

Analysis of Direct Searches for Discontinuous Functions

L. N. Vicente* A. L. Custódio†

November 24, 2010

Abstract

It is known that the Clarke generalized directional derivative is nonnegative along the limit directions generated by directional direct-search methods at a limit point of certain subsequences of unsuccessful iterates, if the function being minimized is Lipschitz continuous near the limit point.

In this paper we generalize this result for discontinuous functions using Rockafellar generalized directional derivatives (upper subderivatives). We show that Rockafellar derivatives are also nonnegative along the limit directions of those subsequences of unsuccessful iterates when the function values converge to the function value at the limit point. This result is obtained assuming that the function is directionally Lipschitz with respect to the limit direction.

It is also possible under appropriate conditions to establish more insightful results by showing that the sequence of points generated by these methods eventually approaches the limit point along the locally best branch or step function (when the number of steps is equal to two).

The results of this paper are presented for constrained optimization and illustrated numerically.

*CMUC, Department of Mathematics, University of Coimbra, 3001-454 Coimbra, Portugal (lnv@mat.uc.pt). Support for this research was provided by FCT under the grants PTDC/MAT/64838/2006 and PTDC/MAT/098214/2008. Part of this work was developed while this author was visiting the Courant Institute of Mathematical Sciences of New York University under a FLAD scholarship.

†Department of Mathematics, FCT-UNL, Quinta da Torre, 2829-516 Caparica, Portugal (alcustodio@fct.unl.pt). Support for this author was also provided by Centro de Matemática e Aplicações da Universidade Nova de Lisboa.

Keywords: Direct-search methods, discontinuity, directionally Lipschitz, lower semicontinuity, generalized directional derivatives, nonsmooth calculus, Lipschitz extensions.

1 Introduction

In this paper we consider a constrained minimization problem posed as

$$\begin{aligned} \min \quad & f(x), \\ \text{s.t.} \quad & x \in \Omega, \end{aligned} \tag{1}$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ represents a nonsmooth, possibly discontinuous, extended-real-valued objective function and $\Omega \subseteq \mathbb{R}^n$ denotes a nonempty constrained or feasible region. Our interest relies on the solution of problem (1) by derivative-free methods, i.e., by methods which make no use of derivatives of the objective function (or of the functions defining the feasible region), and in particular by direct-search methods (DSM).

We are particularly interested in the situation where the objective function is discontinuous at a limit point of the sequence of iterates generated by a direct-search method. Problems where the objective function is discontinuous (and no access to derivatives is available) appear in several applications. One example is the Omega function [8] which measures the performance of an asset, or of a portfolio of assets, by the ratio of the weighted gains (above a given threshold) over the weighted losses (below the threshold). Such function exhibits numerous discontinuities and recent studies involving its derivative-free optimization are reported in [15, 18]. Another example arises in the tuning of algorithmic parameters for a given method/code (see [6] for instances where DSM have been applied to solve such problems) — the resulting objective functions are likely to exhibit all sorts of discontinuities given the way that typically a method/code responds to changes in its parameters. DSM have also been used for automatic error analysis [12, 13], a process in which the computer is used to analyze the accuracy or stability of a numerical computation (and examples have been provided where the objective function is discontinuous). Many engineering design problems (which are likely to form the core of the mainstream applications of derivative-free optimization) lead to objective functions involving discontinuities and limited or no access to derivatives (one such application in aircraft design which was recently drawn to our attention is reported in [3]).

DSM can be classified as either directional or simplicial [10, Chapters 7 and 8]. In this paper we are interested in directional DSM and will consider their iterations organized around a search step (optional) and a poll step.

We will essentially concentrate on the poll step since it is responsible for the global convergence properties of the resulting algorithm. A poll step consists of evaluating the objective function at a set of points defined by a positive spanning set or, in some methods different from the ones studied in our paper, defined by a set of positive generators of a cone related to the constraints. A successful poll step occurs when at least one poll point exhibits a function value lower (or sufficiently lower) than the current one (infeasible points are considered to have an infinite objective function value).

A number of directional DSM consider a finite number of such sets of directions and are referred to as pattern search or generalized pattern search. Although some of our basic results apply to these methods, we will focus on those directional DSM, like mesh adaptive direct search (MADS) [5] and generating set search (GSS) [16] (under sufficient decrease), which are entitled to use an infinite number of poll directions. A byproduct of our investigation will actually extend the analysis of GSS type methods to the nonsmooth case (which, to the best of our knowledge, cannot be found in the literature even for the continuous case, although there exist results once a specific form of nondifferentiability is assumed like those in [7]).

It is possible to prove for these classes of methods the existence of a subsequence of unsuccessful iterates (i.e., unsuccessful poll steps) converging to a limit point and driving the step size parameter to zero (this parameter is also called the mesh size parameter and it basically controls the displacement along a direction). At these refining subsequences one can consider limits of normalized poll directions which are then called refining directions. Audet and Dennis [5] proved that if the objective function is Lipschitz continuous near the limit point x_* , then the Clarke-Jahn directional derivative is nonnegative along an appropriate refining direction v :

$$f_C^\circ(x_*; v) = \limsup_{\substack{x' \rightarrow x_*, x' \in \Omega \\ t \downarrow 0, x' + tv \in \Omega}} \frac{f(x' + tv) - f(x')}{t} \geq 0. \quad (2)$$

This derivative is essentially the Clarke generalized directional derivative [9] extended by Jahn [14] to the constrained setting. The refining direction must belong to the hypertangent cone to Ω at x_* , represented by $H_\Omega(x_*)$. If the corresponding set of refining directions for x_* is dense in the unit sphere, then these derivatives are proved to be nonnegative for all directions in the tangent cone to Ω at x_* (this cone is represented by $T_\Omega(x_*)$ and is the closure of the hypertangent cone $H_\Omega(x_*)$). A similar result had already been proved for unconstrained optimization and generalized pattern search [4].

To our knowledge, not too much is known about the behavior of DSM

when applied to discontinuous functions, besides the fact that any limit of a refining subsequence is the limit of unsuccessful iterations (*poll centers*) on meshes or lattices that get infinitely fine (see [5]).

In this paper we will show that the nonnegativity of generalized derivatives along refining directions can be extended to the Rockafellar upper subderivative [20] (generalized by us to the constrained case),

$$f^\uparrow(x_*; v) = \limsup_{\substack{x' \rightarrow_f x_*, x' \in \Omega \\ t \downarrow 0}} \inf_{\substack{v' \rightarrow v \\ x' + tv' \in \Omega}} \frac{f(x' + tv') - f(x')}{t} \geq 0, \quad (3)$$

whenever, at the point x_* , the function f is lower semicontinuous and directionally Lipschitz in Ω with respect to a direction v belonging to the hypertangent cone $H_\Omega(x_*)$. Such a result applies to functions discontinuous at x_* . The notation $x' \rightarrow_f x$ represents $x' \rightarrow x$ and $f(x') \rightarrow f(x)$. The function f is said to be directionally Lipschitz at x with respect to $v \in H_\Omega(x)$ if

$$f^+(x; v) = \limsup_{\substack{x' \rightarrow_f x, x' \in \Omega \\ t \downarrow 0}} \sup_{\substack{v' \rightarrow v \\ x' + tv' \in \Omega}} \frac{f(x' + tv') - f(x')}{t} < +\infty.$$

Examples of directionally Lipschitz functions are given in [20, Section 6]. In Section 5 we describe several functions discontinuous at a point that are directionally Lipschitz at that point with respect to certain directions.

In this paper we will also show that when f is lower semicontinuous at a point x (but not necessarily continuous) and directionally Lipschitz in Ω at the point with respect to $v \in H_\Omega(x)$, one has $f^\uparrow(x; v) = f^+(x; v) = f_R^\circ(x; v)$, where

$$f_R^\circ(x; v) = \limsup_{\substack{x' \rightarrow_f x, x' \in \Omega \\ t \downarrow 0, x' + tv \in \Omega}} \frac{f(x' + tv) - f(x')}{t}. \quad (4)$$

This result was originally established by Rockafellar [20] for the case $\Omega = \mathbb{R}^n$. Also by extending the results of Rockafellar [20] for the constrained setting $\Omega \neq \mathbb{R}^n$, we will show, under appropriate conditions, that the upper subderivative $f^\uparrow(x; v)$, when v is in the tangent cone $T_\Omega(x)$, is the limit inferior of derivatives $f_R^\circ(x; w)$ where $w \in H_\Omega(x)$. This analysis will allow us then to state a result for directions in the tangent cone $T_\Omega(x_*)$ but not necessarily in the hypertangent cone $H_\Omega(x_*)$.

These results apply to discontinuous functions but they do not provide information about the ability of the algorithms to locally identify the best

branch or step function. It is possible, however, to prove that the algorithms have the capability to generate an infinite number of iterates in such a step, provided the number of steps is two around the limit point and the function has some continuity properties in each step (essentially the step domains must have nonempty interiors and one must be able to extend the function, in a certain Lipschitz continuous way, from a step domain to a neighborhood of the limit point).

This paper follows the line of others where nonsmooth calculus (in particular Clarke calculus) has been used to analyze the asymptotic properties of the sequence of iterates generated by DSM of directional type (besides the above cited papers [4, 5], see also [2, 11]).

We organize the material of this paper in the following way. In Section 2 we describe the algorithmic setting for direct-search methods, and gather the necessary material about the globalization strategies that we consider and about the notions of refining subsequences and directions. The main asymptotic results of this paper are contained in Section 3 for (possibly discontinuous) functions directionally Lipschitz with respect to certain directions. We leave to an appendix all the auxiliary nonsmooth calculus background needed for these results. In Section 4 we study the behavior of the algorithm for step discontinuities. We illustrate a number of our results and assumptions in Section 5, numerically and for problems in two dimensions. The paper is concluded in Section 6 with some final remarks.

2 Algorithmic framework and behavior of the step size

2.1 Algorithmic description

Our algorithmic description follows the one in [10, Chapter 7] for the unconstrained case. This framework will encompass both the MADS methodology (based on integer lattices and where a simple decrease on the function value suffices to identify a new iterate) and general directional DSM based on randomly generated normalized directions and sufficient decrease for acceptance of new iterates. Each iteration of the algorithm is organized around a search step (optional) and a poll step. The evaluation process of the poll step is opportunistic moving to a poll point once simple or sufficient decrease is found, depending on the variant being used.

