

# Second-Order Convergence of Mesh-Adaptive Direct Search

Mark A. Abramson

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>

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

Charles.Audet@gerad.ca, <http://www.gerad.ca/Charles.Audet>

August 17, 2005

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

**Abstract:** A previous analysis of second-order behavior of 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.

---

\*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.

# 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,

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

with  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  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  $\psi_\Omega$  is the indicator function for  $\Omega$ ; *i.e.*, it is zero on  $\Omega$ , and infinity elsewhere. If a point  $x$  is not in  $\Omega$ , 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 algorithms GPS [3, 22] 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, in which positive spanning directions are chosen in a random fashion, and almost sure first-order convergence is obtained in [4]. A similar first-order analysis is done in [14] for the DIRECT algorithm.

This paper extends the MADS analysis to show important and surprisingly strong second-order convergence 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 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.

The paper is outlined as follows. The MADS algorithm is briefly described in Section 2, with first-order convergence properties restated in Section 3. Section 4 introduces the generalized Hessian [15] with some associated properties, followed by necessary and sufficient second-order optimality conditions and convergence results. Illustrative examples are also presented. Section 5 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 [15] 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  $f_\Omega$  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 [13] (*i.e.*, nonnegative linear combinations of  $D$  must span  $\mathbb{R}^n$ ), and each direction  $d_j \in D$ ,  $j = 1, 2, \dots, n_D$  must be constructed as the product  $G\bar{z}_j$ , where  $G \in \mathbb{R}^{n \times n}$  is a nonsingular generating matrix, and  $\bar{z}_j \in \mathbb{Z}^n$  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  $k$ .

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  $f_\Omega$  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]).

If the SEARCH step fails to find an improved mesh point, the POLL step is invoked, in which  $f_\Omega$  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  $\Delta_k^p$ , called the *poll size parameter*, which satisfies  $\Delta_k^m \leq \Delta_k^p$  for all  $k$ , and also

$$\lim_{k \in K} \Delta_k^m = 0 \Leftrightarrow \lim_{k \in K} \Delta_k^p = 0 \text{ for any infinite subset of indices } K. \quad (2)$$

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  $\mathbb{R}^n$ .

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  $\Delta_k^m$  is updated according to the rule,

$$\begin{aligned} \Delta_{k+1}^m &= \tau^{w_k} \Delta_k^m \\ \text{for some } w_k &\in \begin{cases} \{0, 1, \dots, w^+\} & \text{if an improved mesh point is found} \\ \{w^-, w^- + 1, \dots, -1\} & \text{otherwise.} \end{cases} \end{aligned} \quad (3)$$

The class of MADS algorithms is stated formally as follows:

A GENERAL MADS ALGORITHM

- **INITIALIZATION:** Let  $x_0 \in \Omega$ ,  $\Delta_0^m \leq \Delta_0^p$ ,  $G$ ,  $Z$ ,  $\tau$ ,  $w^-$  and  $w^+$  satisfy the requirements given above. Set the iteration counter  $k$  to 0.
- **SEARCH AND POLL STEP:** Perform the SEARCH and possibly the POLL steps (or only part of them) until an improved mesh point  $x_{k+1}$  is found on the mesh  $M_k$  (see Definition 2.1).
  - **OPTIONAL SEARCH:** Evaluate  $f_\Omega$  on a finite subset of trial points on the mesh  $M_k$ .
  - **LOCAL POLL:** Evaluate  $f_\Omega$  on the frame  $P_k$  (see Definition 2.2).
- **PARAMETER UPDATE:** Update  $\Delta_{k+1}^m$  according to equation (3), and  $\Delta_{k+1}^p$  according to (2). Increase  $k \leftarrow k + 1$  and go back to the SEARCH and POLL step.

### 3 Existing First-Order Convergence Theory

Before presenting new results, it is important to state what is currently known about the convergence properties of MADS. The results given in this section are taken from [4] and only rely on the mild assumptions that a feasible initial point  $x_0$  is provided, and has a finite objective function value, and that all iterates  $\{x_k\}$  lie in a compact set. First some definitions are restated, and then the main results are given.

#### 3.1 Refining Subsequences

The following proposition is based on one by Torczon [22] for pattern search methods. It has been reproved for other cases in several papers (*e.g.*, see [4, 5, 17]).

