Convergence towards a local minimum by direct search methods with a covering step

Pierre-Yves Bouchet [a, b]

Charles Audet [a, b]

Loïc Bourdin [c]

[a] *École Polytechnique de Montréal, Montréal (Qc), Canada, H3T 1J4*

[b] *GERAD, Montréal (Qc), Canada, H3T 1J4*

[c] *XLIM Research Institute, University of Limoges, 87060 France*

pierre-yves.bouchet@polymtl.ca
charles.audet@gerad.ca
loic.bourdin@unilim.fr

January 2024
Revised: July 2024
Les Cahiers du GERAD
G–2024–05

Abstract : This paper introduces a new step to the *Direct Search Method* (DSM) to strengthen its convergence analysis. By design, this so-called *covering step* may ensure that, for any refined point of the sequence of incumbent solutions generated by the resulting cDSM (*covering DSM*), the set of all evaluated trial points is dense in a neighborhood of that refined point. We prove that this additional property guarantees that all refined points are local solutions to the optimization problem. This new result holds true even for a discontinuous objective function, under a mild assumption that we discuss in details. We also provide a practical construction scheme for the covering step that works at low additional cost per iteration. Finally, we show that the covering step may be adapted to classes of algorithms differing from the DSM.

Keywords: Discontinuous optimization, nonsmooth optimization, derivative-free optimization, Direct Search Method, convergence, local solution

Acknowledgements: We thank Dr. Solène Kojtych for her insightful comments when we were investigating the properties of the revealing step. She deserves credit for Theorem 1 since she also foresaw that the revealing step could strengthen the convergence analysis of DSM. This work was funded by Audet's NSERC Canada Discovery Grant 2020–04448.

1 Introduction

Consider the optimization problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x), \tag{\textbf{P}}$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is possibly discontinuous. A few *derivative-free optimization* (DFO) methods [7, 12, 14] are studied in a discontinuous context, and usually by assuming some major additional structure on f [11, 21]. To our best knowledge, only the *Direct Search Method* (DSM) [7, Part 3] is rigorously supported by a thorough convergence analysis in the presence of discontinuities under light assumptions about f [3, 20]. The DSM addresses Problem (**P**) by generating sequences $(x^k)_{k\in\mathbb{N}}$ of incumbent solutions and $(\delta^k)_{k\in\mathbb{N}}$ of poll radii, where $(x^k, \delta^k) \in X \times \mathbb{R}^*_+$ for all $k \in \mathbb{N}$. The literature about DSM extracts *refining subsequences* from $(x^k, \delta^k)_{k\in\mathbb{N}}$ and studies their associated *refined points*. A subsequence $(x^k, \delta^k)_{k\in K^*}$, where $K^* \subseteq \mathbb{N}$ is infinite, is said to be *refining* [4, 5] if all iterations indexed by $k \in K^*$ *fail* (that is, $x^{k+1} = x^k$), $(\delta^k)_{k\in K^*}$ converges to zero and $(x^k)_{k\in K^*}$ converges to a limit x^* named the *refined point*. It is proved in [20] that, for all sets $K^* \subseteq \mathbb{N}$ indexing a refining subsequence, the corresponding refined point x^* satisfies a necessary optimality condition expressed in term of the Rockafellar derivative [19, 20] of f at x^*, provided that $(f(x^k))_{k\in K^*}$ converges to $f(x^*)$. Then, our previous work [3] extends [20] to ensure this last requirement under the assumption that all refining subsequences admit the same refined point. These results are valid for two standard classes of DSM: the *mesh*-based method [7, Part 3] and the *sufficient decrease*-based method [12, Section 7.7].

This paper has two main goals and two related auxiliary ones. The first main goal is to strengthen the above convergence analysis, via the addition of a new step to the DSM to ensure that all refined points are local minima. For this purpose, we introduce the `covering` DSM (cDSM), relying on the so-called `covering` step which aims to ensure that the set of all evaluated trial points is dense in a neighborhood of any refined point. We refer to Property 1 for details. Then, we prove that Property 1 implies that, for all $K^* \subseteq \mathbb{N}$ indexing a refining subsequence with refined point denoted by x^*, either $x^* \in X$ is a local solution to Problem (**P**), or $x^* \notin X$ is such that $(f(x^k))_{k\in K^*}$ converges to the infimum of f over a neighborhood of x^*. This result is formalized in Theorem 1. The related auxiliary goal is to propose a practical construction scheme for the `covering` step ensuring Property 1 by design. This scheme fits in both the mesh-based cDSM and the sufficient decrease-based cDSM, and the additional cost per iteration it induces is low. Our second main goal is to show that the `covering` step is compatible with many classes of methods, to allow for a convergence analysis close to Theorem 1. Theorem 2 formalizes this study. Our related auxiliary goal is to prove that our assumption about Problem (**P**), involved in Theorems 1 and 2, cannot be relaxed in general and is weaker than in former work. This paper originates from the corresponding author's PhD thesis [10, Chapter 4] (in French).

Note that the `covering` step generalizes the `revealing` step from [2, 3]. Actually, our initial motivation was to better study this `revealing` step. Its goal in [2] is to reveal local discontinuities, but we observed in [3] that, when the so-called `revealing` DSM (rDSM) generates a unique refined point, the `revealing` step provides the density of the trial points in a neighborhood of the refined point. However, [2, 3] fail to deduce the local optimality of the refined point. Thus, the current work originally aimed to state this property. Yet, we eventually found that the `revealing` step admits a generalization providing the density of the set of trial points around an arbitrary number of refined points. The formalization of this generalization and the study of its properties constitute the core of the present work. Note that we decided to change the terminology because, in comparison with the name `revealing`, we believe that the name `covering` better captures what this step actually does.

This paper is organized as follows. Section 2 formalizes the cDSM and states Theorem 1. Section 3 proves Theorem 1. Section 4 provides a construction scheme for the `covering` step. Section 5 states and proves Theorem 2. Section 6 discusses our assumption about Problem (**P**). Section 7 discusses our work and its possible extensions. In addition, Appendix A contains supplementary materials and Appendix B contains proofs of some auxiliary results.

Notation: We denote by $\mathbb{S}^n$ the unit sphere of $\mathbb{R}^n$. For all $r \in \mathbb{R}_+^*$ and all $x \in \mathbb{R}^n$, we denote by $\mathcal{B}_r(x)$ the open ball of radius r centered at x, and by $\mathcal{B}_r \triangleq \mathcal{B}_r(0)$. For all $x \in \mathbb{R}^n$ and all $\mathcal{S} \subseteq \mathbb{R}^n$, we denote by $\{x\} + \mathcal{S} \triangleq \{x + s : s \in \mathcal{S}\}$, $f(\mathcal{S}) \triangleq \{f(s) : s \in \mathcal{S}\}$ and $\mathrm{dist}(x, \mathcal{S}) \triangleq \inf\{\|x - s\| : s \in \mathcal{S}\}$. For all $\mathcal{S} \subseteq \mathbb{R}^n$, $\mathcal{S}$ is said to be *ample* if $\mathcal{S} \subseteq \mathrm{cl}(\mathrm{int}(\mathcal{S}))$, or *locally thin* if there exists $\mathcal{N} \subseteq \mathbb{R}^n$ open so that $\mathcal{S} \cap \mathcal{N} \neq \emptyset = \mathrm{int}(\mathcal{S}) \cap \mathcal{N}$. Note that a set is either ample or locally thin (see Proposition 7), and that the definition of an ample set is equivalent to that of a *semi-open* set introduced in [16]. For all $x \in \mathbb{R}^n$ and all $\mathcal{S} \subseteq \mathbb{R}^n$ nonempty and closed, we denote by $\mathrm{round}(x, \mathcal{S}) \triangleq \mathrm{argmin}\, \mathrm{dist}(x, \mathcal{S})$. For all collections $(\mathcal{S}_i)_{i=1}^N$ of subsets of $\mathbb{R}^n$ (with $N \in \mathbb{N}^* \cup \{+\infty\}$), their union is denoted by $\sqcup_{i=1}^N \mathcal{S}_i$ when the sets are pairwise disjoint. Finally, for all $(a, b) \in \mathbb{Z}^2$, we denote by $[\![a, b]\!] \triangleq \mathbb{Z} \cap [a, b]$.

2 Formal `covering` step and convergence result of `cDSM`

This section formalizes the `cDSM` as Algorithm 1 and its convergence analysis in Theorem 1. Theorem 1 is based on Property 1 provided by the `covering` step, and on Assumption 1 regarding the objective function f. The proof of Theorem 1 follows in Section 3.

The `cDSM` matches the usual `DSM` in most of its aspects. The only novelty lies in the `covering` step, with its parameter $r \in \mathbb{R}_+^*$, detailed in Section 4. Our choices for the other steps and parameters follow simple instances of mesh-based `DSM` [7, Part 3] and sufficient decrease-based `DSM` [12, Section 7.7] respectively, but they may be designed as in many such `DSM`. We discuss the `cDSM` in Remark 1.

Algorithm 1 `cDSM` (`covering DSM`) **solving Problem (P).**

Initialization:
- set a covering radius $r \in \mathbb{R}_+^*$; set the trial points history $\mathcal{H}^0 \triangleq \emptyset$;
- set the incumbent solution and poll radius $(x^0, \delta^0) \in X \times \mathbb{R}_+^*$; set $\underline{\delta}^0 \triangleq \delta^0$;
- set $\tau \in]0, 1[\cap \mathbb{Q}$, and set $\mathcal{M} : \mathbb{R}_+ \to 2^{\mathbb{R}^n}$ and $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ according to one of
  - the *mesh-based DSM*: $\mathcal{M}(\nu) \triangleq \min\{\nu, \frac{\nu^2}{\delta^0}\}\mathbb{Z}^n$ and $\rho(\nu) \triangleq 0$ for all $\nu \in \mathbb{R}_+$;
  - the *sufficient decrease-based DSM*: $\mathcal{M}(\nu) \triangleq \mathbb{R}^n$ and $\rho(\nu) \triangleq \min\{\nu, \frac{\nu^2}{\delta^0}\}$ for all $\nu \in \mathbb{R}_+$.