As we will see later in the convergence theory, the set of directions used for polling is not required to positively span \mathbb{R}^n (although for coherence with the smooth case we will write it so in the algorithm below) and not

necessarily drawn from a finite set of directions. The algorithm requires an initial feasible point with finite objective function value.

To make the algorithmic description shorter we will make use of the extreme barrier function

$$f_{\Omega}(x) = \begin{cases} f(x) & \text{if } x \in \Omega, \\ +\infty & \text{otherwise.} \end{cases}$$

Following the terminology in [16], $\rho : (0, +\infty) \rightarrow (0, +\infty)$ will represent a forcing function, i.e., a continuous and non decreasing function satisfying $\rho(t)/t \rightarrow 0$ when $t \downarrow 0$. Typical examples of forcing functions are $\rho(t) = t^{1+a}$, for $a > 0$. To write the algorithm in general terms we will use $\bar{\rho}(\cdot)$ to either represent a forcing function $\rho(\cdot)$ or the constant, zero function. A relatively minor difference from the presentation below to what is in [10, Chapter 7] is the use of $\bar{\rho}(\alpha_k \|d_k\|)$ instead of $\bar{\rho}(\alpha_k)$ — this will be discussed in Section 2.3.

Algorithm 2.1 (Directional direct-search method)

Initialization

Choose $x_0 \in \Omega$ with $f(x_0) < +\infty$, $\alpha_0 > 0$, $0 < \beta_1 \leq \beta_2 < 1$, and $\gamma \geq 1$.
Let \mathcal{D} be a (possibly infinite) set of positive spanning sets.

For $k = 0, 1, 2, \dots$

1. **Search step:** Try to compute a point with $f_{\Omega}(x) < f(x_k) - \bar{\rho}(\alpha_k)$ by evaluating the function f at a finite number of points (in an integer lattice or mesh if $\bar{\rho}(\cdot) = 0$, see Section 2.2). If such a point is found then set $x_{k+1} = x$, declare the iteration and the search step successful, and skip the poll step.
2. **Poll step:** Choose a positive spanning set D_k from the set \mathcal{D} . Order the set of poll points $P_k = \{x_k + \alpha_k d : d \in D_k\}$. Start evaluating f_{Ω} at the poll points following the chosen order. If a poll point $x_k + \alpha_k d_k$ is found such that $f_{\Omega}(x_k + \alpha_k d_k) < f(x_k) - \bar{\rho}(\alpha_k \|d_k\|)$ then stop polling, set $x_{k+1} = x_k + \alpha_k d_k$, and declare the iteration and the poll step successful. Otherwise declare the iteration (and the poll step) unsuccessful and set $x_{k+1} = x_k$.
3. **Step size parameter update:** If the iteration was successful then maintain or increase the step size parameter: $\alpha_{k+1} \in [\alpha_k, \gamma \alpha_k]$. Otherwise decrease the step size parameter: $\alpha_{k+1} \in [\beta_1 \alpha_k, \beta_2 \alpha_k]$.

The global convergence of directional DSM is heavily based on the analysis of the behavior of the step size parameter α_k which must approach zero as an indication of some form of stationarity. There are essentially two known ways of enforcing the existence of a subsequence of step size parameters converging to zero in DSM of directional type. One way is by ensuring that all new iterates lie on an integer lattice (rigorously speaking only when the step size is bounded away from zero). The other form consists of imposing a sufficient decrease on the acceptance of new iterates. In the former case we need the iterates to lie in a bounded set and in the latter situation the objective function must be bounded below.

Assumption 2.1 *The level set $L(x_0) = \{x \in \Omega : f(x) \leq f(x_0)\}$ is bounded. The function f is bounded below in $L(x_0)$.*

2.2 Integer lattices (MADS)

Generalized pattern search makes use of a finite set of directions $D = \mathcal{D}$ which satisfy appropriate integrality requirements for globalization by integer lattices.

Assumption 2.2 *The set D of positive spanning sets is finite and the elements of D are of the form $G\bar{z}_j$, $j = 1, \dots, |D|$, where $G \in \mathbb{R}^{n \times n}$ is a nonsingular matrix and each \bar{z}_j is a vector in \mathbb{Z}^n .*

Given the type of non-smoothness and discontinuity of the objective function which we would like to consider in this paper, we need to make use of an infinite set of directions \mathcal{D} dense (after normalization) in the unit sphere. MADS makes use of such a set of directions but, since it is also based on globalization by integer lattices, the set \mathcal{D} must then be generated from a finite set D satisfying Assumption 2.2 (which will be guaranteed by the first requirement of the next assumption).

Assumption 2.3 *Let D represent a finite set of positive spanning sets satisfying Assumption 2.2.*

The set \mathcal{D} is so that the elements $d_k \in D_k \subseteq \mathcal{D}$ satisfy the following conditions:

1. d_k is a nonnegative integer combination of the columns of D .
2. The distance between x_k and the point $x_k + \alpha_k d_k$ tends to zero if and only if α_k does:

$$\lim_{k \in K} \alpha_k \|d_k\| = 0 \iff \lim_{k \in K} \alpha_k = 0,$$

for any infinite subsequence K .

3. The limits of all convergent subsequences of $\bar{D}_k = \{d_k/\|d_k\| : d_k \in D_k\}$ are positive spanning sets for \mathbb{R}^n .

We remark that the third requirement in the above definition is not used in the convergence theory of DSM of directional type for nonsmooth objective functions, but is nonetheless included for consistency with the smooth case and because it is part of the MADS original presentation [5].

In addition to Assumptions 2.2 and 2.3, the update of the step size parameter must conform to some form of integrality.

Assumption 2.4 *The step size parameter is updated as follows: Choose a rational number $\tau > 1$, a nonnegative integer $m^+ \geq 0$, and a negative integer $m^- \leq -1$. If the iteration is successful, the step size parameter is maintained or increased by taking $\alpha_{k+1} = \tau^{m_k^+} \alpha_k$, with $m_k^+ \in \{0, \dots, m^+\}$. Otherwise, the step size parameter is decreased by setting $\alpha_{k+1} = \tau^{m_k^-} \alpha_k$, with $m_k^- \in \{m^-, \dots, -1\}$.*

Note that these rules respect those of Algorithm 2.1 by setting $\beta_1 = \tau^{m^-}$, $\beta_2 = \tau^{-1}$, and $\gamma = \tau^{m^+}$.

Finally, the search step is restricted to points in a previously (implicitly defined) mesh or grid. Note that poll points must also lie on the mesh, but this requirement is trivially satisfied from the definition of the mesh M_k given below (i.e., one trivially has $P_k \subset M_k$).

Assumption 2.5 *The search step in Algorithm 2.1 only evaluates points in*

$$M_k = \bigcup_{x \in S_k} \{x + \alpha_k D z : z \in \mathbb{N}_0^{|D|}\},$$

where S_k is the set of all the points evaluated by the algorithm previously to iteration k .

The following result was originally proved by Torczon [21] for pattern search and extended later to generalized pattern search [4] and MADS [5].

Theorem 2.1 *Let Assumption 2.1 hold. Algorithm 2.1 under Assumptions 2.3–2.5 and $\bar{\rho}(\cdot) = 0$ (MADS) generates a sequence of iterates satisfying*

$$\liminf_{k \rightarrow +\infty} \alpha_k = 0.$$

2.3 Sufficient decrease

An alternative to the use of integer lattices is to impose sufficient rather than simple decrease as a criterion for accepting new iterates. This can be simply achieved by selecting $\bar{\rho}(\cdot)$ as a forcing function in Algorithm 2.1. We will need the following assumption (which, note, was already part of Assumption 2.3).

Assumption 2.6 *The distance between x_k and the point $x_k + \alpha_k d_k$ tends to zero if and only if α_k does:*

$$\lim_{k \in K} \alpha_k \|d_k\| = 0 \iff \lim_{k \in K} \alpha_k = 0,$$

for any infinite subsequence K .

The result below (Theorem 2.2) is relatively classic in nonlinear optimization when using some form of sufficient decrease. It is proved in [16] and in [10, Section 7.7] in the context of directional DSM for unconstrained optimization when using $\bar{\rho}(\alpha_k)$ instead of $\bar{\rho}(\alpha_k \|d_k\|)$. However, Assumption 2.6 allows the proof to easily go through for the latter case.

Theorem 2.2 *Let Assumption 2.1 hold. Algorithm 2.1, when $\bar{\rho}(\cdot)$ is a forcing function and Assumption 2.6 holds, generates a sequence of iterates satisfying*

$$\liminf_{k \rightarrow +\infty} \alpha_k = 0.$$

Note that such a result is derived under a very weak assumption on the set of directions \mathcal{D} . We are free to use, for instance, a normalized set of directions \mathcal{D} dense in the unit sphere.

2.4 Refining subsequences and directions

The type of stationarity results which can be derived for DSM of directional type are established at limit points of the so-called refining subsequences (a concept formalized in [4]).

Definition 2.1 *A subsequence $\{x_k\}_{k \in K}$ of iterates corresponding to unsuccessful poll steps is said to be a refining subsequence if $\{\alpha_k\}_{k \in K}$ converges to zero.*

One can ensure for the two algorithmic settings of this paper (Sections 2.2 and 2.3) the existence of a convergent refining subsequence. Such a result is a simple and known consequence of Assumption 2.1, Theorems 2.1 or 2.2, and the scheme that updates the step size parameter (see, e.g., [10, Section 7.3]).

Theorem 2.3 *Let Assumption 2.1 hold. Consider a sequence of iterates generated by Algorithm 2.1 under the scenarios of either Section 2.2 (MADS) or Section 2.3 (sufficient decrease). Then there is at least a convergent refining subsequence $\{x_k\}_{k \in K}$.*

The type of directions along which appropriate directional derivatives will be proved nonnegative are the so-called refining directions (a notion formalized in [5]).