**Proposition 3.1** *The poll and mesh size parameters produced by a MADS algorithm satisfy*

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

Proposition 3.1 ensures the existence of infinitely many minimal frame centers, since  $\Delta_k^m$  only shrinks when a minimal frame is found. The following definition, taken from [4], is needed for later results.

**Definition 3.2** *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.*

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 was proved for MADS in [4].

## 3.2 Cones and Generalized Derivatives

The following four definitions [9, 16, 21] are needed because they are used in the statements of the main theorems.

**Definition 3.3** 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  $\epsilon > 0$  such that

$$y + tw \in \Omega \quad \text{for all } y \in \Omega \cap B_\epsilon(x), \quad w \in B_\epsilon(v) \quad \text{and} \quad 0 < t < \epsilon. \quad (4)$$

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

**Definition 3.4** 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 the closure of  $\Omega$  if for every sequence  $\{y_k\}$  of elements of  $\Omega$  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  $\Omega$  at  $x$  is called the Clarke tangent cone to  $\Omega$  at  $x$ .

**Definition 3.5** 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 the closure of  $\Omega$  if there exists a sequence  $\{y_k\}$  of elements of  $\Omega$  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  $\Omega$  at  $x$  is called the contingent cone (or sequential Bouligand tangent cone) to  $\Omega$  at  $x$ .

**Definition 3.6** The set  $\Omega$  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 [21] showed that if  $T_{\Omega}^H(x)$  is nonempty, then  $T_{\Omega}^H(x) = \text{int}(T_{\Omega}^{Cl}(x))$  and therefore  $T_{\Omega}^{Cl}(x) = \text{cl}(T_{\Omega}^H(x))$  [20].

In order to establish the results of this section, we apply a generalization of the Clarke [9] directional derivative, as presented in [16], 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

$$f^{\circ}(x; v) := \limsup_{\substack{y \rightarrow x, y \in \Omega \\ t \downarrow 0, y + tv \in \Omega}} \frac{f(y + tv) - f(y)}{t}. \quad (5)$$

If  $T_{\Omega}^H(x) \neq \emptyset$  and  $v \in T_{\Omega}^{Cl}(x)$ , then an equivalent definition [5] is given by

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

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

**Definition 3.7** *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  $\Omega$ , if  $f^{\circ}(\hat{x}; v) \geq 0$  for every direction  $v$  in the Clarke, or contingent cone of  $\Omega$  at  $\hat{x}$ , respectively.*

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

If  $\Omega = \mathbb{R}^n$  or  $\hat{x}$  lies in the interior of  $\Omega$ , then a stationary point as described by Definition 3.7 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  $\hat{x}$ , which is defined by

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

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

### 3.3 First-Order Results

We now restate the main results from [4]. Theorem 3.9 is a restatement of four different theorems found in [4], and is the main result of this section. It requires Theorem 3.8.

**Theorem 3.8** *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 3.9** *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  $\Omega$ .*
2. *If  $f$  is strictly differentiable at  $\hat{x}$ , then  $\hat{x}$  is a Clarke KKT stationary point of  $f$  on  $\Omega$ .*

*Furthermore, if  $\Omega$  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  $\Omega$ .*
2. *If  $f$  is strictly differentiable at  $\hat{x}$ , then  $\hat{x}$  is a contingent KKT stationary point of  $f$  on  $\Omega$ .*

## 4 New Second-Order Convergence Theory

This section contains a second-order convergence theory for MADS. Section 4.1, recalls the generalized Hessian and identifies some useful properties, including a second-order Taylor series representation of non-smooth functions. Section 4.2 presents second-order necessary and sufficient conditions for optimality for set-constrained optimization. Finally, in Section 4.3, conditions are established, 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 second-order convergence, we need nonsmooth notions of second derivatives and a description of second-order optimality conditions. Generalized second-order directional derivatives are developed in [10] and [15], consistent with the Clarke [9] calculus for first order derivatives. In this paper, we apply the Hirriart-Urruty *et al.* [15] definition of a generalized Hessian, given as follows.

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

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

By construction,  $\partial^2 g(x)$  is a nonempty, compact, and convex set of symmetric matrices [15]. 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 [15], 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  $x_k \rightarrow x$  and  $A_k \rightarrow A$  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 [15].

**Theorem 4.2** *Let  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  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*

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

We should note that this result applies to any point in an open set in  $\mathbb{R}^n$ , and we wish to apply it to a point in  $\Omega$  that may lie on the boundary. However, in our results given in the next section, we make certain local smoothness assumptions on  $f$  that are independent of  $\Omega$ . Thus, we can apply this result to any direction in  $\mathbb{R}^n$ , and still get the requisite matrix  $A_x$ .