For $k \in \mathbb{N}$ **do**:
- `covering` **step**:
  - set $\mathcal{D}_{\mathtt{C}}^k \subseteq \mathcal{M}(\underline{\delta}^k) \cap \mathrm{cl}(\mathcal{B}_r)$ nonempty and finite; set $\mathcal{T}_{\mathtt{C}}^k \triangleq \{x^k\} + \mathcal{D}_{\mathtt{C}}^k$; set $t_{\mathtt{C}}^k \in \mathrm{argmin}\, f(\mathcal{T}_{\mathtt{C}}^k)$;
  - if $f(t_{\mathtt{C}}^k) < f(x^k) - \rho(\underline{\delta}^k)$, then set $t^k \triangleq t_{\mathtt{C}}^k$ and $\mathcal{T}_{\mathtt{S}}^k = \mathcal{T}_{\mathtt{P}}^k \triangleq \emptyset$ and skip to the `update` step;
- `search` **step**:
  - set $\mathcal{D}_{\mathtt{S}}^k \subseteq \mathcal{M}(\underline{\delta}^k)$ empty or finite; if $\mathcal{T}_{\mathtt{S}}^k \triangleq \{x^k\} + \mathcal{D}_{\mathtt{S}}^k$ is nonempty, then set $t_{\mathtt{S}}^k \in \mathrm{argmin}\, f(\mathcal{T}_{\mathtt{S}}^k)$;
  - if also $f(t_{\mathtt{S}}^k) < f(x^k) - \rho(\underline{\delta}^k)$, then set $t^k \triangleq t_{\mathtt{S}}^k$ and $\mathcal{T}_{\mathtt{P}}^k \triangleq \emptyset$ and skip to the `update` step;
- `poll` **step**:
  - set $\mathcal{D}_{\mathtt{P}}^k \subseteq \mathcal{M}(\underline{\delta}^k) \cap \mathrm{cl}(\mathcal{B}_{\delta^k})$ a positive basis of $\mathbb{R}^n$; set $\mathcal{T}_{\mathtt{P}}^k \triangleq \{x^k\} + \mathcal{D}_{\mathtt{P}}^k$; set $t_{\mathtt{P}}^k \in \mathrm{argmin}\, f(\mathcal{T}_{\mathtt{P}}^k)$;
  - if $f(t_{\mathtt{P}}^k) < f(x^k) - \rho(\underline{\delta}^k)$, then set $t^k \triangleq t_{\mathtt{P}}^k$, otherwise set $t^k \triangleq x^k$;
- `update` **step**:
  - set $\mathcal{H}^{k+1} \triangleq \mathcal{H}^k \cup \mathcal{T}^k$, where $\mathcal{T}^k \triangleq \mathcal{T}_{\mathtt{C}}^k \cup \mathcal{T}_{\mathtt{S}}^k \cup \mathcal{T}_{\mathtt{P}}^k$; set $x^{k+1} \triangleq t^k$;
  - set $\delta^{k+1} \triangleq \frac{1}{\tau}\delta^k$ if $x^k \neq t^k$ and $\delta^{k+1} \triangleq \tau\delta^k$ otherwise; set $\underline{\delta}^{k+1} \triangleq \min\{\underline{\delta}^k, \delta^{k+1}\}$.

Algorithm 1 generates a sequence $(x^k, \delta^k, \mathcal{H}^k)_{k \in \mathbb{N}}$ from which we may define

$$\mathcal{R} \triangleq \left\{ \lim_{k \in K^*} x^k : K^* \subseteq \mathbb{N} \text{ indexes a refining subsequence of } (x^k, \delta^k)_{k \in \mathbb{N}} \right\} \qquad \text{and} \qquad \mathcal{H} \triangleq \bigcup_{k \in \mathbb{N}} \mathcal{H}^k,$$

and a carefully constructed `covering` step ensures that these two sets satisfy the next Property 1. A practical construction scheme to indeed meet this property is given in Section 4.3.

Property 1 (Dense covering provided by Algorithm 1 with a well designed `covering` step)**.** The inclusion

$$\mathcal{B}_r(x^*) \subseteq \mathrm{cl}(\mathcal{H})$$

holds for all $x^* \in \mathcal{R}$, where $r \in \mathbb{R}_+^*$ is the `covering` radius chosen in the **initialization** step.

Theorem 1 also requires the following assumptions about f. Assumptions 1.a) and 1.b) match usual assumptions used in the literature, while the unusual Assumption 1.c) is discussed in Section 6.

Assumption 1 (on the objective function f)**.** The objective function f in Problem (**P**) is such that
a) f is bounded below and has bounded sublevel sets;
b) the restriction $f_{|X} : X \to \mathbb{R}$ is lower semicontinuous;
c) X admits a partition $X = \sqcup_{i=1}^{N} X_i$ (where $N \in \mathbb{N}^* \cup \{+\infty\}$) such that, for all $i \in [\![1, N]\!]$, X_i is an *ample continuity set of* f (that is, X_i is ample and the restriction $f_{|X_i} : X_i \to \mathbb{R}$ is continuous).

Theorem 1. Under Assumption 1, Algorithm 1 generates at least one refining subsequence and, if Property 1 holds, then, for all $K^* \subseteq \mathbb{N}$ indexing a refining subsequence, the corresponding refined point x^* satisfies

$$\lim_{k \in K^*} f(x^k) \;=\; \left\{ \begin{array}{ll} \min f(\mathcal{B}_r(x^*)) = f(x^*) & \text{if } x^* \in X, \\ \inf f(\mathcal{B}_r(x^*)) & \text{if } x^* \notin X. \end{array} \right.$$

We conclude this section with two remarks about the `cDSM` and Theorem 1 respectively.

Remark 1. First, for ease of presentation, Algorithm 1 allows little freedom in the designs of the mesh $\mathcal{M}$, the sufficient decrease function ρ and the rule to update the poll radius δ^k. Nevertheless, their designs may follow any mesh-based `DSM` [7, Part 3] and any sufficient decrease-based `DSM` [12, Section 7.7]. A variant of Algorithm 1 with the generic definition of all these elements appears as Algorithm 3 in Appendix A.1. Theorem 1 remains valid when Algorithm 1 is replaced by Algorithm 3. Second, the mesh is defined in the space of directions in Algorithm 1, while in the literature it is usually defined in the space of variables directly, but centered at the incumbent solution. This variation simplifies our presentation and is transparent. Third, the `covering` step fits in the framework of the `search` step. Consequently, the `cDSM` inherits all the properties of the usual `DSM`. Last, we stress that the `cDSM` relies on the smallest poll radius $\underline{\delta}^k$ in addition to the current poll radius δ^k. This dependency in $(\underline{\delta}^k)_{k \in \mathbb{N}}$ is mandatory in the proof of Theorem 1 (precisely, in Proposition 2).

Remark 2. When Assumption 1 holds, Property 1 ensures that all refined points are local solutions to Problem (**P**), in the usual sense for those lying in X and in a generalized sense for the others. In practice, most `DSM` usually generate exactly one refined point which is moreover a local solution to Problem (**P**). Nevertheless, to our best knowledge, no `DSM` from the literature ensures Property 1 (and thus satisfies Theorem 1). Then, at the time of writing, the only `DSM` that is guaranteed to generate a local solution under Assumption 1 is the `cDSM` relying the `covering` step defined in Section 4.

3 Proof of Theorem 1

Preliminary results. The `covering` step is a specific `search` step, thus Algorithm 1 inherits all the properties of the usual `DSM`. Hence, the next Proposition 1 holds. It is stated as [7, Theorem 8.1] for the mesh-based `DSM` and as [12, Corollary 7.2] for the sufficient decrease-based `DSM`.

Proposition 1. Under Assumption 1.a), Algorithm 1 generates at least one refining subsequence.

The rest of this section relies on Proposition 2, which shows what Property 1 provides and settles the ground for the proof of Theorem 1. The proof of Proposition 2 involves an auxiliary topological claim that is proved in Proposition 8 in Appendix B.

Proposition 2. Under Assumption 1, if Algorithm 1 satisfies Property 1, then, for all refined points x^*, it holds that $\lim_{k \in \mathbb{N}} f(x^k) \leq f(x)$ for all $x \in \mathcal{B}_r(x^*)$.

Proof. First, $f^* \triangleq \lim_{k \in \mathbb{N}} f(x^k) \in \mathbb{R}$ exists since $(f(x^k))_{k \in \mathbb{N}}$ decreases by construction and is bounded below by Assumption 1.a). Second, let x^* be a refined point generated by Algorithm 1 satisfying Property 1 and take $x \in \mathcal{B}_r(x^*)$. The result holds if $x \notin X$, so assume that $x \in X_i$ for some $i \in [\![1, N]\!]$. Let $(y_\ell)_{\ell \in \mathbb{N}}$ converging to x with $y_\ell \in \mathcal{B}_r(x^*) \cap X_i \cap \mathcal{H} \setminus \{x\}$ for all $\ell \in \mathbb{N}$ (it exists by Proposition 8.d) applied with $\mathcal{S}_1 \triangleq \mathcal{B}_r(x^*)$, $\mathcal{S}_2 \triangleq X_i$ and $\mathcal{S}_3 \triangleq \mathcal{H}$). Let $\kappa(\ell) \triangleq \min\{k \in \mathbb{N} : y_\ell \in \mathcal{T}^k\}$ for all $\ell \in \mathbb{N}$. Then $(\kappa(\ell))_{\ell \in \mathbb{N}}$ diverges to $+\infty$, since $\mathcal{T}^k$ is finite for all $k \in \mathbb{N}$ and every subsequence of $(y_\ell)_{\ell \in \mathbb{N}}$ takes infinitely many values (since $(y_\ell)_{\ell \in \mathbb{N}}$ converges to x with $y_\ell \neq x$ for all $\ell \in \mathbb{N}$). Moreover, $(\underline{\delta}^{\kappa(\ell)})_{\ell \in \mathbb{N}}$ converges to 0 since $\liminf_{k \in \mathbb{N}} \delta^k = 0$ as a consequence of Proposition 1. In addition, for all $\ell \in \mathbb{N}$ we have $f(t^{\kappa(\ell)}) \geq f(x^{\kappa(\ell)}) - \rho(\underline{\delta}^{\kappa(\ell)})$ if iteration $\kappa(\ell)$ fails and $f(t^{\kappa(\ell)}) = f(x^{\kappa(\ell)+1})$ otherwise, and $f(y_\ell) \geq f(t^{\kappa(\ell)})$ by construction. Hence, for all $\ell \in \mathbb{N}$ we have $f(y_\ell) \geq f(x^{\kappa(\ell)+1}) - \rho(\underline{\delta}^{\kappa(\ell)})$. The result follows by taking $\ell \to +\infty$, since $(f(y_\ell))_{\ell\in\mathbb{N}}$ converges to $f(x)$ by continuity of $f_{|X_i}$ and $(f(x^{\kappa(\ell)+1}))_{\ell\in\mathbb{N}}$ converges to f^* and $(\rho(\underline{\delta}^{\kappa(\ell)}))_{\ell\in\mathbb{N}}$ converges to 0. □

