular at \hat{x} , then by Definition 6, $T_{\Omega}^{Cl}(\hat{x}) = T_{\Omega}^{Co}(\hat{x})$, and the final two results follow directly from Definition 7. ■

The next result is similar to Theorem 10 but considers the limit of EXTENDED POLL endpoints \hat{z} instead of \hat{x} .

Theorem 11 *Assume that $\hat{y} \in \mathcal{N}(\hat{x})$, $T_{\Omega}^H(\hat{z}) \neq \emptyset$, and the set of refining directions is asymptotically dense in $T_{\Omega}^H(\hat{z})$.*

1. *If f is Lipschitz near \hat{z} with respect to the continuous variables, then \hat{z} is a Clarke stationary point of f on Ω with respect to the continuous variables.*
2. *If f is strictly differentiable at \hat{z} with respect to the continuous variables, then \hat{z} is a Clarke KKT stationary point of f on Ω with respect to the continuous variables.*

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

1. *If f is Lipschitz near \hat{z} with respect to the continuous variables, then \hat{z} is a contingent stationary point of f on Ω with respect to the continuous variables.*
2. *If f is strictly differentiable at \hat{z} with respect to the continuous variables, then \hat{z} is a contingent KKT stationary point of f on Ω .*

Proof. The proof is identical to that of Theorem 10, but with \hat{z} replacing \hat{x} . ■

We have shown conditions under which the limit points \hat{x} and \hat{z} satisfy certain necessary conditions for optimality. We now tie these results together with the notion of local optimality given in Definition 1.

Remark 12 *If $\hat{y} \in \mathcal{N}(\hat{x})$ is the limit point of discrete neighbors $\{y_k\}_{k \in K}$, where $y_k \in \mathcal{N}(x_k)$ for $k \in K$, then Theorem 9 ensures that $f(\hat{x}) \leq f(\hat{y})$. If this inequality is strict, then \hat{x} is locally optimal (i.e., Definition 1 is satisfied) with respect to \hat{y} .*

On the other hand, suppose $f(\hat{x}) = f(\hat{y})$. Then, since the EXTENDED POLL triggers are bounded away from zero, an EXTENDED POLL step must have been performed around y_k for infinitely many $k \in K$. Since $f(x_k) \leq f(z_k) \leq f(y_k)$ for all $k \in K$, it follows that $f(\hat{x}) = f(\hat{z}) = f(\hat{y})$, and Theorem 10 ensures first-order stationarity at \hat{z} with respect to the continuous variables.

4 Concluding Remarks

This paper fills an important gap in the convergence theory for the class of MADS algorithms. We have introduced a new class of MADS algorithms for mixed variable optimization problems and proved that it possesses appropriate convergence properties, which are consistent with previous results for less general algorithms. We hope that these results will serve as a springboard for extending MADS to other classes of problems, such as those with stochastic noise and multiple objectives.

Acknowledgments

Work of the first author was supported by the Air Force Office of Scientific Research (AFOSR) and Los Alamos National Laboratory. Work of the second author was supported by AFOSR, ExxonMobil Upstream Research Company, and NSERC.

The views expressed in this paper are those of the authors and do not reflect the official policy or position of the U.S. Air Force, Department of Defense, or U.S. Government.

References

1. K. Abhishekand, S. Leyffer, and J. T. Linderoth. Modeling without categorical variables: A mixed-integer nonlinear program for the optimization of thermal insulation systems. Technical Report Preprint ANL/MCS-P1434-0607, Argonne National Laboratory, 2007.
2. M. A. Abramson. *Pattern Search Algorithms for Mixed Variable General Constrained Optimization Problems*. PhD thesis, Department of Computational and Applied Mathematics, Rice University, August 2002.
3. M. A. Abramson and C. Audet. Second-order convergence of mesh adaptive direct search. *SIAM Journal on Optimization*, 17(2):606–619, 2006.
4. M. A. Abramson, C. Audet, and J. E. Dennis, Jr. Generalized pattern searches with derivative information. *Math. Programming, Series B*, 100(1):3–25, 2004.
5. M. A. Abramson, C. Audet, and J. E. Dennis, Jr. Filter pattern search algorithms for mixed variable constrained optimization problems. *Pacific Journal of Optimization*, to appear.
6. C. Audet and J. E. Dennis, Jr. Pattern search algorithms for mixed variable programming. *SIAM Journal on Optimization*, 11(3):573–594, 2000.
7. C. Audet and J. E. Dennis, Jr. Analysis of generalized pattern searches. *SIAM Journal on Optimization*, 13(3):889–903, 2003.
8. C. Audet and J. E. Dennis, Jr. A pattern search filter method for nonlinear programming without derivatives. *SIAM Journal on Optimization*, 14(4):980–1010, 2004.
9. C. Audet and J. E. Dennis, Jr. Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on Optimization*, 17(2):188–217, 2006.
10. C. Audet and D. Orban. Finding optimal algorithmic parameters using the mesh adaptive direct search algorithm. *SIAM Journal on Optimization*, 17(3):642–664, 2006.
11. 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. *Struct. Optim.*, 17(1):1–13, February 1999.
12. 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.
13. I. D. Coope and C. J. Price. Frame-based methods for unconstrained optimization. *J. Optim. Theory Appl.*, 107(2):261–274, 2000.
14. C. Davis. Theory of positive linear dependence. *Amer. J. Math.*, 76(4):733–746, 1954.
15. J. Jahn. *Introduction to the Theory of Nonlinear Optimization*. Springer, Berlin, 1994.
16. M. Kokkolaras, C. Audet, and J. E. Dennis, Jr. Mixed variable optimization of the number and composition of heat intercepts in a thermal insulation system. *Optimization and Engineering*, 2(1):5–29, 2001.
17. T. G. Kolda, R. M. Lewis, and V. Torczon. Optimization by direct search: new perspectives on some classical and modern methods. *SIAM Review*, 45(3):385–482, 2003.
18. R. M. Lewis and V. Torczon. A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Optimization*, 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 Journal on Optimization*, 7(1):1–25, February 1997.