## 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 [15] 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  $\epsilon \geq 0$  for which  $x + tv \in \Omega$  for all  $0 \leq t < \epsilon$ . The set of feasible directions for  $\Omega$  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  $\Omega$  is regular at  $x$ , then  $\text{cl}(T_\Omega^H(x)) = \text{cl}(T_\Omega^F(x)) = T_\Omega^{Cl}(x) = T_\Omega^{Co}(x)$ . 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  $\Omega = \mathbb{R}^2 \setminus \{(-\frac{1}{k}, b) \in \mathbb{R}^2 : b \in \mathbb{R}, k = 1, 2, \dots\}$ , 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). 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

$$0 \leq f(x_k) - f(x^*) = \frac{1}{k} \nabla f(x^*)^T v + \frac{1}{2k^2} v^T A_k v = \frac{1}{2k^2} v^T A_k v, \quad (6)$$

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  $\bar{x}_k \rightarrow x^*$ , 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 (6) 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 or contingent cone, as the following example shows.

**Example 4.5** Consider the quadratic optimization problem, in which  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  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

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

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), and suppose that  $T_{\Omega}^H(x^*) \neq \emptyset$  and that  $\Omega$  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).

**Proof.** Let  $x^* \in \Omega$  be a contingent stationary point for the optimization problem defined in (1), and suppose that  $T_{\Omega}^H(x^*) \neq \emptyset$  and that  $\Omega$  is convex near  $x^*$ . Suppose also that  $f$  is continuously differentiable with Lipschitz derivatives near  $x^*$ .

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  $\Omega$  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

$$\begin{aligned} 0 \geq f(y_k) - f(x^*) &= (y_k - x^*)^T \nabla f(x^*) + \frac{1}{2} (y_k - x^*)^T A_k (y_k - x^*) \\ &\geq \frac{1}{2} (y_k - x^*)^T A_k (y_k - x^*). \end{aligned} \quad (7)$$

Now, since  $\bar{x}_k \rightarrow x^*$ , 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 (7) 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  $v^T \nabla f(x^*) = 0$ .  $\blacksquare$

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

**Example 4.7** Consider the quadratic optimization problem, in which  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  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  $\Omega$  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,  $(\epsilon, -\epsilon^2)$  belongs to the strict interior of  $\Omega$  for all  $\epsilon \neq 0$ , and  $f(\epsilon, -\epsilon^2) = -\epsilon^2 < 0 = f(x)$ .

### 4.3 Second-Order Convergence 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  $\Omega$  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.** Under the hypotheses of the Theorem, 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  $v^T Av < 0$  for all  $A \in \partial^2 f(x)$  and for all  $x$  in the open ball  $B_\varepsilon(\hat{x})$  of radius  $\varepsilon$  centered at  $\hat{x}$ .

Let  $K$  denote the set of indices of unsuccessful iterations. Regularity of  $\Omega$ , together with the assumption that  $T_\Omega^H(\hat{x}) \neq \emptyset$  guarantee that  $\text{cl}(T_\Omega^H(\hat{x})) = T_\Omega^{Co}(\hat{x}) = \text{cl}(T_\Omega^F(\hat{x}))$ . 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

$$f(x_k + \Delta_k^p d_k) - f(x_k) = \Delta_k^p d_k^T \nabla f(x_k) + \frac{1}{2} (\Delta_k^p)^2 d_k^T A_k^+ d_k, \quad (8)$$

$$f(x_k - \Delta_k^p d_k) - f(x_k) = -\Delta_k^p d_k^T \nabla f(x_k) + \frac{1}{2} (\Delta_k^p)^2 d_k^T A_k^- d_k \quad (9)$$