Proof of Theorem 1. Consider that Assumption 1 is satisfied. Proposition 1 states that at least one refining subsequence is generated. Assume that Property 1 holds. Let $K^* \subseteq \mathbb{N}$ indexing a refining subsequence, x^* denoting its refined point, and let $f^* \triangleq \lim_{k\in K^*} f(x^k) = \lim_{k\in\mathbb{N}} f(x^k)$. Let us show that $f^* = \inf f(\mathcal{B}_r(x^*))$. We have $f^* \geq \inf f(\mathcal{B}_r(x^*))$ since $x^k \in \mathcal{B}_r(x^*)$ for all $k \in K^*$ large enough, and $f^* \leq \inf f(\mathcal{B}_r(x^*))$ is proved by contradiction: if $f^* > \inf f(\mathcal{B}_r(x^*))$, then there exists $x^\sharp \in \mathcal{B}_r(x^*)$ such that $f^* > f(x^\sharp)$, but then $f^* > f(x^\sharp) \geq f^*$ by Proposition 2. Note that this already concludes the proof of the case where $x^* \notin X$. Now assume that $x^* \in X$. Then $f^* \geq f(x^*)$ by Assumption 1.b) and $f^* \leq f(x^*)$ by Proposition 2, so $f(x^*) = f^* = \inf f(\mathcal{B}_r(x^*)) = \min f(\mathcal{B}_r(x^*))$. □

4 Discussion on the `covering` step and Property 1

This section discusses the `covering` step. Section 4.1 highlights the differences between the `covering` step and the `revealing` step it is inspired from. Section 4.2 provides a sufficient condition about a construction scheme for the `covering` step to ensure Property 1. Section 4.3 provides a tractable construction scheme for the `covering` step which checks this sufficient condition.

4.1 The `revealing` step provides Property 1 a posteriori in some cases only

As stated in Section 1, the `covering` step is inspired from the `revealing` step in a `DSM` that we hereafter call the `revealing DSM` (`rDSM`) [3, Algorithm 1]. The `rDSM` satisfies Property 1 when exactly one refined point is generated [3, Lemma 2]. Nevertheless, we show in this section that the `rDSM` may fail to ensure Property 1 when more than one refined point is generated.

Expressed in our notation, the `revealing` step in [3] relies on a sequence $(\mathcal{D}_\ell)_{\ell\in\mathbb{N}}$ with $\mathcal{D}_\ell \subset \mathrm{cl}(\mathcal{B}_r)$ finite for all $\ell \in \mathbb{N}$ and such that $\mathcal{B}_r \subseteq \mathrm{cl}(\cup_{\ell\in\mathbb{N}}\mathcal{D}_\ell)$. For all $k \in \mathbb{N}$, it defines $\mathcal{D}^k_{\mathtt{C}} \triangleq \mathrm{round}(\mathcal{D}_{\ell(k)}, \mathcal{M}(\underline{\delta}^k))$ as the rounding of $\mathcal{D}_{\ell(k)}$ onto $\mathcal{M}(\underline{\delta}^k)$, where $\ell(k)$ denotes the number of iterations indices $u < k$ such that $\underline{\delta}^{u+1} < \underline{\delta}^u$. Thus, the `revealing` step ensures that

$$\mathcal{B}_r \subseteq \mathrm{cl}\left(\underset{k\in\mathbb{N}}{\cup} \mathcal{D}^k_{\mathtt{C}}\right)$$

when the `rDSM` generates at least one refining subsequence, which is certified a priori by Proposition 1. Nevertheless, this property states only the dense intersection of the trial directions history with $\mathcal{B}_r$. This may not lead to Property 1 when the `rDSM` generates more than one refined point. The following Example 1, illustrated in Figure 1, confirms this observation.

Example 1. Consider the objective function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x) \triangleq \|x\|_\infty$ for all $x \in \mathbb{R}^2$ and the algorithmic parameters $r \triangleq 1$, $x^0 \triangleq -3\mathbb{1}$ where $\mathbb{1} \triangleq (1,1)$, $\delta^0 \triangleq 1$, $\mathcal{M}(\nu) \triangleq \min\{\nu,\nu^2\}\mathbb{Z}^2$ and $\rho(\nu) \triangleq 0$ for all $\nu \in \mathbb{R}^*_+$, $\Lambda \triangleq \{\frac{1}{2}\}$ and $\Upsilon \triangleq \{1\}$. Define the `poll` step as $\mathcal{D}^k_{\mathtt{P}} \triangleq \{(\pm\delta^k, 0), (0, \pm\delta^k)\}$ for all $k \in \mathbb{N}$. Define the `revealing` step such that the sequence $(\mathcal{D}_\ell)_{\ell\in\mathbb{N}}$ satisfies

$$\forall q \in \mathbb{N}, \qquad \mathcal{D}_{4q} \subset \mathbb{R}_+ \times \mathbb{R}_+, \quad \mathcal{D}_{4q+1} \subset \mathbb{R}_- \times \mathbb{R}_+, \quad \mathcal{D}_{4q+2} \subset \mathbb{R}_- \times \mathbb{R}_-, \quad \mathcal{D}_{4q+3} \subset \mathbb{R}_+ \times \mathbb{R}_-.$$

Define the `search` step at each iteration $k \in \mathbb{N}$ so that $\mathcal{T}^k_{\mathtt{S}} \triangleq \emptyset$ if $k \notin 3\mathbb{N}$, and so that $\mathcal{T}^{3q}_{\mathtt{S}} \triangleq \{t^{3q}_{\mathtt{S}}\}$ for all $k = 3q \in 3\mathbb{N}$, where

$$t^{3q}_{\mathtt{S}} \triangleq (-1)^q \left(1 + 2^{-q}\right) \mathbb{1}.$$

This instance of `rDSM` has a predictable behavior. Using induction, one can show that

$$\forall q \in \mathbb{N}, \quad \left\{ \begin{array}{llll} x^{3q} = (-1)^{q-1}\left(1+2^{-(q-1)}\right)\mathbb{1}, & \delta^{3q} = 4^{-q}, & \mathcal{D}_{\mathtt{C}}^{3q} = \mathcal{D}_{2q}, & \texttt{search}\text{ success,} \\ x^{3q+1} = (-1)^{q}\quad\left(1+2^{-q}\right)\mathbb{1}, & \delta^{3q+1} = 4^{-q}, & \mathcal{D}_{\mathtt{C}}^{3q+1} = \mathcal{D}_{2q}, & \text{iteration fails,} \\ x^{3q+2} = (-1)^{q}\quad\left(1+2^{-q}\right)\mathbb{1}, & \delta^{3q+2} = \frac{1}{2}4^{-q}, & \mathcal{D}_{\mathtt{C}}^{3q+2} = \mathcal{D}_{2q+1}, & \text{iteration fails.} \end{array} \right.$$

Thus, there are two refined points $x_+^* = \mathbb{1}$ and $x_-^* = -\mathbb{1}$, and none is a local minimizer of f. However, this does not contradict Theorem 1 because Property 1 is not satisfied. Indeed, it holds that

$$\mathrm{cl}(\mathcal{H}) \;\cap\;]-1,1[^{\,2} = \emptyset \quad \text{while} \quad \forall r \in \mathbb{R}_+^*,\; \mathcal{B}_r(x_+^*) =]1-r,1+r[^{\,2} \text{ and } \mathcal{B}_r(x_-^*) =]-1-r,-1+r[^{\,2}.$$

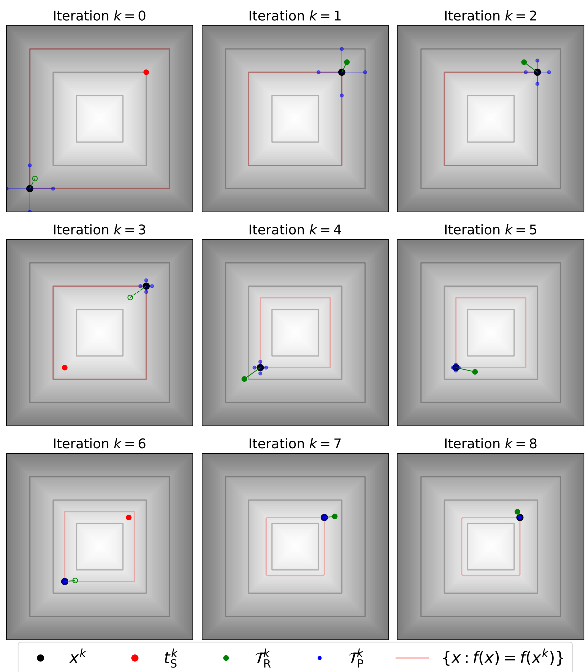

Figure 1: First eight iterations of the `rDSM` in the context of Example 1. The set $\mathcal{T}_{\mathtt{R}}^k$ of `revealing` trial points is dotted at each iteration $k \in 3\mathbb{N}$ to highlight that these points are actually not evaluated by the `rDSM`. Indeed, the `search` step trial point $t_{\mathtt{S}}^k$ is evaluated before and makes the iteration a success.