Definition 2.2 *Let x_* be the limit point of a convergent refining subsequence. If the limit $\lim_{k \in L} d_k / \|d_k\|$ exists, where $L \subseteq K$ and $d_k \in D_k$, and if $x_k + \alpha_k d_k \in \Omega$, for sufficiently large $k \in L$, then this limit is said to be a refining direction for x_* .*

Some of the results of this paper will require for the refining subsequence under consideration that the associated set of refining directions for x_* is dense in the unit sphere (an assumption stronger than just saying that the normalized set of directions \mathcal{D} is dense in the unit sphere).

3 Results for discontinuous functions using the directionally Lipschitz property

A reader not familiar with nonsmooth calculus, in particular with the paper of Rockafellar [20], might find difficulties in grasping the essentials of the upper subderivative $f^\uparrow(x; v)$ and the notion of directionally Lipschitz with respect to a vector. Reading the Sections 2.4 and 2.9 of the book [9] will definitely help. However, these concepts have an immediate geometrical insight. Let us consider the unconstrained case $\Omega = \mathbb{R}^n$ for the purposes of the discussion in the current paragraph and the next two ones. In fact, one has that the epigraph $\text{epi}(f^\uparrow(x; \cdot))$ of the upper subderivative (for a fixed point x and as a function of the directions) coincides with the tangent cone $T_{\text{epi } f}(x, f(x))$ of the epigraph ($\text{epi } f$) of f at $(x, f(x))$, i.e., $\text{epi}(f^\uparrow(x; \cdot)) = T_{\text{epi } f}(x, f(x))$. It is known that this result remains true when f is not lower semicontinuous, for a definition of upper subderivative (see the Appendix) which reduces to (3) under lower semicontinuity. Also, such a result has been generalized by us, in the Appendix, to the case $\Omega \neq \mathbb{R}^n$.

The above characterization of the tangent cone of the epigraph of a function allows us to immediately see, for $n = 1$ or $n = 2$, when the upper subderivative is finite or equal to $+\infty$ (the definition of tangent cone is given later, but a geometrical intuition of a tangent cone suffices for the moment).

In fact, this geometric characterization is precisely the reason why this directional derivative is called upper subderivative (it allows us to look up and determine the vectors in the tangent cone of the epigraph of a function).

Now, let $D_{f,\Omega}(x)$ denote the set of vectors with respect to which f is directionally Lipschitz. It is also known that $D_{f,\Omega}(x) = \text{int}(\{w : f^\uparrow(x; w) < +\infty\})$ (a result also generalized by us to the case $\Omega \neq \mathbb{R}^n$ and also valid without lower semicontinuity). Thus, from the characterization $\text{epi}(f^\uparrow(x; \cdot)) = T_{\text{epi } f}(x, f(x))$, one can also easily compute $D_{f,\Omega}(x)$ for examples in one or two dimensions, gaining insight for the n -dimensional case. In Section 5 we report what is $D_{f,\Omega}(x)$ at $x = (0, 0)$ for all the four examples listed there.

Our first convergence result addresses the case where a refining direction is in the hypertangent cone to Ω at the limit point (see [9, Page 57]).

Definition 3.1 *A vector v is said to be hypertangent to Ω at x if there exists an $\epsilon > 0$ such that*

$$x' + tv' \in \Omega \quad \text{for all } x' \in \Omega \cap B(x; \epsilon), \quad v' \in B(v; \epsilon), \quad \text{and } t \in (0, \epsilon).$$

The set of all hypertangent vectors to Ω at x is called the hypertangent cone to Ω at x and is represented by $H_\Omega(x)$.

It is easy to see that the hypertangent cone is convex. Note that the definition of hypertangency used by Rockafellar [20] is different from the one used by Clarke [9] which later became the standard (and was used for the analysis of MADS [5]).

As we have seen before, the existence of a convergent refining subsequence $\{x_k\}_{k \in K}$ is guaranteed by Theorem 2.3. It is then possible to state this condition as an assumption for deriving asymptotic results at limit points.

Theorem 3.1 *Consider a refining subsequence $\{x_k\}_{k \in K}$ converging to $x_* \in \Omega$ and a refining direction v for x_* in $H_\Omega(x_*)$. Assume that f is lower semicontinuous at x_* and directionally Lipschitz at x_* with respect to v . Assume further that $\lim_{k \in K} f(x_k) = f(x_*)$. Then $f^\uparrow(x_*; v) = f^+(x_*; v) = f_R^\circ(x_*; v) \geq 0$.*

Proof. Since f is directionally Lipschitz at x_* with respect to v , we have

that $f^+(x_*; v) < +\infty$. Now let $v = \lim_{k \in L} d_k / \|d_k\|$, with $L \subseteq K$. Thus,

$$\begin{aligned}
f^+(x_*; v) &= \limsup_{\substack{x' \rightarrow_f x_*, x' \in \Omega \\ t \downarrow 0}} \sup_{\substack{v' \rightarrow v \\ x' + tv' \in \Omega}} \frac{f(x' + tv') - f(x')}{t} \\
&\geq \limsup_{k \in L} \frac{f(x_k + \alpha_k \|d_k\| (d_k / \|d_k\|)) - f(x_k)}{\alpha_k \|d_k\|} \\
&= \limsup_{k \in L} \frac{f(x_k + \alpha_k d_k) - f(x_k) + \bar{\rho}(\alpha_k \|d_k\|)}{\alpha_k \|d_k\|} - \frac{\bar{\rho}(\alpha_k \|d_k\|)}{\alpha_k \|d_k\|} \geq 0.
\end{aligned}$$

The first inequality follows from $\{x_k\}_{k \in L}$ being a feasible refining subsequence with $\lim_{k \in L} f(x_k) = f(x_*)$ and the fact that $x_k + \alpha_k d_k$ is feasible for $k \in L$ sufficiently large. The limit $\lim_{k \in L} \bar{\rho}(\alpha_k \|d_k\|) / (\alpha_k \|d_k\|)$ is 0 for both globalization strategies (Sections 2.2 and 2.3). In the case of MADS (Section 2.2)¹, one uses $\bar{\rho}(\cdot) = 0$. When imposing sufficient decrease (Section 2.3), it follows directly from the properties of a forcing function and from Assumption 2.6.

The fact that $f^\uparrow(x_*; v) = f^+(x_*; v) = f_R^\circ(x_*; v)$ is showed in the Appendix (Theorem A.2). ■

Now we need to address the case where the directions are in the tangent cone to Ω at the limit point but not necessarily in the hypertangent cone.

Definition 3.2 *A vector v is said to be tangent to Ω at x if for all sequences $\{y_k\} \subset \Omega$ converging to x and for all sequences $\{t_k\}$ with $t_k \downarrow 0$, there exists a sequence of vectors $\{w_k\}$ converging to v such that $y_k + t_k w_k \in \Omega$ for all k .*

The set of all tangent vectors to Ω at x is called the tangent cone to Ω at x and is represented by $T_\Omega(x)$.

The tangent cone $T_\Omega(x)$ is the closure of the hypertangent cone $H_\Omega(x)$. It can be also defined by the limit inferior of a multifunction (see the Appendix for details).

Let $D_{f,\Omega}(x)$ denote the set of vectors in $H_\Omega(x)$ with respect to which f is directionally Lipschitz. Note that in the presence of constraints ($\Omega \neq \mathbb{R}^n$) the definition of $D_{f,\Omega}(x)$ makes only sense for vectors in $H_\Omega(x)$.

Theorem 3.2 *Consider a refining subsequence $\{x_k\}_{k \in K}$ converging to $x_* \in \Omega$. Let v be in $T_\Omega(x_*)$ but not necessarily in $H_\Omega(x_*)$ (which, in turn, is*

¹Note that MADS and the imposition of sufficient decrease are compatible. In fact, this is true because we are using $\bar{\rho}(\alpha_k \|d_k\|)$ in the statement of Algorithm 2.1 (instead of $\bar{\rho}(\alpha_k)$).

assumed nonempty). Assume that f is lower semicontinuous at x_* and $f^\uparrow(x_*; v) < +\infty$. Assume further that $\lim_{k \in K} f(x_k) = f(x_*)$.

If f is directionally Lipschitz with respect to all directions in the intersection of a ball centered at v with $H_\Omega(x_*)$ and the set of refining directions for x_* is dense in this intersection, then $f^\uparrow(x_*; v) \geq 0$.

Proof. First we apply Theorem 3.1 to obtain that $f_R^\circ(x_*; w) \geq 0$ for all the refining directions w in the intersection of the ball centered at v with $D_{f,\Omega}(x_*)$. Then, from the result proved in the Appendix (Theorem A.2),

$$f^\uparrow(x_*; v) = \liminf_{\substack{w \rightarrow v \\ w \in D_{f,\Omega}(x_*)}} f_R^\circ(x_*; w) \geq 0.$$

■

4 Identification of the best branch for discontinuous functions

We are now interested in studying the behavior of directional DSM when the objective function is defined by several branches or steps, in particular to know if the algorithm can identify the locally best step. We will give an affirmative answer provided the number of steps is two, the step domains have nonempty interiors, and their borders exhibit a minimum of regularity (the exterior cone property stated below in Definition 4.1).

The condition given next covers a wide range of discontinuities.

Assumption 4.1 *The function f is such that there exists a neighborhood B of x_* (a limit point of a refining subsequence) which admits a finite partition*

$$B = \bigcup_{i=1}^{n_B} B_i,$$

such that, for all $i \in \{1, \dots, n_B\}$,

1. $\text{int}(B_i) \neq \emptyset$,
2. $\text{cl}(B_i)$ has the exterior cone property (see Definition 4.1),
3. f is Lipschitz continuous in $\text{int}(B_i)$ and can be continuously extended from $\text{int}(B_i)$ to ∂B_i .

We will refer to the B_i 's as either step domains or partition sets. It can be easily seen that if we extend a Lipschitz continuous function in the interior of a set to the boundary, in a continuous way, the extension is also Lipschitz continuous on the closure of the set.