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  $\Delta_k \rightarrow 0^+$  and  $x_k \rightarrow \hat{x}$ , 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 (8) and (9) and substituting  $d_k = \|d_k\|w_k$  yields

$$\frac{1}{\Delta_k^p \|d_k\|} \left[ \frac{f(x_k + \Delta_k^p \|d_k\|w_k) - f(x_k)}{\Delta_k^p \|d_k\|} + \frac{f(x_k - \Delta_k^p \|d_k\|w_k) - f(x_k)}{\Delta_k^p \|d_k\|} \right] = w_k^T A_k w_k, \quad (10)$$

where  $A_k = \frac{1}{2}(A_k^+ + A_k^-)$ . Furthermore, since  $w_k \rightarrow v$  and  $v^T Av < 0$ , 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 (10) 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 (10) (for all sufficiently large  $k \in K$ ). Thus it must be the case that

$$\frac{f(x_k - \Delta_k^p \|d_k\|w_k) - f(x_k)}{\Delta_k^p \|d_k\|} \leq \gamma < 0 \quad (11)$$

for all sufficiently large  $k \in K$ . Taking the limit of (11) as  $k \rightarrow \infty$  in  $K$  yields  $\nabla f(\hat{x})^T(-v) < 0$ , or  $\nabla f(\hat{x})^T v > 0$ , which contradicts the assumption that  $\nabla f(\hat{x})^T v = 0$ . ■

We should note that, since the proof of Theorem 4.8 uses a second-order central difference formula, it might appear to the reader that the result depends on the usage of  $2n$  positive spanning directions at each iteration. However, this is not the case. In the proof, while an argument is made about function values at  $x_k - \Delta_k^p d_k$ , the function  $f$  is never actually evaluated there. This is an important clarification because it is entirely possible that  $x_k - \Delta_k^p d_k$  is never feasible, meaning that  $f$  is never evaluated there.

The following result shows that the sufficiency conditions of Theorem 4.6 can be satisfied by a subsequence of MADS iterates. As expected, the hypotheses are stronger 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  $T_\Omega^H(\hat{x}) \neq \emptyset$ ,  $\Omega$  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  $\Omega$ .*

**Proof.** Since  $f$  is 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 3.9 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. ■

Second order results for GPS are presented in [1]. They are not as strong as those presented here for MADS. For example, GPS with the standard coordinate directions applied to the unconstrained objective function  $f(a, b) = -(ab)^2$  with  $x_0^T = (0, 0)$  stalls at  $x_k^T = (0, 0)$  for every  $k$ , which is the global maximizer of  $f$ . It is shown in [1] that this behavior may only occur when a local maximizer is reached in a finite number of iterations, and that the feasible poll directions are directions where  $f$  is constant. On the other hand, Theorem 4.8 ensures that this behavior is not possible with MADS if the poll directions are diverse enough.

The following unconstrained example shows that the GPS iterates may converge in an infinite number of iterations to a saddle point. MADS cannot converge to that saddle point.

**Example 4.10** *Consider an instance of GPS where  $D_k = D$  is constant throughout all iterations, and is composed of  $2n$  directions, namely plus and minus the standard coordinate directions  $e_1$  and  $e_2$ . 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 parameter is kept constant. On other successful iterations, the mesh size is multiplied by 8. The GPS parameters are  $G = I$  (the identity matrix),  $Z = D = [I; -I]$ ,  $\tau = 2$ ,  $w^- = -4$  and  $w^+ = 3$ .*

*We consider 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.

Consider the unconstrained quadratic optimization problem in which the polynomial objective function in  $\mathbb{R}^2$  is  $f(a, b) = a^2 + b^2 + 3ab$ . The point  $(0, 0)$  is a saddle point, at which the descent directions lie in the cone generated by  $a = \frac{1}{2}b(-3 \pm \sqrt{5})$ . The initial parameters are  $x_0^T = (1, 1)$  with  $f(x_0) = 5$  and  $\Delta_0 = 8$ . Figure 1 displays the first iterates generated by the algorithm. The figure also displays some level sets of  $f$ .