Example 1 shows that the `revealing` step from [3] focuses only on the asymptotic density of the trial directions, and that this may not translate to the density of the trial points in some neighborhoods of the refined points. The original instance of the `revealing` step provided in [2] differs from those in [3] (in [2], the `revealing` step draws at each iteration a direction in $\mathrm{cl}(\mathcal{B}_r)$ according to the independent uniform distribution), but it also fails to provide Property 1 when more than one refined point exists. Accordingly, in Section 4.2, we study schemes for the `covering` step that ensure instead the dense intersection of the trial points with a neighborhood of all refined points.

4.2 Sufficient condition to ensure a priori that the `cDSM` satisfies Property 1

This section focuses on `covering` step instances relying only, at each iteration $k \in \mathbb{N}$, on the current couple $(x^k, \underline{\delta}^k)$ and the current history $\mathcal{H}^k \triangleq \cup_{\ell<k}\mathcal{T}^\ell$ (where $\mathcal{T}^\ell$ denotes the set of all trial points evaluated at iteration $\ell \in \mathbb{N}$). Proposition 3 proves that, if the `covering` step relies on a `covering` oracle from Definition 1, illustrated in Figure 2, then all executions of the `cDSM` satisfy Property 1.

Definition 1 (`covering` **oracle).** For all $r \in \mathbb{R}_+^*$, a function $\mathbb{O} : \mathbb{R}^n \times 2^{\mathbb{R}^n} \to 2^{\mathbb{R}^n}$ is said to be a *`covering` oracle* of radius r if there exists $\beta \in]0,1]$ such that, for all points $x \in \mathbb{R}^n$ and all sets $\emptyset \neq \mathcal{S} \subseteq \mathbb{R}^n$,

$$\mathbb{O}(x,\mathcal{S}) \text{ is compact and } \emptyset \neq \mathbb{O}(x,\mathcal{S}) \subseteq \mathrm{cl}(\mathcal{B}_r) \qquad \text{and} \qquad \frac{\max_{d\in\mathbb{O}(x,\mathcal{S})} \mathrm{dist}\,(x+d,\mathcal{S})}{\max_{d\in\mathrm{cl}(\mathcal{B}_r)} \mathrm{dist}\,(x+d,\mathcal{S})} \;\geq\; \beta.$$

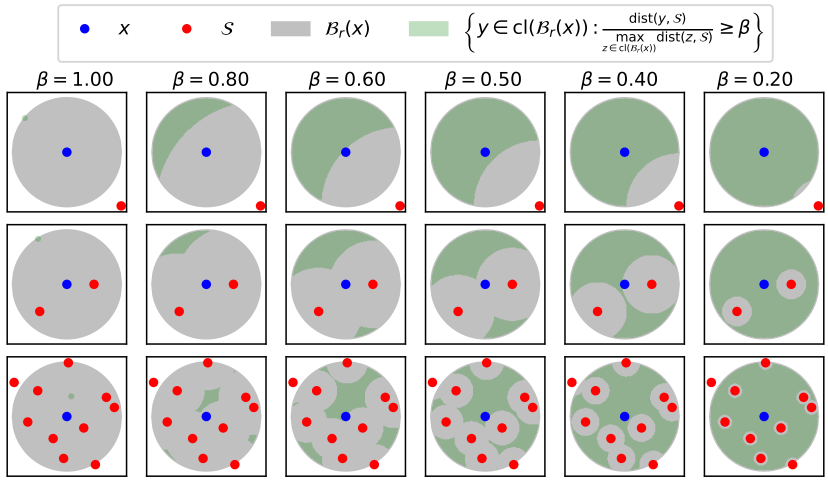

Figure 2: Illustration of Definition 1. A function $\mathbb{O}$ satisfies Definition 1 at $(x, \mathcal{S})$ for the value β if $\mathbb{O}(x, \mathcal{S})$ contains at least one direction d such that $x + d$ lies in the set represented in green.

Proposition 3. For all functions $\mathbb{O}$ satisfying Definition 1, if Algorithm 1 defines its `covering` step as

$$\forall k \in \mathbb{N}, \qquad \mathcal{D}_{\mathtt{C}}^k \triangleq \text{round}\left(\mathbb{O}\left(x^k, \mathcal{H}^k\right), \mathcal{M}(\underline{\delta}^k)\right),$$

then it ensures Property 1.

Proof. Consider the framework of Proposition 3. Let $x^* \in \mathcal{R}$ and $K^* \subseteq \mathbb{N}$ indexing an associated refining subsequence, such that $\max_{d\in\mathbb{R}^n} \|d - \text{round}(d, \mathcal{M}(\underline{\delta}^k))\| \leq r$ and $x^k \in \text{cl}(\mathcal{B}_r(x^*))$ for all $k \in K^*$. For all $k \in \mathbb{N}$, let $d_{\mathbb{O}}^k \in \mathbb{O}(x^k, \mathcal{H}^k)$ such that $\text{dist}(x^k + d_{\mathbb{O}}^k, \mathcal{H}^k) = \max_{d\in\mathbb{O}(x^k,\mathcal{H}^k)} \text{dist}(x^k + d, \mathcal{H}^k)$. Let also $d_{\mathtt{C}}^k \triangleq \text{round}(d_{\mathbb{O}}^k, \mathcal{M}(\underline{\delta}^k)) \in \mathcal{D}_{\mathtt{C}}^k$ and $t_{\mathtt{C}}^k \triangleq x^k + d_{\mathtt{C}}^k \in \mathcal{T}_{\mathtt{C}}^k \subseteq \mathcal{T}^k$. Observe that, when $k \to +\infty$ in K^*,

$$\left\|x^* - x^k\right\| \to 0, \qquad \left\|d_{\mathbb{O}}^k - d_{\mathtt{C}}^k\right\| \to 0 \qquad \text{and} \qquad \min_{\ell<k} \left\|t_{\mathtt{C}}^k - t_{\mathtt{C}}^\ell\right\| \to 0.$$

The first two claims result from K^* indexing a refining subsequence. The third claim is proved by contradiction. Assume that there exists $K \subseteq K^*$ infinite and $\varepsilon > 0$ such that $\min_{\ell<k} \left\|t_{\mathtt{C}}^k - t_{\mathtt{C}}^\ell\right\| \geq \varepsilon$ for all $k \in K$. Then, $\|t_{\mathtt{C}}^k - t_{\mathtt{C}}^\ell\| \geq \varepsilon$ for all $(k,\ell) \in K^2$ with $\ell < k$. Nevertheless, for all $k \in K$, it holds that $\|t_{\mathtt{C}}^k - x^*\| = \|x^k + d_{\mathtt{C}}^k + d_{\mathbb{O}}^k - d_{\mathbb{O}}^k - x^*\| \leq \|x^k - x^*\| + \|d_{\mathbb{O}}^k\| + \|d_{\mathtt{C}}^k - d_{\mathbb{O}}^k\| \leq 3r$, so all elements of $(t_{\mathtt{C}}^k)_{k\in K}$ lie in the compact set $\text{cl}(\mathcal{B}_{3r}(x^*))$. This situation contradicts the Bolzano-Weierstrass theorem applied to $(t_{\mathtt{C}}^k)_{k\in K}$. Then, to conclude the proof, remark that, for all $k \in K^*$,

$$\begin{aligned}
\max_{d\in\mathcal{B}_r} \text{dist}\left(x^* + d, \mathcal{H}\right) &\leq \max_{d\in\text{cl}(\mathcal{B}_r)} \text{dist}\left(x^* + d, \mathcal{H}^k\right) \\
&\leq \left\|x^* - x^k\right\| + \max_{d\in\text{cl}(\mathcal{B}_r)} \text{dist}\left(x^k + d, \mathcal{H}^k\right) \\
&\leq \left\|x^* - x^k\right\| + \frac{1}{\beta} \max_{d\in\mathbb{O}(x^k,\mathcal{H}^k)} \text{dist}\left(x^k + d, \mathcal{H}^k\right) \\
&= \left\|x^* - x^k\right\| + \frac{1}{\beta} \text{dist}\left(x^k + d_{\mathbb{O}}^k, \mathcal{H}^k\right) \\
&\leq \left\|x^* - x^k\right\| + \frac{1}{\beta}\left(\left\|d_{\mathbb{O}}^k - d_{\mathtt{C}}^k\right\| + \text{dist}\left(x^k + d_{\mathtt{C}}^k, \mathcal{H}^k\right)\right) \\
&= \left\|x^* - x^k\right\| + \frac{1}{\beta}\left\|d_{\mathbb{O}}^k - d_{\mathtt{C}}^k\right\| + \frac{1}{\beta} \text{dist}\left(t_{\mathtt{C}}^k, \underset{\ell<k}{\cup} \mathcal{T}^\ell\right) \\
&\leq \left\|x^* - x^k\right\| + \frac{1}{\beta}\left\|d_{\mathbb{O}}^k - d_{\mathtt{C}}^k\right\| + \frac{1}{\beta} \min_{\ell<k} \left\|t_{\mathtt{C}}^k - t_{\mathtt{C}}^\ell\right\|.
\end{aligned}$$

Then, taking the limit as $k \to +\infty$ in K^* leads to the desired result that $\mathcal{B}_r(x^*) \subseteq \text{cl}(\mathcal{H})$. □

Proposition 3 depends only on the oracle $\mathbb{O}$ driving the `covering` step, so it may be checked a priori (prior to the execution of the `cDSM`). In contrast, Property 1 must be checked a posteriori, as this requires the whole sequence $(x^k, \delta^k, \mathcal{H}^k)_{k\in\mathbb{N}}$ to compute $\mathcal{R}$ and $\mathcal{H}$ and check Property 1. Thus, Proposition 3 allows for a simple way to ensure a priori that Property 1 will be satisfied. The construction scheme for the `covering` step we introduce in Section 4.3 relies on Definition 1 and Proposition 3.

4.3 Construction of a `covering` step instance usable in practice

In practice, an instance of the `covering` step must satisfy two criteria. For theoretical consistency, the resulting trial points history $\mathcal{H}$ must satisfy Property 1, and for practical efficiency, the number of `covering` trial points must be small at each iteration. This section proposes a scheme that meets these two criteria. We follow the guideline from Section 4.2, that is, $\mathcal{D}^k_{\mathtt{C}} \triangleq \mathrm{round}(\mathbb{O}\left(x^k, \mathcal{H}^k\right), \mathcal{M}(\underline{\delta}^k))$ at each $k \in \mathbb{N}$, where $\mathbb{O}$ is a `covering` oracle from Definition 1. Our scheme designs a tractable expression for $\mathbb{O}$, valid for all `covering` radii $r \in \mathbb{R}^*_+$.