Proposition 4.1 *Let f be a locally Lipschitz function in $\text{int}(S)$, with Lipschitz constant L . The continuous extension of f from S to $\text{cl}(S)$ is locally Lipschitz continuous with Lipschitz constant $3L + 2$.*

Proof. Let $x \in \partial S$ and consider a neighborhood N of x . The value of the extended function in any point $y \in \partial S \cap N$ can be given as the limit of $\{f(y^j)\}$ for any sequence of points $\{y^j\} \subseteq \text{int}(S) \cap N$ converging to y .

Let us consider two points z and w in $\partial S \cap N$. Let $\epsilon \leq \|z - w\|$. Then, there exist $z_\epsilon, w_\epsilon \in \text{int}(S) \cap N$, with $\|z - z_\epsilon\| \leq \epsilon$ and $\|w - w_\epsilon\| \leq \epsilon$, such that:

$$\begin{aligned} |f(z) - f(w)| &\leq |f(z) - f(z_\epsilon)| + |f(z_\epsilon) - f(w_\epsilon)| + |f(w_\epsilon) - f(w)| \\ &\leq \epsilon + L\|z_\epsilon - w_\epsilon\| + \epsilon \leq \epsilon + 3L\|z - w\| + \epsilon \\ &\leq (2 + 3L)\|z - w\|. \end{aligned}$$

The case where one point is in $\text{int}(S) \cap N$ and the other in $\partial S \cap N$ can be proved analogously (the result is trivial when both points are in $\text{int}(S) \cap N$).

■

The precise form of the exterior cone property which we will use is stated below. Note that the border of a set exhibiting this property cannot contain singularities like cusps, in other words, points like the origin when the set is of the form $\mathbb{R}^2 \setminus \{(x, y) \in \mathbb{R}^2 : -x^2 \leq y \leq x^2, x \geq 0\}$. The exterior cone property holds for a set when it is possible to ‘stick’ a cone with nonempty interior to the complementary of the set at any point of its boundary.

Definition 4.1 *A set S has the exterior cone property if at any point $z \in \partial S$ there exists a cone C_z with nonempty interior, an angle $\theta_z > 0$, and a neighborhood N_z of z such that $E_z = N_z \cap \{z' = z + c, c \in C_z, c \neq 0\} \subset \mathbb{R}^n \setminus S$ and any angle between all the vectors in $E_z - \{z\}$ and all the vectors in $S_z - \{z\}$, with $S_z = S \cap N_z$, is larger than θ_z .*

We will also need the following auxiliary result. Essentially we extend a Lipschitz continuous function from a set to \mathbb{R}^n by first extending it (at a given point on the boundary of the set) to an ‘exterior cone’ of nonempty interior along which the function is strictly decreasing. In this way, our extension to \mathbb{R}^n is guaranteed to decrease strictly for a set of directions of nonempty interior.

Proposition 4.2 *Let S be a set with the exterior cone property and g a Lipschitz continuous function in S . Let also $z \in \partial S \cap S$.*

Then there exists an extension \tilde{g} of g from S to \mathbb{R}^n which is Lipschitz continuous in \mathbb{R}^n and locally strictly decreasing along all directions emanating from z and belonging to a cone with nonempty interior.

Proof. Consider the sets C_z , N_z , E_z , and S_z as in Definition 4.1. Define an auxiliary function \tilde{g} which coincides with g in S_z and is linear and strictly decreasing from z to the interior of E_z . We will show first that this function is Lipschitz continuous in $S_z \cup E_z$. Let L_1 be the Lipschitz constant of $g = \tilde{g}$ in S_z and L_2 the Lipschitz constant of \tilde{g} in E_z . Now consider a point $y \in S_z$ and a point $w \in E_z$. One can derive that

$$\begin{aligned} |\tilde{g}(y) - \tilde{g}(w)| &\leq |g(y) - g(z)| + |\tilde{g}(w) - \tilde{g}(z)| \\ &\leq \max\{L_1, L_2\} (\|y - z\| + \|w - z\|) \\ &\leq \max\{L_1, L_2\} M \|y - w\|, \end{aligned}$$

where the last equality follows from $\|y - z\| + \|w - z\| \leq M \|y - w\|$, with $M = \sqrt{2}\pi/\theta_z$. The cases where both points lie in S_z or E_z are straightforward to analyze. We then obtain that $|\tilde{g}(y) - \tilde{g}(w)| \leq \max\{L_1, L_2\} \max\{1, M\} \|y - w\|$, for all y and w in $S_z \cup E_z$.

It is known that any Lipschitz function in a set can be extended to the whole space with the same Lipschitz constant (see [17, Theorem 1]). Thus, one can now extend \tilde{g} from $S_z \cup E_z$ to \mathbb{R}^n , and in particular to N_z , with the same Lipschitz constant. ■

We are now ready for the main result of this section. Recall that the existence of a convergent refining subsequence $\{x_k\}_{k \in K}$ is guaranteed by Theorem 2.3. The technique of Proposition 4.2 is used to reach a contradiction in the proof of Theorem 4.1 below, thus guaranteeing (at a limit point where the function is discontinuous) an infinite number of poll points (corresponding to unsuccessful iterations) in the complementary of the step domain where those unsuccessful iterates belong to.

Theorem 4.1 *Consider a refining subsequence $\{x_k\}_{k \in K}$ converging to $x_* \in \Omega$ (and note that Assumption 2.1 is required for the existence of such a subsequence). Assume that f is lower semicontinuous at x_* and satisfies Assumption 4.1. Let the sets of refining directions for x_* corresponding to any infinite subsequence of K be dense in the unit sphere.*

If x_ belongs to the interior of a partition set in $\{B_1, \dots, B_{n_B}\}$, then $f_C^\circ(x_*; v) \geq 0$ for all refining directions $v \in T_\Omega(x_*)$ (assuming here also that $H_\Omega(x_*)$ is nonempty).*

Otherwise, there exists a subsequence $K' \subset K$ and a partition set $B' \in \{B_1, \dots, B_{n_B}\}$ such that (i) $\{x_k\}_{k \in K'} \subset \text{cl}(B')$, (ii) there is an infinite number of poll points, corresponding to iterates in K' , in $\text{int}(B')$, and (iii) there is an infinite number of poll points, corresponding to iterates in K' , in $\mathbb{R}^n \setminus \text{cl}(B')$.

Proof. The proof is done for the case of MADS (Section 2.2) but the case of sufficient decrease (Section 2.3) is obtained from this one with minor modifications.

Consider first the neighborhood B guaranteed by Assumption 4.1. If x_* belongs to $\text{int}(B_l)$, for some $l \in \{1, \dots, n_B\}$, the Lipschitz continuity of f near x_* would allow to apply the known results from [5].

So, let us assume that x_* belongs to the boundary of B_l , for some $l \in \{1, \dots, n_B\}$. If all the iterates after a certain order lie in borders of the step domains, then note that the theorem can be easily established, since the partition is finite and we assume density of the sets of refining directions for x_* corresponding to any infinite subsequence of K . Otherwise, since the partition is finite, by passing to a subsequence $K_1 \subset K$ if necessary, one can state the existence of an $i \in \{1, \dots, n_B\}$ such that $x_* \in \partial B_i$ and $\{x_k\}_{k \in K_1} \subset \text{int}(B_i)$ with $K_1 \subset K$. Also, given that B_i has a nonempty interior and that the set of refining directions for x_* corresponding to K_1 is dense in the unit sphere, there must exist an infinite number of poll points associated with a subsequence $K_2 \subset K_1$ belonging to $\text{int}(B_i)$.

By using Assumption 4.1 and Proposition 4.1, we can extend f from B_i to $\text{cl}(B_i)$ in a continuous way and ensuring that the extended function \tilde{f} is Lipschitz continuous in $\text{cl}(B_i)$.

Let us assume that all poll points associated with the refining subsequence belong to $\text{cl}(B_i)$. We will see that this leads us to a contradiction. So, let us assume that there exists a $\bar{k} \in K_1$ such that $x_k + \alpha_k d \in \text{cl}(B_i)$ for all $k \in K_1$ with $k \geq \bar{k}$ and for all $d \in D_k$. We now apply Proposition 4.2 using $S = \text{cl}(B_i)$, $g = \tilde{f}$, and $z = x_*$. Let \tilde{f} be the extended function (and L its Lipschitz constant). We then obtain that

$$\begin{aligned} & \tilde{f}_C^\circ(x_*; v) \\ & \geq \limsup_{k \in K_1} \frac{\tilde{f}(x_k + \alpha_k d_k) - \tilde{f}(x_k)}{\alpha_k \|d_k\|} + \frac{\tilde{f}(x_k + \alpha_k \|d_k\| v) - \tilde{f}(x_k + \alpha_k d_k)}{\alpha_k \|d_k\|} \\ & \geq \limsup_{k \in K_2} \frac{f(x_k + \alpha_k d_k) - f(x_k)}{\alpha_k \|d_k\|} - \frac{L \alpha_k \|d_k\| \|v - (d_k / \|d_k\|)\|}{\alpha_k \|d_k\|} \\ & \geq 0, \end{aligned}$$

for all refining directions v , which is a contradiction since these directions are

dense in the unit sphere and \tilde{f} is locally strictly decreasing from x_* along all directions in a cone of nonempty interior.

So, one can build a sequence of points $K_3 \subset K_1$ for which there exists $d_k \in D_k$ such that

$$f_\Omega(x_k + \alpha_k d_k) \geq f(x_k), \quad x_k + \alpha_k d_k \notin \text{cl}(B_i), \quad (5)$$

for all $k \in K_3$.

The proof is completed by setting $B' = B_i$ and $K' = K_1$. ■

When the number of steps is equal to two ($n_B = 2$) it is possible to prove a stronger result. Such a result is an obvious consequence of the existence of an infinite number of poll points (corresponding to unsuccessful iterations) in the other step domain, i.e., in the step domain complementary to the one where those unsuccessful iterates belong to.

Corollary 4.1 *Under Assumption 2.1 and the assumptions of Theorem 4.1 and when $n_B = 2$, there exists a subsequence $K_* \subset K$ and a partition set $B_* \in \{B_1, B_2\}$ such that, when x_* is in the border of the two partition sets,*

1. B_* satisfies the properties stated for B' in Theorem 4.1,
2. $\lim_{k \in K_*} f(x_k) = f(x_*)$.

Proof. The proof of the corollary is a continuation of the proof of the Theorem 4.1.