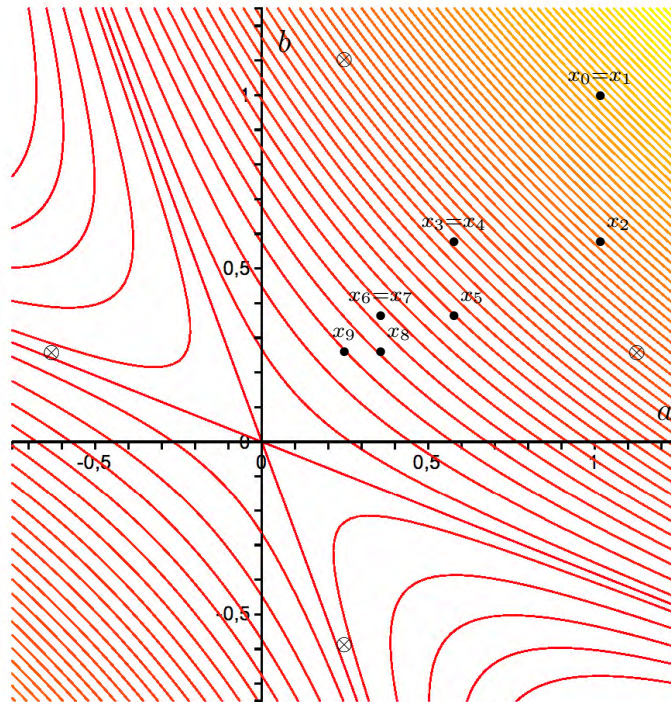


Figure 1: Initial iterates of GPS

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 4.11** For any integer  $\ell \geq 0$ , the GPS iterates are such that  $x_{3\ell} = x_{3\ell+1} = (2^{-\ell}, 2^{-\ell})$ ,  $x_{3\ell+2} = (2^{-\ell}, 2^{-\ell-1})$ , 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  $x_k = (2^{-\ell}, 2^{-\ell})$ . The current objective function value is  $f(x_k) = 5 \times 4^{-\ell}$ . Table 1 details the objective function values at the poll points for iterations  $k = 3\ell, 3\ell + 1$  and  $3\ell + 2$ . The trial points that improve the incumbent appear in shaded boxes.

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

Table 1: Iterations  $k = 3\ell$  to  $k = 3(\ell + 1)$

This table shows that the iterate 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 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 that, under conditions weaker than standard Newton assumptions: our necessity result only requires that  $f$  be continuously differentiable with Lipschitz derivatives near a limit point, and twice strict differentiability of  $f$  is required to ensure that MADS generates a subsequence converging to a local minimizer for problem (1).

We should point out that these results are obtained by making the assumption that the set of directions used by the MADS algorithm is asymptotically dense in the hypertangent cone. In practice, this is not possible because, as the algorithm converges, it would require increasingly more poll directions; in fact, an infinite number in the limit.

To get around this problem, an implementable instance of MADS is introduced in [4], 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 easy to implement, the drawback is that asymptotic denseness of the directions is achieved *almost surely* (i.e., with probability one). Thus the convergence results both in [4] and in this paper are also achieved *almost surely*. This is a weaker measure of convergence, but it works well in practice. 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.

## References

- [1] M. A. Abramson. Second order behavior of pattern search. Technical Report TR04-03, Department of Computational and Applied Mathematics, Rice University, Houston Texas, 2004. To appear in *SIAM J. Optim.*
- [2] M. A. Abramson, C. Audet, and J. E. Dennis, Jr. Generalized pattern searches with derivative information. *Mathematical Programming, Series B*, 100:3–25, 2004.
- [3] C. Audet and J. E. Dennis, Jr. Analysis of generalized pattern searches. *SIAM Journal on Optimization*, 13(3):889–903, 2003.
- [4] C. Audet and J. E. Dennis, Jr. Mesh adaptive direct search algorithms for constrained optimization. Technical Report G–2004–04, Les Cahiers du GERAD, Montréal, 2004.
- [5] 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.
- [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 SIOPT.
- [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 Optimization*, 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 Journal on Control and Optimization*, 28(4):789–809, 1990.
- [11] I. D. Coope and C. J. Price. Frame-based methods for unconstrained optimization. *Journal of Optimization Theory and Applications*, 107(2):261–274, 2000.
- [12] I. D. Coope and C. J. Price. On the convergence of grid-based methods for unconstrained optimization. *SIAM Journal on Optimization*, 11(4):859–869, 2001.



- [13] C. Davis. Theory of positive linear dependence. *American Journal of Mathematics*, 76:733–746, 1954.
- [14] D.E. Finkel and C.T. Kelley. Convergence analysis of the direct algorithm. Technical report, NCSU Mathematics Department, 2004.
- [15] 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.
- [16] J. Jahn. *Introduction to the Theory of Nonlinear Optimization*. Springer, Berlin, 1994.
- [17] R. M. Lewis and V. Torczon. Pattern search methods for linearly constrained minimization. *SIAM Journal on Optimization*, 10(3):917–941, 2000.
- [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. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.
- [21] R. T. Rockafellar. Generalized directional derivatives and subgradients of nonconvex functions. *Canad. J. Math.*, 32(2):257–280, 1980.
- [22] V. Torczon. On the convergence of pattern search algorithms. *SIAM Journal on Optimization*, 7(1):1–25, February 1997.