The baseline scheme we suggest for the `covering` step relies on the following oracle:

$$\forall x \in \mathbb{R}^n, \quad \forall \mathcal{S} \subseteq \mathbb{R}^n, \qquad \mathbb{O}(x, \mathcal{S}) \triangleq \operatorname*{argmax}_{d\in\mathrm{cl}(\mathcal{B}_r)} \mathrm{dist}\left(x + d, \mathcal{S}\right). \tag{1}$$

Oracle 1 is a `covering` oracle, as it satisfies Definition 1 with $\beta \triangleq 1$. Then, Proposition 3 claims that a `covering` step instance relying on Oracle 1 ensures Property 1. At each $k \in \mathbb{N}$, this instance selects $\mathcal{D}^k_{\mathtt{C}}$ as the set of all $d^k_{\mathtt{C}} \in \mathrm{cl}(\mathcal{B}_r)$ such that $t^k_{\mathtt{C}} \triangleq x^k + d^k_{\mathtt{C}}$ is the farthest possible from the set $\mathcal{H}^k$ of all past trial points (or, in the case of the mesh-based `cDSM`, $\mathcal{D}^k_{\mathtt{C}}$ is the rounding of all these directions $d^k_{\mathtt{C}}$ onto the mesh). Moreover, Property 1 remains valid if we actually compute only one such direction. The computation of $\mathbb{O}$ is costly when $\mathcal{S}$ contains many elements, but it may be alleviated.

Computing Oracle 1 is a continuous and piecewise affine problem, for all $(x, \mathcal{S}) \in \mathbb{R}^n \times 2^{\mathbb{R}^n}$. This problem admits smooth surrogates, such as $\overline{\mathbb{O}}(x, \mathcal{S}) \triangleq \mathrm{argmax}_{d\in\mathrm{cl}(\mathcal{B}_r)} \Sigma_{s\in\mathcal{S}} \frac{-1}{\|x+d-s\|}$. Moreover, the computation of $\mathbb{O}(x^k, \mathcal{H}^k)$ at each iteration $k \in \mathbb{N}^*$ may start from the point $t^{k-1}_{\mathtt{C}}$ calculated at the preceding iteration. Also, in the case of the mesh-based `cDSM`, the set $\mathbb{O}(x, \mathcal{S})$ is rounded onto the mesh, so it may be more relevant to restrict the search to direction lying on the mesh directly. For all $(x, \mathcal{S}) \in \mathbb{R}^n \times 2^{\mathbb{R}^n}$ and all mesh parameters $\nu \in \mathbb{R}^*_+$, we may seek for the directions $d \in \mathcal{M}(\nu) \cap \mathrm{cl}(\mathcal{B}_r)$ maximizing $\mathrm{dist}(x + d, \mathcal{S})$. This problem is combinatorial because of the discrete nature of the mesh, but it may be solved using a *distance transform algorithm* such as [17]. This algorithm works in a number of operations linear with the cardinality of $\mathcal{M}(\nu) \cap \mathrm{cl}(\mathcal{B}_r)$, and most are feasible in parallel.

Oracle 1 may be adapted to ease its computation. Let $\alpha \in]0, 1]$ and define

$$\mathbb{O}_\alpha(x, \mathcal{S}) \triangleq \left\{ d_\alpha \in \mathrm{cl}(\mathcal{B}_r) : \frac{\mathrm{dist}\left(x + d_\alpha, \mathcal{S}\right)}{\max\limits_{d\in\mathrm{cl}(\mathcal{B}_r)} \mathrm{dist}\left(x + d, \mathcal{S}\right)} \geq \alpha \right\}, \tag{2}$$

for all $(x, \mathcal{S}) \in \mathbb{R}^n \times 2^{\mathbb{R}^n}$ with $\mathcal{S} \neq \emptyset$. Oracle 2 satisfies Definition 1 with $\beta \triangleq \alpha$, then it is a `covering` oracle. The set $\mathbb{O}_\alpha(x, \mathcal{S})$ usually contains infinitely many elements, but recall that in practice we do not need to compute more than one. A simple heuristic approach to localize such an element d_α may use a grid search on a grid thin enough, or some *space-filling sequences* such as the Halton sequence [13]. It is presumably easier to compute an element of the set defined by Oracle 2 than one of the set defined by Oracle 1, especially when α is chosen close to 0 and the points in $\mathcal{S}$ are well spread into $\mathrm{cl}(\mathcal{B}_r(x))$.

Oracles 1 and 2 may be altered for some marginal gain, by setting $\delta r \in \mathbb{R}^*_+$ small and considering

$$\mathbb{O}^{\delta r}(x, \mathcal{S}) \triangleq \mathbb{O}\left(x, \mathcal{S} \cap \mathrm{cl}(\mathcal{B}_{r+\delta r}(x))\right) \qquad \text{and} \qquad \mathbb{O}^{\delta r}_\alpha(x, \mathcal{S}) \triangleq \mathbb{O}_\alpha\left(x, \mathcal{S} \cap \mathrm{cl}(\mathcal{B}_{r+\delta r}(x))\right) \tag{3}$$

respectively, for all $x \in \mathbb{R}^n$ and all $\mathcal{S} \subseteq \mathbb{R}^n$ intersecting $\mathrm{cl}(\mathcal{B}_{r+\delta r}(x))$. These alterations reduce the number of elements to consider in the computation of the point-set distance, and they remain `covering` oracles (Definition 1 holds for $\mathbb{O}^{\delta r}$ with $\beta \triangleq \frac{\delta r}{2r+\delta r}$, and for $\mathbb{O}^{\delta r}_\alpha$ with $\beta \triangleq \alpha\frac{\delta r}{2r+\delta r}$).

As desired, the four schemes above ensure a priori that any `cDSM` relying on them satisfies Property 1 a posteriori, and that the additional cost per iteration induced by the `covering` step is small. Indeed, our schemes construct a `covering` step instance evaluating 1 point per iteration, while for comparison the `poll` step considers at least $n+1$ points per iteration since it relies on a positive basis of $\mathbb{R}^n$. Moreover, in a blackbox context, the cost to compute Oracle 1 or any of its alterations may be negligible anyway, since the bottleneck in this context is the cost to evaluate $f(x)$ for any $x \in X$ while the computation of Oracle 1 involves no call to the objective function.

Let us stress that, in practice, we may perform a `revealing` step such as in [2, 3] instead of a `covering` step relying on a `covering` oracle. We refer to Section 4.1 for details about the `revealing` step. Indeed, the `revealing` step ensures Property 1 when the algorithm generates exactly one refined point [3, Lemma 2], which is the usual behavior observed in practice. Moreover, the computational cost required by the `revealing` step is almost null. Nevertheless, despite its more expensive cost, Oracle 1 ensures that the trial points are well spread in a neighborhood of the current incumbent solution at each iteration. This contrasts with the `revealing` step, which offers no such guarantee.

Let us discuss also the `covering` radius r. All $r \in \mathbb{R}^*_+$ are accepted, but fine-tuning this value is a problem-dependent concern. Two default choices when no information about Problem (**P**) is available are $r \triangleq \frac{\delta^0}{10}$ or $r \triangleq \delta^0$. An extreme case occurs when r is sufficiently large so that the sublevel set of f of level $f(x^0)$ is included in $\mathrm{cl}(\mathcal{B}_r(x^0))$ (Assumption 1.a) ensures that this is always possible). In this case, Theorem 1 ensures that the `cDSM` returns a global solution to Problem (**P**). However, an overly large value of r is impracticable. The larger r is, the more likely the `cDSM` eventually escapes poor local solutions, while in contrast, the smaller r is, the faster the `covering` step covers $\mathcal{B}_r(x^*)$ well. In order to ensure that all points in $\mathcal{B}_r(x^*)$ are at distance at most $0 < \varepsilon \leq r$ of a `covering` trial point, the `covering` step must generate at least $(\frac{r}{\varepsilon})^n$ trial points in $\mathcal{B}_r(x^*)$ (and space them evenly). For a fixed value of ε, this number rapidly grows as r increases. Then, the value of r impacts the reliability of the `cDSM` as follows: the `cDSM` asymptotically covers any ball of radius r it visits infinitely often with smaller balls of any radius $0 < \varepsilon \leq r$, but to do so, each of these balls must be visited during at least $(\frac{r}{\varepsilon})^n$ iterations. One may also consider a sequence $(r^k)_{k\in\mathbb{N}}$ instead of a fixed r, provided that $\underline{r} \triangleq \inf_{k\in\mathbb{N}} r^k > 0$. In that case, one must replace $\mathcal{B}_r(x^*)$ by $\mathcal{B}_{\underline{r}}(x^*)$ in Property 1 and Theorem 1.

5 Formal `covering` step suited to many classes of `DFO` methods

Section 2 states Theorem 1 for the `DSM` only. Nevertheless, it is possible to preserve a similar theorem by considering many other `DFO` algorithms differing from `DSM`, provided that they are enhanced with a `covering` step. More precisely, a convergence result similar to Theorem 1 holds when the `cDSM` is replaced by any algorithm fitting in the broad algorithmic framework given as Algorithm 2.

Algorithm 2 Generic `DFO` algorithm with a `covering` step to solve Problem (P).

Initialization:
 set a covering radius $r \in \mathbb{R}^*_+$ and a `covering` oracle $\mathbb{O}$ (that is, $\mathbb{O}$ satisfying Definition 1);
 set the incumbent solution $x^0 \in X$ and the trial points history $\mathcal{H}^0 \triangleq \emptyset$.