First we note that $\{f(x_k)\}$ is decreasing and bounded below and thus it converges, say to f_* . Since f is lower semicontinuous, $f(x_*) \leq f_*$.

We can now show that it is along B' (see the proof of Theorem 4.1) that the value of f is attained, i.e., that $f_* = \lim_{k \in K'} f(x_k) = f(x_*)$. If this was not true, then there would exist an $\epsilon > 0$ and a bordering B'' (since $n_B = 2$, the remaining one) and a neighborhood N of x_* for which $f(y) > f(z) + \epsilon$, for all $y \in B' \cap N \cap \Omega$ and $z \in B'' \cap N \cap \Omega$. But this contradicts (5). ■

5 Numerical illustrations

To illustrate the ability of Algorithm 2.1 in finding local minimizers for lower semicontinuous functions we ran some examples in MATLAB. We included examples which violate some of the assumptions required to ensure convergence. Four problems of the form (1) were considered, where $\Omega = [-1, 1] \times [-1, 1]$ was partitioned into a finite number of disjoint subsets $\Omega = \bigcup_{i=1}^{n_B} \Omega_i$, with $n_B = 2$ in three of the cases and $n_B = 4$ in the last problem. The minimizer is unique and corresponds to $x_* = (0, 0)$. Figures 1–4

depict plots of each one of the functions considered. Functions f_1 , f_3 , and f_4 are discontinuous at $x_* = (0, 0)$. Function f_2 is continuous at $x_* = (0, 0)$ but not Lipschitz continuous near this point. We call the attention of the reader to the first three paragraphs of Section 3 for instructions on how to determine whether a function is directionally Lipschitz with respect to a given direction at a given point.

Problem 1: A lower semicontinuous function of the form

$$f_1(x) = \begin{cases} x_1^2 + x_2^2 & \text{if } \frac{x_1}{2} \leq x_2 \leq 2x_1, \\ 10 + x_1^2 + x_2^2 & \text{otherwise,} \end{cases}$$

and steps Ω_1 and Ω_2 with nonempty interior. Near x_* , the distance between function values in the two steps remains constant.

The upper subderivative $f^\uparrow(x_*; v)$ is finite for all directions $v \in \mathbb{R}^2$ such that $v_1/2 \leq v_2 \leq 2v_1$ and coincides there with traditional directional derivatives. Thus, the function is directionally Lipschitz at x_* with respect to all directions $w \in \mathbb{R}^2$ such that $w_1/2 < w_2 < 2w_1$.

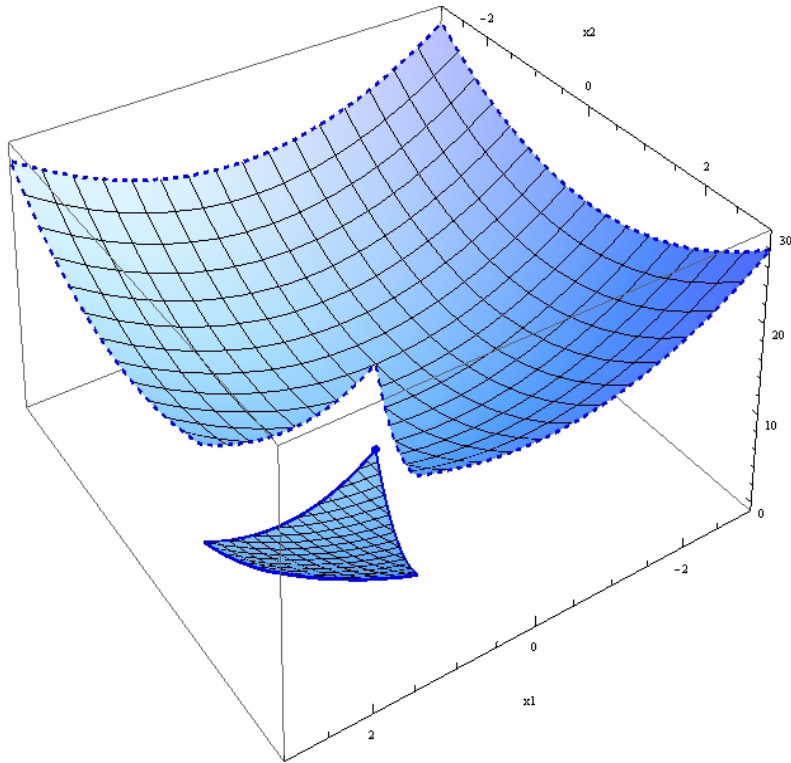


Figure 1: Plot of function f_1 .

Problem 2: A lower semicontinuous function of the form

$$f_2(x) = \begin{cases} 10x_1^2 + 10x_2^2 & \text{if } x_1 < 0, \\ 10x_1^2 + x_2^2 & \text{otherwise,} \end{cases}$$

and steps Ω_1 and Ω_2 with nonempty interior. Near x_* , the distance between function values in the steps converges to zero.

The upper subderivative $f^\uparrow(x_*; v)$ is finite for all directions $v \in \mathbb{R}^2$ such that $v_1 \geq 0$ and coincides there with traditional directional derivatives. The function is directionally Lipschitz at x_* with respect to all directions $w \in \mathbb{R}^2$ such that $w_1 > 0$.

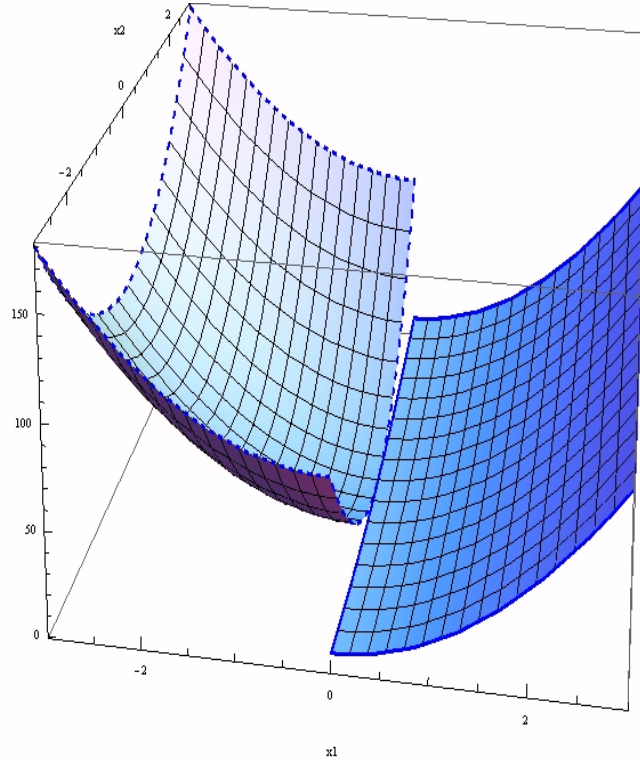


Figure 2: Plot of function f_2 .

Problem 3: A lower semicontinuous function of the form

$$f_3(x) = \begin{cases} x_1^2 + x_2^2 & \text{if } x_2 = 2x_1, \\ 10 + x_1^2 + x_2^2 & \text{otherwise,} \end{cases}$$

where one of the steps has empty interior.

The upper subderivative $f^\uparrow(x_*; v)$ is finite only along the directions $v \in \mathbb{R}^2$ such that $v_2 = 2v_1$ and coincides there with traditional directional derivatives. Thus, the function is not directionally Lipschitz at x_* since there is no direction with respect to which it is directionally Lipschitz.

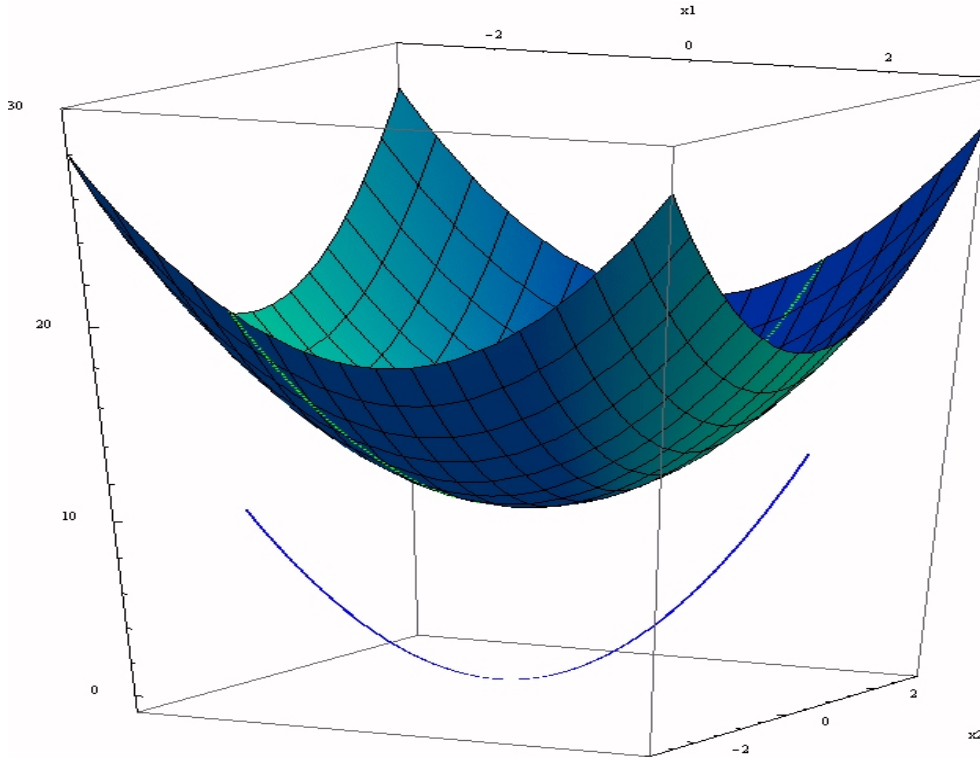


Figure 3: Plot of function f_3 .

Problem 4: A lower semicontinuous function of the form

$$f_4(x) = \begin{cases} x_1^2 + x_2^2 & \text{if } \frac{x_1}{2} \leq x_2 \leq 2x_1, \\ 5 + x_1^2 + x_2^2 & \text{if } x_1 \leq 0 \wedge x_2 \leq 0 \wedge (x_1, x_2) \neq (0, 0), \\ 10 + x_1^2 + x_2^2 & \text{if } x_2 < \frac{x_1}{2} \wedge x_1 > 0, \\ 15 + x_1^2 + x_2^2 & \text{otherwise,} \end{cases}$$

and steps Ω_i , $i \in \{1, 2, 3, 4\}$, with nonempty interior. Near x_* , the distance between function values in any of the steps remains constant. The number of steps considered exceeds two, which violates one of the conditions required in Section 4 to establish the asymptotic results.

The upper subderivative $f^\uparrow(x_*; v)$ is finite for all directions $v \in \mathbb{R}^2$ such that $v_1/2 \leq v_2 \leq 2v_1$ and coincides there with traditional directional deriva-

tives. Thus, the function is directionally Lipschitz at x_* with respect to all directions $w \in \mathbb{R}^2$ such that $w_1/2 < w_2 < 2w_1$.

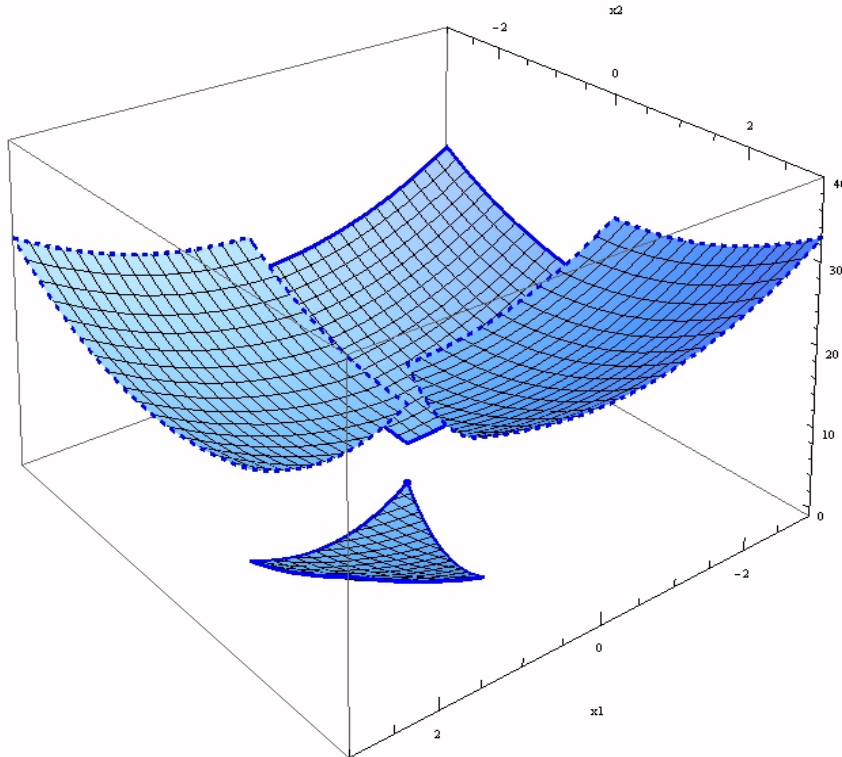


Figure 4: Plot of function f_4 .

We tested NOMADm [1], version 4.6, a MATLAB implementation of MADS (which fits into Algorithm 2.1, see Section 2.2) and a very simple implementation of Algorithm 2.1 with a globalization strategy based on sufficient decrease (as in Section 2.3). In both algorithms, the search step was empty, the initial step size parameter was set to one, and the run stopped once the step size reached the threshold 10^{-7} . In the implementation of the variant which requires sufficient decrease, the poll set D_k was set equal to $[Q_k - Q_k]$, where Q_k is an orthogonal matrix computed by randomly generating the first column. In MADS, the positive spanning set considered corresponds to the implementation LTMADS, with a total of $2n$ directions. Since our main concern is (proper) convergence rather than efficiency, the poll points were evaluated following the consecutive order of storage. As a forcing function, in the case of the sufficient decrease variant, we considered $\rho(t) = t^2$ (other variants were tested, but with worse results in what concerns the total number of function evaluations required).

Given the random behavior of both algorithms, a sequence of 10 runs was considered for each problem. The initial point was set to $x_0 = (-0.4, -0.5)$. A summary of the computational experiments is reported in Table 1.

function	Algorithm 2.1			
	MADS		Suff. Decrease	
	#failures	#fevals	#failures	#fevals
f_1	0	233	0	175.6
f_2	0	193.9	0	494.6
f_3	10	173.7	10	144.4
f_4	2	220.6	1	177.7

Table 1: Number of failures in identifying a local minimizer and corresponding average number of function evaluations required.

When any of the algorithms failed to converge to the function minimizer, the final iterate corresponded generally to a point near the minimizer of the function when restricted to a higher step. The exception occurred with f_3 where there were cases of convergence to points lying on the line of discontinuity.

In order to assess the dependency of the results from the initial point provided to the methods, we considered a grid of 100 equally spaced points in $[-1, -0.1] \times [-1, -0.1]$, and ran Algorithm 2.1 with a globalization strategy based on sufficient decrease. For each of the four problems and for each of the initial points, we ran the algorithm 10 times, yielding a total of 1000 runs for each problem. The number of failures in detecting the minimizer is reported in Table 2. The nature of the results remains the same if we consider a grid of equally spaced starting points in $\Omega = [-1, 1] \times [-1, 1]$.

function	#failures — for 10^{-7}	#failures — for 10^{-10}
f_1	2	0
f_2	0	0
f_3	1000	1000
f_4	61	44

Table 2: Number of failures of Algorithm 2.1 (sufficient decrease variant, see Section 2.3), for a sequence of 1000 runs, starting from different initial points, and for two different stopping tolerances (10^{-7} and 10^{-10}).

The numerical results support the theoretical analysis developed in the

previous sections. Failures in locating the minimizer occur only when at least one of the assumptions required for establishing convergence is violated. Note that the failures in f_1 when the stopping tolerance is 10^{-7} are due to the fact that not enough directions were generated. There is no discrepancy between these results and the theory because the latter assumes a set of refining directions dense in the unit sphere. In fact, by letting the algorithm run longer (stopping tolerance of 10^{-10}) these failures disappear while the ones for f_4 do not.

6 Final remarks

In this paper we tried to shed some light on the convergence properties of direct-search methods (DSM) of directional type for lower semicontinuous functions not necessarily continuous. We divided our analysis into two main parts. In the first part, we derive results for refining directions with respect to which the function is directionally Lipschitz (but not necessarily continuous) at the limit point of the underlying refining subsequence. These results were derived for the constrained case which forced us to redo the analysis in [20] for upper subderivatives in the presence of constraints.

In the second part of the analysis, we considered a class of discontinuous functions and showed that when the number of branches or steps is two and the function has some continuity properties in each step, these DSM identify the best local step around the limit point. The problem in extending this result to more than two local steps or branches lies on the fact that the speed at which the poll points approach the border of a step domain can be slower than the speed at which these points approach the iterates. We were able to prove, by extending continuously the function and taking limits, that an infinite number of poll points jump out of the step domain. However, they could only visit a neighbor step and thus one can only infer results when the number of steps is equal to two.

A Appendix

In this section we provide the rigorous definitions of the various generalized directional derivatives used throughout this paper.

Definition of upper subderivative

The upper subderivative (3) was defined by Rockafellar [20] for the case $\Omega = \mathbb{R}^n$. To extend it to the constrained case $\Omega \neq \mathbb{R}^n$, let $g(s, y)$ be an

extended-real-valued function defined on $(\mathbb{R}^n \times \mathbb{R} \times [0, +\infty)) \times \mathbb{R}^n$. Let also $s \in S \subset \mathbb{R}^n \times \mathbb{R} \times [0, +\infty)$. Define

$$h(s, y) = \limsup_{\substack{s' \rightarrow s \\ s' \in S}} \inf_{\substack{y' \rightarrow y \\ y' \in \Gamma_\Omega(p(s'))}} g(s', y')$$

as

$$\sup_{Y \in N(y)} \inf_{U \in N(s)} \sup_{s' \in S \cap U} \inf_{y' \in \Gamma_\Omega(p(s')) \cap Y} g(s', y'),$$

where $N(y)$ and $N(s)$ denote, respectively, a family of sufficiently small neighborhoods around y and s , $p(\cdot)$ denotes the projection from $\mathbb{R}^n \times \mathbb{R} \times [0, +\infty)$ onto $\mathbb{R}^n \times [0, +\infty)$, and

$$\Gamma_\Omega(x, t) = \begin{cases} t^{-1}(\Omega - x) & \text{if } t > 0, \\ \mathbb{R}^n & \text{if } t = 0. \end{cases}$$

To define the upper subderivative $f^\uparrow(x; v)$ one proceeds similarly as in [20] and chooses

$$g(x', \alpha', t, v') = \begin{cases} \frac{f(x'+tv') - \alpha'}{t} & \text{if } t > 0, \\ -\infty & \text{if } t = 0, \end{cases} \quad (6)$$

$s = (x, f(x), 0)$, $s' = (x', \alpha', t)$, $y = v$, and $y' = v'$. In the constrained case, however, one has now $S = \text{epi}(f)(\Omega) \times [0, +\infty)$, where $\text{epi}(f)(\Omega)$ is the epigraph of f restricted to Ω . These choices result then in the definition $f^\uparrow(x; v) = h((x, f(x), 0), v)$. We use the following expression to more easily grasp the essential of the definition of the upper subderivative $f^\uparrow(x; v)$:

$$f^\uparrow(x; v) = \limsup_{\substack{(x', \alpha') \downarrow_f x, x' \in \Omega \\ t \downarrow 0}} \inf_{\substack{v' \rightarrow v \\ x' + tv' \in \Omega}} \frac{f(x' + tv') - \alpha'}{t}.$$

The notation $(x', \alpha') \downarrow_f x$ represents $(x', \alpha') \rightarrow (x, f(x))$ with $\alpha' \geq f(x')$.