For $k \in \mathbb{N}$ **do**:
 `covering` **step**:
  set $\mathcal{D}^k_{\mathtt{C}} \triangleq \mathbb{O}(x^k, \mathcal{H}^k)$; set $\mathcal{T}^k_{\mathtt{C}} \triangleq \{x^k\} + \mathcal{D}^k_{\mathtt{C}}$ and $t^k_{\mathtt{C}} \in \operatorname{argmin} f(\mathcal{T}^k_{\mathtt{C}})$;
  if $f(t^k_{\mathtt{C}}) < f(x^k)$, then set $t^k \triangleq t^k_{\mathtt{C}}$ and $\mathcal{T}^k_{\mathtt{O}} \triangleq \emptyset$ and skip to the `update` step;
 `optional` **step**:
  set $\mathcal{D}^k_{\mathtt{O}} \subseteq \mathbb{R}^n$ empty or finite; if $\mathcal{T}^k_{\mathtt{O}} \triangleq \{x^k\} + \mathcal{D}^k_{\mathtt{O}}$ is nonempty, then set $t^k_{\mathtt{O}} \in \operatorname{argmin} f(\mathcal{T}^k_{\mathtt{O}})$;
  if $f(t^k_{\mathtt{O}}) < f(x^k)$, then set $t^k \triangleq t^k_{\mathtt{O}}$, otherwise set $t^k \triangleq x^k$;
 `update` **step**:
  set $x^{k+1} \triangleq t^k$; set $\mathcal{H}^{k+1} \triangleq \mathcal{H}^k \cup \mathcal{T}^k_{\mathtt{C}} \cup \mathcal{T}^k_{\mathtt{O}}$.

Algorithm 2 generates a sequence $(x^k, \mathcal{H}^k)_{k\in\mathbb{N}}$ from which we may define

$$\mathcal{A} \triangleq \left\{ \lim_{k\in K^*} x^k : K^* \subseteq \mathbb{N} \text{ indexes a converging subsequence of } (x^k)_{k\in\mathbb{N}} \right\} \qquad \text{and} \qquad \mathcal{H} \triangleq \bigcup_{k\in\mathbb{N}} \mathcal{H}^k,$$

and an equivalent of Property 1 holds, since the `covering` step follows a `covering` oracle. More precisely, it holds that $\mathcal{B}_r(x^*) \subset \mathrm{cl}(\mathcal{H})$ for all $x^* \in \mathcal{A}$.

Unlike the `cDSM`, Algorithm 2 does not project the oracle trial points onto a mesh and does not discriminate the trial points via a sufficient decrease function. Moreover, it involves no sequence of poll radii and no mandatory `poll` step, so it cannot be studied via the notion of refining subsequence. Nevertheless, its convergence analysis, formalized in Theorem 2 is similar to those of the `cDSM`. Also, Theorem 2 is valid for the `cDSM`, provided that its `covering` step is defined exactly as in Algorithm 2.

Theorem 2. Under Assumption 1, all $x^* \in \mathcal{A} \neq \emptyset$ generated by Algorithm 2 and all $K^* \subseteq \mathbb{N}$ such that $(x^k)_{k \in K^*}$ converges to x^* satisfy

$$\lim_{k \in K^*} f(x^k) \;=\; \begin{cases} \min f(\mathcal{B}_r(x^*)) = f(x^*) & \text{if } x^* \in X, \\ \inf f(\mathcal{B}_r(x^*)) & \text{if } x^* \notin X. \end{cases}$$

Proof. Consider Algorithm 2 under Assumption 1. First, Algorithm 2 generates $\mathcal{A} \neq \emptyset$ since all elements of $(x^k)_{k \in \mathbb{N}}$ lie in the closure of the sublevel set of f of level $f(x^0)$, which is compact because of Assumption 1.a). Second, we match the notation of Section 3 by denoting by $\mathcal{M}(\nu) \triangleq \mathbb{R}^n$ and $\rho(\nu) \triangleq 0$ for all $\nu \in \mathbb{R}_+$, and we set $\underline{\delta}^k \triangleq 0$ for all $k \in \mathbb{N}$. This proof consists in reading former propositions (and their proofs) to observe that they remain valid by replacing a refining subsequence by a converging one in this setting. The proof of Proposition 3 requires a set $K^* \subseteq \mathbb{N}$ indexing a refining sequence only to ensure that $(\underline{\delta}^k)_{k \in K^*}$ converges to 0, so it remains valid in our current setting. Then, Proposition 3 claims that the trial points history $\mathcal{H}$ satisfies $\mathcal{B}_r(x^*) \subseteq \mathrm{cl}(\mathcal{H})$ for all $x^* \in \mathcal{A}$. Similarly, the proofs of Proposition 2 and of Theorem 1 remain valid for our current setting, since again a set $K^* \subseteq \mathbb{N}$ indexing a refining sequence is required in these proofs only to ensure that $(\underline{\delta}^k)_{k \in K^*}$ converges to 0. Hence, Theorem 2 is proved by reading the proof of Theorem 1 in the current setting. □

In practice, Algorithm 2 allows skipping the `covering` step and allows a perturbation of the `covering` oracle (for example, by a projection onto a mesh). However, the `covering` step must be performed infinitely often, in the sense that for all $x^* \in \mathcal{A}$, there must exists $K^* \subseteq \mathbb{N}$ such that $(x^k)_{k \in K^*}$ converges to x^* and the `covering` step is performed at all iterations $k \in K^*$. Moreover, the `covering` oracle must not be excessively perturbed by, for example, rounding infinitely often onto a coarse mesh. Precisely, if Algorithm 2 relies on a positive sequence $(\nu^k)_{k \in \mathbb{N}}$ and sets $\mathcal{D}^k_{\mathtt{C}} \triangleq \mathrm{round}(\mathbb{O}(x^k, \mathcal{H}^k), \mathcal{M}(\nu^k))$ or accepts $x^{k+1} \triangleq t^k_{\mathtt{C}}$ only if $f(t^k_{\mathtt{C}}) < f(x^k) - \rho(\nu^k)$ at each iteration $k \in \mathbb{N}$, then it is mandatory that $(\nu^k)_{k \in \mathbb{N}}$ decreases to 0. This echoes our comment in Remark 1 that the `cDSM` relies on the sequence of smallest poll radii $(\underline{\delta}^k)_{k \in \mathbb{N}}$ in addition to those of the current poll radii $(\delta^k)_{k \in \mathbb{N}}$. Indeed, the mesh-based `cDSM` rounds the `covering` oracle into the mesh parameterized by $(\underline{\delta}^k)_{k \in \mathbb{N}}$ and the sufficient decrease-based `cDSM` discriminates trial points by a sufficient decrease depending on $(\underline{\delta}^k)_{k \in \mathbb{N}}$, and (unlike the sequence $(\delta^k)_{k \in \mathbb{N}}$) the sequence $(\underline{\delta}^k)_{k \in \mathbb{N}}$ is guaranteed to converge to 0 under Assumption 1.

We conclude this section by stressing that Theorem 2 is applicable in wide range of situations. Indeed, Section 6.2 below shows that Assumption 1 is quite weak, and Section 4.3 provides an explicit scheme for the `covering` oracle compatible with many `DFO` algorithms. Moreover, Theorem 2 shows that the `covering` step is, roughly speaking, a self-sufficient algorithmic step that ensures the local optimality of all accumulation points returned by any algorithm it is fitted into.

6 Discussion on Assumption 1.c)

This section discusses our novel assumption describing the continuity sets of f, that is, Assumption 1.c). In Section 6.1, we prove that Assumption 1.c) is strictly weaker than similar assumptions considered in former work [2, 3, 20]. In Section 6.2, we show that Assumption 1.c) is tight.

6.1 Comparison of Assumption 1.c) with similar assumptions in prior work

In this section, we compare Assumption 1.c) to similar assumptions considered by former work [2, 3, 20]. Precisely, we show that Assumption 1.c) is strictly weaker than either [2, Assumption 4.4] and [3, Assumption 1]. The work [20] is not considered since [3] is an extension of it.

First, let us compare Assumption 1.c) to [2, Assumption 4.4], recalled below as Assumption 2. We prove in Proposition 4 that Assumption 1.c) is strictly weaker than Assumption 2.

Assumption 2 (Assumption 4.4 in [2])**.** There exists $N \in \mathbb{N}^* \cup \{+\infty\}$ nonintersecting open sets X_i such that $\mathrm{cl}(X) = \cup_{i=1}^{N} \mathrm{cl}(X_i)$ and $f_{|X_i}$ is continuous for all $i \in [\![1, N]\!]$ and, for all $x \in X$, there exists $j \in [\![1, N]\!]$ such that $x \in \mathrm{cl}(X_j)$ and $f_{|X_j \cup \{x\}}$ is continuous.

Proposition 4. Assumption 1.c) is strictly weaker than Assumption 2.

Proof. Suppose that Assumption 2 holds. Denote by $(X_i)_{i=1}^{N}$ the family it provides. For all $i \in [\![1, N]\!]$, let $\mathrm{cl}_f(X_i) \triangleq \{x \in \mathrm{cl}(X_i) : f_{|X_i \cup \{x\}} \text{ is continuous}\}$. Let $I(x) \triangleq \min\{i \in [\![1, N]\!] : x \in \mathrm{cl}_f(X_i)\}$ for all $x \in X$. Then let $Y_i \triangleq \{x \in X : I(x) = i\}$ for every $i \in [\![1, N]\!]$. Thus $(Y_i)_{i=1}^{N}$ satisfies all the requirements in Assumption 1.c) (see Proposition 9), so Assumption 1.c) is weaker than Assumption 2. Now, to prove that Assumption 2 is not weaker than Assumption 1.c), consider the case

$$f : \begin{cases} X \triangleq [-1, 1] \setminus \{0\} & \to & \mathbb{R} \\ x & \mapsto & \dfrac{1}{i} \text{ if } |x| \in \left]\dfrac{1}{i+1}, \dfrac{1}{i}\right] \text{ for some } i \in \mathbb{N}^*. \end{cases}$$

The continuity sets of f are $X_i \triangleq [\frac{-1}{i}, \frac{-1}{i+1}[\cup]\frac{1}{i+1}, \frac{1}{i}]$ for all $i \in \mathbb{N}^*$. Assumption 1.c) holds since X_i is ample for all $i \in \mathbb{N}^*$. Nevertheless Assumption 2 does not hold since the continuity sets must be adapted as $Y_i \triangleq \mathrm{int}(X_i)$ for all $i \in \mathbb{N}^*$ to be open, but then $\mathrm{cl}(X) = [-1, 1] \neq \cup_{i=1}^{\infty} \mathrm{cl}(Y_i) = [-1, 1] \setminus \{0\}$. □

Second, let us compare Assumption 1.c) to [3, Assumption 1], reformulated below in Assumption 3. We prove in Proposition 5 that Assumption 1.c) is strictly weaker than Assumption 3.