When f is lower semicontinuous at x , the derivative $f^\uparrow(x; v)$ can be equivalently defined by

$$f^\uparrow(x; v) = \limsup_{\substack{x' \rightarrow_f x, x' \in \Omega \\ t \downarrow 0}} \inf_{\substack{v' \rightarrow v \\ x' + tv' \in \Omega}} \frac{f(x' + tv') - f(x')}{t},$$

where, recall, $x' \rightarrow_f x$ represents $x' \rightarrow x$ and $f(x') \rightarrow f(x)$.

A characterization of the epigraph of the upper subderivative

The following proposition extends [20, Proposition 1] to the constrained case $\Omega \neq \mathbb{R}^n$.

Proposition A.1 *For each $s' \in S \subset \mathbb{R}^n \times \mathbb{R} \times [0, +\infty)$, let $\Gamma(s')$ denote the set in $\mathbb{R}^n \times \mathbb{R}$ which is the epigraph of $y \rightarrow g(s', y)$ restricted to $\Gamma_\Omega(p(s'))$:*

$$\Gamma(s') = \text{epi}(g(s', \cdot))(\Gamma_\Omega(p(s'))).$$

Let also

$$\Delta(s) = \liminf_{\substack{s' \rightarrow s \\ s' \in S}} \Gamma(s').$$

Then $\Delta(s)$ is the epigraph of $y \rightarrow h(s, y)$ restricted to $T_\Omega(s)$:

$$\text{epi}(h(s, \cdot))(T_\Omega(s)) = \Delta(s).$$

Proof. From its definition, the point (y, β) is in $\Delta(s)$ if and only if

$$\begin{aligned} \forall Y \in N(y), \forall \epsilon > 0, \exists U \in N(s) : \forall s' \in S \cap U, \exists (y', \beta') : \\ y' \in \Gamma_\Omega(p(s')) \cap Y, \beta' \in (\beta - \epsilon, \beta + \epsilon), g(s', y') \leq \beta'. \end{aligned}$$

which is equivalent to

$$\begin{aligned} \forall Y \in N(y), \forall \epsilon > 0, \exists U \in N(s) : \forall s' \in S \cap U, \exists y' : \\ y' \in \Gamma_\Omega(p(s')) \cap Y, g(s', y') \leq \beta + \epsilon. \end{aligned}$$

Thus, (y, β) is in $\Delta(s)$ if and only if $y \in T_\Omega(s)$ and

$$\forall Y \in N(y), \forall \epsilon > 0, \exists U \in N(s) : \sup_{s' \in S \cap U} \inf_{y' \in \Gamma_\Omega(p(s')) \cap Y} g(s', y') \leq \beta + \epsilon.$$

This last condition is the same as saying that $h(s, y) \leq \beta$. ■

Note that the epigraph of $y \rightarrow g(s', y)$ restricted to $\Gamma_\Omega(p(s'))$ in the case (6) considered for the upper subderivatives is:

$$\Gamma(x', \alpha', t) = \begin{cases} t^{-1} (\text{epi}(f)(\Omega) - (x', \alpha')) & \text{if } t > 0, \\ \mathbb{R}^n \times \mathbb{R} & \text{if } t = 0. \end{cases}$$

Thus, from Proposition A.1,

$$\liminf_{\substack{(x', \alpha') \downarrow_f x, x' \in \Omega \\ t \downarrow 0}} \Gamma(x', \alpha', t) = \text{epi}(f^\uparrow(x; \cdot))(T_\Omega(x)).$$

On the other hand, from the definition of tangent cone

$$\liminf_{\substack{(x', \alpha') \downarrow_f x, x' \in \Omega \\ t \downarrow 0}} \Gamma(x', \alpha', t) = T_{\text{epi}(f)(\Omega)}(x, f(x)).$$

Thus,

$$\text{epi}(f^\uparrow(x; \cdot))(T_\Omega(x)) = T_{\text{epi}(f)(\Omega)}(x, f(x)). \quad (7)$$

The relation (7) extends, to the constrained case, the part of [20, Theorem 2] which we need for what comes in Theorem A.2 below.

Definitions of other generalized directional derivatives

To define the generalized directional derivative $f_R^\circ(x; v)$ introduced in (4), in the constrained case, one first considers

$$h(s, y) = \limsup_{\substack{s' \rightarrow s \\ s' \in S : y \in \Gamma_\Omega(p(s'))}} g(s', y)$$

as

$$\inf_{U \in \mathcal{N}(s)} \sup_{s' \in S \cap U : y \in \Gamma_\Omega(p(s'))} g(s', y).$$

The derivative is then defined as $f_R^\circ(x; v) = h((x, f(x), 0), v)$ by setting g as in (6), $s = (x, f(x), 0)$, $s' = (x', \alpha', t)$, and $y = v$ and, given the constrained case, $S = \text{epi}(f)(\Omega) \times [0, +\infty)$. We will also use a more friendly description for this definition:

$$f_R^\circ(x; v) = \limsup_{\substack{(x', \alpha') \downarrow_f x, x' \in \Omega \\ t \downarrow 0, x' + tv \in \Omega}} \frac{f(x' + tv) - \alpha'}{t}.$$

When f is lower semicontinuous at x , the derivative $f_R^\circ(x; v)$ can be equivalently defined by

$$f_R^\circ(x; v) = \limsup_{\substack{x' \rightarrow_f x, x' \in \Omega \\ t \downarrow 0, x' + tv \in \Omega}} \frac{f(x' + tv) - f(x')}{t}.$$

Finally, if f is Lipschitz continuous near x , this derivative coincides with the Clarke-Jahn generalized directional derivative (2):

$$f_R^\circ(x; v) = f_C^\circ(x; v) = \limsup_{\substack{x' \rightarrow x, x' \in \Omega \\ t \downarrow 0, x' + tv \in \Omega}} \frac{f(x' + tv) - f(x')}{t}.$$

A characterization for the upper subderivatives

We reproduce below, in the space \mathbb{R}^m , the result in [9, Theorem 2.4.8] (originally proved in [20, Corollary 2, Page 268]). Recall, from Definition 3.1, the notion of a vector hypertangent to a set at a point of the set.

Proposition A.2 *Let $C \subset \mathbb{R}^m$ and $y \in C$. Suppose there is at least one $w \in H_C(y)$. Then, $T_C(y) = \text{cl}(H_C(y))$.*

Theorem A.1 *The function f is directionally Lipschitz at x with respect to $v \in H_\Omega(x)$ if and only if, for some $\beta \in \mathbb{R}$, $(v, \beta) \in H_{\text{epi}(f)(\Omega)}(x, f(x))$.*

Proof. The proof follows the one in [9, Proposition 2.9.3]. Among other changes one has to replace $\text{epi } f$ by $\text{epi}(f)(\Omega)$, the epigraph of f restricted to Ω .

First we assume that $(v, \beta) \in H_{\text{epi}(f)(\Omega)}(x, f(x))$. From the definition of hypertangency, there exists an $\epsilon > 0$ such that

$$x' + tv' \in \Omega \quad \text{for all } x' \in \Omega \cap B(x; \epsilon), \quad v' \in B(v; \epsilon), \quad \text{and } t \in (0, \epsilon)$$

and

$$\alpha + t\beta' \geq f(x' + tv')$$

for all $\alpha \geq f(x')$, $\alpha \in (f(x) - \epsilon, f(x) + \epsilon)$, $\beta' \in (\beta - \epsilon, \beta + \epsilon)$, and, again, $t \in (0, \epsilon)$. From the latter condition we immediately infer that $f^+(x; v)$ is bounded by $\beta + \epsilon$ and thus finite.

Now let us prove the other implication and assume that the function f is directionally Lipschitz at x with respect to $v \in H_\Omega(x)$. Let β be any number such that $\beta - \epsilon' \geq f^+(x; v)$ for some $\epsilon' > 0$. Then, there exists $\epsilon \in (0, \epsilon')$ such that, for all $x' \in \Omega \cap B(x; \epsilon)$, $v' \in B(v; \epsilon)$, $\alpha \geq f(x')$, $\alpha \in (f(x) - \epsilon, f(x) + \epsilon)$, $\beta' \in (\beta - \epsilon, \beta + \epsilon)$, and $t \in (0, \epsilon)$, one has (note that $\beta - \epsilon' < \beta - \epsilon < \beta'$)

$$x' + tv' \in \Omega, \quad \frac{f(x' + tv') - \alpha}{t} \leq \beta',$$

which proves that $(v, \beta) \in H_{\text{epi}(f)(\Omega)}(x, f(x))$. ■

Let us recall that, following the notation in [9, Section 2.9], $D_{f,\Omega}(x)$ denotes the set of vectors in $H_\Omega(x)$ with respect to which f is directionally Lipschitz. Finally, we prove the results needed for Theorems 3.1 and 3.2.

Theorem A.2 *Let f be an extended-real-valued function and x a point in Ω with $f(x) < +\infty$. Suppose $D_{f,\Omega}(x) \neq \emptyset$. Then,*

$$D_{f,\Omega}(x) = \text{int}(\{w \in H_\Omega(x) : f^\uparrow(x; w) < +\infty\}) \quad (8)$$

and $f^\uparrow(x; \cdot)$ is continuous and coincides with $f^+(x; \cdot)$ and $f_R^\circ(x; \cdot)$ in $D_{f,\Omega}(x)$.
Furthermore, if $v \in T_\Omega(x)$ and $f^\uparrow(x; v) < +\infty$, then

$$f^\uparrow(x; v) = \liminf_{\substack{w \rightarrow v \\ w \in D_{f,\Omega}(x)}} f^+(x; w) = \liminf_{\substack{w \rightarrow v \\ w \in D_{f,\Omega}(x)}} f_R^\circ(x; w). \quad (9)$$

Proof. Part of the proof follows the lines of [9, Theorem 2.9.5]. Again, one has to replace $\text{epi } f$ by $\text{epi}(f)(\Omega)$, among other quantities.