Assumption 3 (Global reformulation of Assumption 1 in [3])**.** The set X admits a partition $X = \sqcup_{i=1}^{N} X_i$ (where $N \in \mathbb{N}^* \cup \{+\infty\}$) such that, for all $i \in [\![1, N]\!]$, X_i is a *continuity set of f with the interior cone property* (that is, X_i satisfies Definition 2 below and $f_{|X_i} : X_i \to \mathbb{R}$ is continuous).

Definition 2 (Interior cone property and exterior cone property)**.** A set $\mathcal{S} \subseteq \mathbb{R}^n$ is said to have the *interior cone property* (ICP) if

$$\forall x \in \partial\mathcal{S}, \quad \exists \begin{cases} \mathcal{U} \subseteq \mathbb{S}^n \text{ nonempty, open in the topology induced by } \mathbb{S}^n \\ \mathcal{K} \triangleq \mathbb{R}_+^* \mathcal{U} \text{ the cone generated by } \mathcal{U}, \quad \mathcal{K}_x \triangleq \{x\} + \mathcal{K} \\ \mathcal{O} \text{ an open neighborhood of 0 in } \mathbb{R}^n, \quad \mathcal{O}_x \triangleq \{x\} + \mathcal{O} \end{cases} : \quad (\mathcal{K}_x \cap \mathcal{O}_x) \subseteq \mathcal{S}.$$

Similarly, $\mathcal{S}$ is said to have the *exterior cone property* (ECP) if $(\mathbb{R}^n \setminus \mathcal{S})$ has the ICP.

Assumption 3 differs from [3, Assumption 1] in three aspects. Let us stress those and argue that they do not spoil the assumption. First, Assumption 3 is a global statement, while [3, Assumption 1] states a local property but for each $x \in X$. Second, [3, Assumption 1] requires Lipschitz-continuity of f on each continuity set, while Assumption 3 calls for continuity only. Indeed [3] requires Lipschitz-continuity only to evaluate some generalized derivatives. We do not consider those in the current work, so we weaken the assumption accordingly. Third, [3, Assumption 1] requires an *exterior* cone property for $X \setminus X_i$ for all $i \in [\![1, N]\!]$, while Assumption 3 demands an *interior* cone property for X_i. Moreover the exterior cone property required in [3, Assumption 1] is [20, Definition 4.1], which differs from Definition 2. However, these two approaches are equivalent (see Proposition 10). Hence, Assumption 3 and [3, Assumption 1] differ only in the nature of the continuity of f on each continuity set.

Proposition 5. Assumption 1.c) is strictly weaker than Assumption 3.

Proof. For all $\mathcal{S} \subseteq \mathbb{R}^n$, if $\mathcal{S}$ has the ICP, then $\mathcal{S}$ is ample, but the reciprocal implication may fail (see Proposition 11). The result follows directly. □

6.2 Tightness of Assumption 1.c)

In this section, we show that Assumption 1.c) is tight, in the sense that the conclusion of Theorem 2 may not hold if Assumption 1.c) is not satisfied.

Proposition 6. If Assumption 1.c) does not hold, then the conclusion of Theorem 2 may not hold.

Proof. Let us develop a counterexample. Let $n = 2$ and $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$\forall x \in X \triangleq \mathbb{R}^2, \qquad f(x) \triangleq \begin{cases} |x_1 - 1| + |x_2| - 1, & \text{if } x \in X_1 \triangleq (\mathbb{R}_- \times \mathbb{R}) \cup (\mathbb{R}_+^* \times \{0\}), \\ |x_1| + |x_2|, & \text{if } x \in X_2 \triangleq X \setminus X_1. \end{cases}$$

Assumptions 1.a) and 1.b) hold, but Assumption 1.c) does not since X_1 is locally thin and thus not ample. Consider an instance of Algorithm 2 such that $(\mathcal{D}_{\mathtt{C}}^k \cup \mathcal{D}_{\mathtt{O}}^k) \cap (\mathbb{R}_+^* \times \{0\}) = \emptyset$ for all $k \in \mathbb{N}$ and starting from the origin. This instance remains at the origin, since it evaluates only points in $\mathrm{cl}(\mathrm{int}(X_1)) \cup X_2 = \mathbb{R}^2 \setminus (\mathbb{R}_+^* \times \{0\})$ and the origin is the global minimizer of the restriction of f to $\mathrm{cl}(\mathrm{int}(X_1)) \cup X_2$. In that situation, the origin is a refined point lying in X but is not a local minimizer of f, which contradicts the conclusion of Theorem 2. □

Assumption 1.c) allows a broad class of discontinuous functions, as it only rejects discontinuous functions for which at least one of the continuity sets is locally thin. A relaxation as a local assumption holding only at the refined points is possible, but such relaxation can be checked only a posteriori since it is impossible to determine the refined points a priori. Hence our framework cannot be broadened using only information available a priori. Nevertheless, we leave the associated adaptation of Theorem 2 as Assertion 1 below. Its proof is similar to those provided in Section 5.

Assertion 1. Under Assumptions 1.a) and 1.b), all $x^* \in \mathcal{A} \neq \emptyset$ generated by Algorithm 2 and all sets $K^* \subseteq \mathbb{N}$ such that $(x^k)_{k \in K^*}$ converges to x^* satisfy

a) if $\mathcal{X}_i$ is ample for all $i \in [\![1, N]\!]$, then $\lim_{k \in K^*} f(x^k) = \inf f(\mathcal{B}_r(x^*))$;

b) if moreover $x^* \in \mathcal{X}_i$ for some $i \in [\![1, N]\!]$, then $\lim_{k \in K^*} f(x^k) = \min f(\mathcal{B}_r(x^*)) = f(x^*)$;

where we denote by $(\mathcal{X}_i)_{i=0}^N$ a partition of $\mathcal{B}_r(x^*)$, for some $N \in \mathbb{N}^* \cup \{+\infty\}$, such that $\mathcal{X}_0 \triangleq \mathcal{B}_r(x^*) \setminus X$ and $\mathcal{X}_i$ is a continuity set of f for all $i \in [\![1, N]\!]$.

7 General comments and main extensions for future work

Theorem 1 is stronger than most results about the `DSM` in three aspects. First, Theorem 1 ensures the local optimality of all refined points while the literature usually checks necessary optimality conditions. Second, Theorem 1 involves Assumption 1 on the problem only, while the `DSM` [20, Theorems 3.1 and 3.2] and the `rDSM` [3, Theorem 1] require also an assumption on the execution of the algorithm (respectively, that $\lim_{k \in K^*} f(x^k) = f(x^*)$ for all $K^* \subseteq \mathbb{N}$ indexing a refining subsequence with refined point x^*, and that a unique refined point is generated). Third, Theorem 1 holds for all refining subsequences. In contrast, the literature usually considers only refining subsequences such that the set of associated refined directions is dense in $\mathbb{S}^n$, and no way to identify such a refining subsequence is provided.

The assumption that f is lower semicontinuous may be relaxed. The proof of Theorem 1 highlights that, if only Assumption 1.b) fails, then $\lim_{k \in K^*} f(x^k) = \inf f(\mathcal{B}_r(x^*))$ holds true for all $x^* \in \mathcal{R}$ and all $K^* \subseteq \mathbb{N}$ indexing an associated refining subsequence. Assuming the lower semicontinuity of f at x^* recovers $\inf f(\mathcal{B}_r(x^*)) = \min f(\mathcal{B}_r(x^*)) = f(x^*)$. A similar claim holds true for Theorem 2.

Only the `covering` step matters to establish the convergence towards local solutions, in the sense formalized in Section 5. The use of a few directions per iteration shares similarities with `DFO` line search algorithms [9] and reduced space algorithms [18]. Many methods fit in Algorithm 2 when enhanced with the `covering` step. Future work may add a `covering` step into Bayesian-based methods, model-based methods, and most `DFO` methods listed in [14]. By coupling the `covering` step with others ideas from the literature, several technical requirements may be alleviated. For example, the requirement in the `cDSM` that $x^0 \in X$ is relaxable when Problem (**P**) is the *extreme barrier* reformulation of a constrained problem with quantifiable constraints [15], using for example a two-phase algorithm [7, Algorithm 12.1] or the *progressive barrier* [6] which iteratively reduces the infeasibility of the incumbent solution.

Although they are optional in Algorithm 2, the `search` and `poll` steps are important in practice. The first allows for global exploration, and the second usually contributes to many successful iterations. In contrast, the `covering` step aims to ensure the asymptotic Property 1, so a poor instance may be inefficient in finite time. We conducted a numerical experiment on problems where the objective functions are alterations of $f \triangleq \|\cdot\|$. It appears that the addition of the `covering` step makes negligible difference in performance when the `cDSM` converges towards a local solution and the objective function is not too erratic near that solution, but that it provides a gain of quality on the returned solution when the objective function has very thin continuity sets. These experiments are reported to Appendix A.2. Future work may investigate these observations and quantify the relevance of each step. Moreover, we only tested our baseline schemes for the `covering` step schemes from Section 4.3. Further investigations and careful implementations may provide more efficient schemes. We may also study how well the `cDSM` performs when Assumption 1.c) holds but its stronger variant Assumption 3 does not. Presumably, the interior cone property from Assumption 3 is important for practical efficiency.

Our schemes for the `covering` oracle in Section 4.3 ensure only that the trial points are well spread around the current incumbent solution at each iteration. Another scheme, for example from an expected improvement [22] or model-based techniques [8], may also seek for points that are relevant candidates or that help to gather the shape of f. Another future work could alter Assumption 1.c) so that a partition of X into continuity sets of f is explicitly provided. A `covering` oracle designed accordingly could therefore evaluate points in all continuity sets of f, even if some are not ample.