From the assumptions of the theorem and from Theorem A.1, one knows that $H_{\text{epi}(f)(\Omega)}(x, f(x)) \neq \emptyset$. We also know that $D_{f,\Omega}(x)$ is the projection of $H_{\text{epi}(f)(\Omega)}(x, f(x))$ on the set of directions \mathbb{R}^n , i.e.,

$$D_{f,\Omega}(x) = P_{\mathbb{R}^n}[H_{\text{epi}(f)(\Omega)}(x, f(x))].$$

Thus, let us apply Proposition A.2 with $C = \text{epi}(f)(\Omega)$ and $y = (x, f(x))$. As a result, we can say that

$$D_{f,\Omega}(x) = P_{\mathbb{R}^n}[\text{int}(T_{\text{epi}(f)(\Omega)}(x, f(x)))].$$

Let us assume first that w is in $P_{\mathbb{R}^n}[\text{int}(T_{\text{epi}(f)(\Omega)}(x, f(x)))]$, which implies the existence of a $\beta \in \mathbb{R}$ such that $(w, \beta) \in \text{int}(T_{\text{epi}(f)(\Omega)}(x, f(x)))$. From (7), $(w, \beta) \in \text{int}(\text{epi}(f^\uparrow(x; \cdot))(T_\Omega(x)))$, from which we can infer that $f^\uparrow(x; w') < +\infty$ for $w' \in H_\Omega(x)$ sufficiently close to w . We have thus proved that $w \in \text{int}\{w \in H_\Omega(x) : f^\uparrow(x; w) < +\infty\}$.

Now let us assume that w is in $\text{int}\{w \in H_\Omega(x) : f^\uparrow(x; w) < +\infty\}$. This means that there exists an $\epsilon > 0$ and a $\beta \in \mathbb{R}$ such that for all $w' \in B(w; \epsilon) \cap H_\Omega(x)$ one has $f^\uparrow(x; w') < \beta - \epsilon$. Thus, for all such $w' \in B(w; \epsilon) \cap H_\Omega(x)$ and $\beta' \in (\beta - \epsilon, \beta + \epsilon)$, one has $(w', \beta') \in \text{epi}(f^\uparrow(x; \cdot))(T_\Omega(x))$. As a result, $(w, \beta) \in \text{int}(\text{epi}(f^\uparrow(x; \cdot))(T_\Omega(x))) = \text{int}(T_{\text{epi}(f)(\Omega)}(x, f(x)))$.

At this point of the proof, we have proved (8). Now, a convex function defined on a convex set, such as $f^\uparrow(x; \cdot)$ defined on $H_\Omega(x)$, is always continuous in the interior of $\{w \in H_\Omega(x) : f^\uparrow(x; w) < +\infty\}$ as long as it is bounded above in a neighborhood of one point (for such a result see, e.g., [19, Theorem 10.1]). However, this is precisely the case because $D_{f,\Omega} \neq \emptyset$.

Now we will show that $f^\uparrow(x; \cdot) = f^+(x; \cdot) = f_R^\circ(x; \cdot)$ in $D_{f,\Omega}(x)$. This fact results from the following expressions for $w \in D_{f,\Omega}(x)$:

$$f^+(x; w) = \inf \{ \beta : (w, \beta) \in H_{\text{epi}(f)(\Omega)}(x, f(x)) \},$$

which is a corollary of Theorem A.1, and

$$f^\uparrow(x; w) = \inf \{ \beta : (w, \beta) \in T_{\text{epi}(f)(\Omega)}(x, f(x)) \},$$

which in turn comes from (7) and the fact that $f^\uparrow(x; w) < +\infty$ in $D_{f,\Omega}(x)$. Since, again, $\text{int}(T_{\text{epi}(f)(\Omega)}(x, f(x))) = H_{\text{epi}(f)(\Omega)}(x, f(x))$, we obtain $f^\uparrow(x; \cdot) = f^+(x; \cdot)$ in $D_{f,\Omega}(x)$. It results then trivially that $f^\uparrow(x; \cdot) = f^+(x; \cdot) = f_R^\circ(x; \cdot)$ in $D_{f,\Omega}(x)$.

Finally, (9) is also a consequence of the above expression for $f^+(x; \cdot)$ in $D_{f,\Omega}(x)$ as well as

$$f^\uparrow(x; v) = \inf \{ \beta : (v, \beta) \in T_{\text{epi}(f)(\Omega)}(x, f(x)) \},$$

for the vector $v \in T_\Omega(x)$ of the statement of theorem (which in turn holds from (7) and the fact that $f^\uparrow(x; v) < +\infty$). To prove it, let us define

$$r(x; v) = \liminf_{\substack{w \rightarrow v \\ w \in D_{f,\Omega}(x)}} f^+(x; w).$$

First we will show that $f^\uparrow(x; v) \leq r(x; v)$. For each $w \in D_{f,\Omega}(x)$ sufficiently close to v , consider β_w^ϵ such that $(\beta_w^\epsilon, w) \in H_{\text{epi}(f)(\Omega)}(x, f(x))$ and $|f^+(x; w) - \beta_w^\epsilon|$ goes to zero with $\epsilon \rightarrow 0$. Let β_v denote the limit of β_w^ϵ when $\epsilon \rightarrow 0$ and $w \rightarrow v$. Since $(\beta_v, v) \in T_{\text{epi}(f)(\Omega)}(x, f(x))$, one has

$$f^\uparrow(x; v) \leq \beta_v \leq (\beta_v - \beta_w^\epsilon) + (\beta_w^\epsilon - f^+(x; w)) + f^+(x; w).$$

We arrive at the conclusion $f^\uparrow(x; v) \leq r(x; v)$ by taking limits in the above derived inequality. Now let us prove that $f^\uparrow(x; v) = r(x; v)$. Let β_{δ_1} be such that $(\beta_{\delta_1}, v) \in T_{\text{epi}(f)(\Omega)}(x, f(x))$ and $|f^\uparrow(x; v) - \beta_{\delta_1}|$ goes to zero as $\delta_1 \rightarrow 0$. Since $H_{\text{epi}(f)(\Omega)}(x, f(x))$ is dense in $T_{\text{epi}(f)(\Omega)}(x, f(x))$, there exists $(\beta_{\delta_1, \delta_2}, v_{\delta_2}) \in H_{\text{epi}(f)(\Omega)}(x, f(x))$ such that $\|(\beta_{\delta_1, \delta_2}, v_{\delta_2}) - (\beta_{\delta_1}, v)\|$ goes to zero when $\delta_2 \rightarrow 0$. Thus,

$$f^+(x; v_{\delta_2}) \leq \beta_{\delta_1, \delta_2} = (\beta_{\delta_1, \delta_2} - \beta_{\delta_1}) + (\beta_{\delta_1} - f^\uparrow(x; v)) + f^\uparrow(x; v),$$

which shows that $f^+(x; v_{\delta_2})$ can get arbitrarily close to $f^\uparrow(x; v)$. The proof is concluded since $f^+(x; \cdot)$ and $f_R^\circ(x; \cdot)$ coincide in $D_{f,\Omega}(x)$. ■

Acknowledgments

We are grateful to José Francisco Rodrigues (University of Lisbon) for having drawn our attention to the extension result in [17]. We also thank Boris Mordukhovich (Wayne State University) and Rui Loja Fernandes (Technical University of Lisbon) for interesting discussions on related topics. Finally, we acknowledge the reports of two anonymous referees which led to an improved version of the paper.

References

- [1] NOMADm. <http://www.gerad.ca/NOMAD/Abramson/nomadm.html>.
- [2] M. A. Abramson and C. Audet. Convergence of mesh adaptive direct search to second-order stationary points. *SIAM J. Optim.*, 17:606–619, 2006.
- [3] S. Arroyo, E. Cramer, P. Frank, and J. Sabino. Simultaneous aircraft design and airline fleet mix optimization using systems of systems optimization. 2009.
- [4] C. Audet and J. E. Dennis Jr. Analysis of generalized pattern searches. *SIAM J. Optim.*, 13:889–903, 2002.
- [5] C. Audet and J. E. Dennis Jr. Mesh adaptive direct search algorithms for constrained optimization. *SIAM J. Optim.*, 17:188–217, 2006.
- [6] C. Audet and D. Orban. Finding optimal algorithmic parameters using derivative-free optimization. *SIAM J. Optim.*, 17:642–664, 2006.
- [7] C. Bogani, M. G. Gasparo, and A. Papini. Generating set search methods for piecewise smooth problems. *SIAM J. Optim.*, 20:321–335, 2009.
- [8] A. Cascon, C. Keating, and W. Shadwick. The Omega function. Technical report, The Finance Development Centre, London, 2002.
- [9] F. H. Clarke. *Optimization and Nonsmooth Analysis*. John Wiley & Sons, New York, 1983. Reissued by SIAM, Philadelphia, 1990.
- [10] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization*. MPS-SIAM Series on Optimization. SIAM, Philadelphia, 2009.
- [11] A. L. Custódio, J. E. Dennis Jr., and L. N. Vicente. Using simplex gradients of nonsmooth functions in direct search methods. *IMA J. Numer. Anal.*, 28:770–784, 2008.
- [12] N. J. Higham. Optimization by direct search in matrix computations. *SIAM J. Matrix Anal. Appl.*, 14:317–333, 1993.
- [13] N. J. Higham. *Accuracy and Stability of Numerical Algorithms*. SIAM, Philadelphia, second edition, 2002.
- [14] J. Jahn. *Introduction to the Theory of Nonlinear Optimization*. Springer-Verlag, Berlin, 1996.

- [15] S. J. Kane, M. C. Bartholomew-Biggs, M. Cross, and M. Dewar. Optimizing Omega. *J. Global Optim.*, 45:153–167, 2009.
- [16] T. G. Kolda, R. M. Lewis, and V. Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Rev.*, 45:385–482, 2003.
- [17] E. J. McShane. Extension of range of functions. *Bull. Amer. Math. Soc.*, 40:837–842, 1934.
- [18] B. Minsky, M. Obradovic, Q. Tang, and R. Thapar. Applying a global optimisation algorithm to fund of hedge funds portfolio optimisation. 2009.
- [19] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.
- [20] R. T. Rockafellar. Generalized directional derivatives and subgradients of nonconvex functions. *Can. J. Math.*, 32:257–280, 1980.
- [21] V. Torczon. On the convergence of pattern search algorithms. *SIAM J. Optim.*, 7:1–25, 1997.