The `covering` step is also compatible with the `DiscoMads` algorithm [2], which designs the original `revealing` step for the purpose to detect discontinuities and repel its incumbent solution from those. This discontinuities detection is more accurate with the `covering` step than with the `revealing` step, as the latter may fail to detect discontinuities when more than one refined point is eventually generated. Also, when the `covering` step relies on Oracle 1, it presumably has better practical guarantees to efficiently detect discontinuities than both the the original `revealing` step from [2] and the adapted `revealing` step from [3]. Then, for reliability reasons, it may be safer to use the `DiscoMads` algorithm with a `covering` step instead of a `revealing` step.

We conclude this paper with the next Table 1. It summarizes the conceptual differences between the usual `DSM` and our `cDSM`, and it highlights the aforementioned ideas to extend our work.

Table 1: Differences between DSM and cDSM, and possible new methods with new covering step goals.

step	method		
	DSM	cDSM	DFO method with involved covering step design
search	Optional. Allows the use of heuristics and globalization strategies.		
poll	Required. Converges towards refined points satisfying necessary optimality conditions.	Optional. But in practice, it performs well in converging towards a good refined point.	
covering	Undefined.	Required. Asymptotically ensures that all refined points are local solutions; low cost per iteration but lacks efficiency in practice.	Required. Asymptotically ensures that all refined points are local solutions; exploits information about the objective function.

A Supplementary materials

A.1 Variant of Algorithm 1 with nonstringent definition of its parameters

Algorithm 3 Generic cDSM (covering DSM) solving Problem (P).

Initialization:

set a covering radius $r \in \mathbb{R}^*_+$ and the trial points history as $\mathcal{H}^0 \triangleq \emptyset$;

set the incumbent solution and poll radius as $(x^0, \delta^0) \in X \times \mathbb{R}^*_+$, and set $\underline{\delta}^0 \triangleq \delta^0$;

set $\mathcal{M} : \mathbb{R}_+ \to 2^{\mathbb{R}^n}$ and $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ and $\Lambda \subset]0,1[$ and $\Upsilon \subset [1,+\infty[$ according to one of

- the *mesh-based DSM*: $\mathcal{M}(\nu) \triangleq \min\{\nu, \frac{\nu^2}{\delta^0}\}M\mathbb{Z}^p$ for all $\nu \in \mathbb{R}_+$, where $p > n$ and $M \in \mathbb{R}^{n\times p}$ positively spans $\mathbb{R}^n$, and $\rho(\cdot) \triangleq 0$, and $\Lambda \subseteq \{\tau^\ell : \ell \in [\![1,m]\!]\}$ and $\Upsilon \subseteq \{\tau^\ell : \ell \in [\![-m,0]\!]\}$ where $\tau \in \mathbb{Q} \cap]0,1[$ and $m \in \mathbb{N}^*$;
- the *sufficient decrease-based DSM*: $\mathcal{M}(\cdot) \triangleq \mathbb{R}^n$, and ρ increasing with $0 < \rho(\nu) \in o(\nu)$ as $\nu \searrow 0$, and $\Lambda \subseteq [\underline{\lambda}, \overline{\lambda}]$ and $\Upsilon \subseteq [\underline{\upsilon}, \overline{\upsilon}]$ where $0 < \underline{\lambda} \leq \overline{\lambda} < 1 \leq \underline{\upsilon} \leq \overline{\upsilon} < +\infty$.

For $k \in \mathbb{N}$ **do**:

covering **step**:

set $\mathcal{D}^k_{\mathtt{C}} \subseteq \mathcal{M}(\underline{\delta}^k) \cap \mathrm{cl}(\mathcal{B}_r)$ nonempty and finite; set $\mathcal{T}^k_{\mathtt{C}} \triangleq \{x^k\} + \mathcal{D}^k_{\mathtt{C}}$; set $t^k_{\mathtt{C}} \in \operatorname{argmin} f(\mathcal{T}^k_{\mathtt{C}})$;

if $f(t^k_{\mathtt{C}}) < f(x^k) - \rho(\underline{\delta}^k)$, then set $t^k \triangleq t^k_{\mathtt{C}}$ and $\mathcal{T}^k_{\mathtt{S}} = \mathcal{T}^k_{\mathtt{P}} \triangleq \emptyset$ and skip to the update step;

search **step**:

set $\mathcal{D}^k_{\mathtt{S}} \subseteq \mathcal{M}(\delta^k)$ empty or finite; if $\mathcal{T}^k_{\mathtt{S}} \triangleq \{x^k\} + \mathcal{D}^k_{\mathtt{S}}$ is nonempty, then set $t^k_{\mathtt{S}} \in \operatorname{argmin} f(\mathcal{T}^k_{\mathtt{S}})$;

if also $f(t^k_{\mathtt{S}}) < f(x^k) - \rho(\underline{\delta}^k)$, then set $t^k \triangleq t^k_{\mathtt{S}}$ and $\mathcal{T}^k_{\mathtt{P}} \triangleq \emptyset$ and skip to the update step;

poll **step**:

set $\mathcal{D}^k_{\mathtt{P}} \subseteq \mathcal{M}(\underline{\delta}^k) \cap \mathrm{cl}(\mathcal{B}_{\delta^k})$ a positive basis of $\mathbb{R}^n$; set $\mathcal{T}^k_{\mathtt{P}} \triangleq \{x^k\} + \mathcal{D}^k_{\mathtt{P}}$; set $t^k_{\mathtt{P}} \in \operatorname{argmin} f(\mathcal{T}^k_{\mathtt{P}})$;

if $f(t^k_{\mathtt{P}}) < f(x^k) - \rho(\underline{\delta}^k)$, then set $t^k \triangleq t^k_{\mathtt{P}}$, otherwise set $t^k \triangleq x^k$;

update **step**:

set $\mathcal{H}^{k+1} \triangleq \mathcal{H}^k \cup \mathcal{T}^k$, where $\mathcal{T}^k \triangleq \mathcal{T}^k_{\mathtt{C}} \cup \mathcal{T}^k_{\mathtt{S}} \cup \mathcal{T}^k_{\mathtt{P}}$; set $x^{k+1} \triangleq t^k$;

set δ^{k+1} as $\delta^{k+1} \in \delta^k\Upsilon$ if $t^k \neq x^k$ and $\delta^{k+1} \in \delta^k\Lambda$ otherwise; set $\underline{\delta}^{k+1} \triangleq \min_{\ell \leq k} \delta^\ell$.

A.2 Report of some numerical experiments regarding the cDSM

This appendix compares the numerical performance of the cDSM with the usual DSM. We consider two problems that illustrate typical situations a DFO algorithm may face in a discontinuous context,

$$\underset{x \in \mathbb{R}^2}{\text{minimize}} \quad \begin{cases} \|x\|_\infty + 1 & \text{if } x_1 > 0, \\ \|x\|_\infty & \text{otherwise,} \end{cases} \tag{$\mathbf{P}^{\mathbf{test}}_1$}$$

and

$$\underset{x \in \mathbb{R}^2}{\text{minimize}} \quad \begin{cases} \|x\|_\infty + 1 & \text{if } x_1 > 0, \\ \|x\|_\infty & \text{if } x_1 \leq 0 \text{ and } \|p(x)\|_2 \leq \min\{\|q(x)\|_2^2, \frac{1}{100}\}, \\ +\infty & \text{otherwise,} \end{cases} \tag{$\mathbf{P}^{\mathbf{test}}_2$}$$

where, for all $x \in \mathbb{R}^2$, $p(x) \triangleq \frac{x^\top a}{\|a\|^2} a$ denotes the projection of x onto the line directed by $a \triangleq (-1, 1)$ and where $q(x) \triangleq x - p(x)$. The objective functions of these two problems are represented on Figure 3.

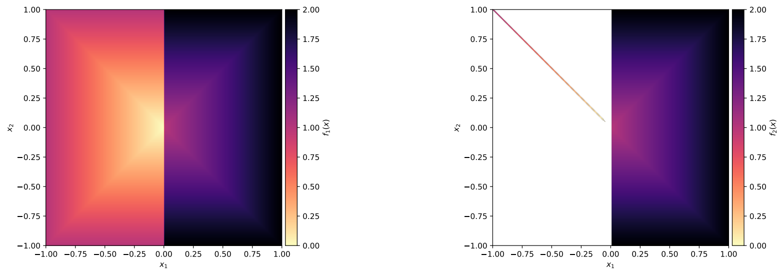

Figure 3: Objective functions of Problems ($\mathbf{P}_1^{\mathbf{test}}$) (on the left plot) and ($\mathbf{P}_2^{\mathbf{test}}$) (on the right plot).

Our experiments study an instance of Algorithm 1. In agreement with the framework of Theorem 2, we define $\mathcal{M}(\nu) \triangleq \mathbb{R}^n$ and $\rho(\nu) \triangleq 0$ for all $\nu \in \mathbb{R}_+$. The incumbent is $x^0 \triangleq (98.7654321, 12.3456789)$, the poll radius is $\delta^0 \triangleq 1$, the `covering` radius is $r \triangleq \frac{\delta^0}{10}$, and the shrinking factor is $\tau \triangleq \frac{1}{2}$. The `covering` step considers, for each iteration $k \in \mathbb{N}$, all directions d on a grid over $\mathrm{cl}(\mathcal{B}_r)$ thin enough and selects one maximizing $\mathrm{dist}(x^k + d, \mathcal{H}^k)$, as in Section 4.3. The `search` step explores a momentum-based strategy with $\mathcal{D}_\mathtt{S}^k \triangleq \{3(x^k - x^{k-1})\}$ for all iteration $k \in \mathbb{N}^*$. The `poll` step follows the orthogonal polling from [1]. The stopping criterion is that either $\delta^k < 10^{-8}$ or $k \geq 300$. We compare 10 executions of this `cDSM` with 10 executions of the associated `DSM` skipping the `covering` step.

For each problem and each execution of each algorithm, we record the smallest objective function value found after the current number of objective function evaluations, and the evolution of the ratio of respectively the number of `search` successes, `covering` successes, `poll` successes and iterations failures over the current number of iterations achieved.

Figure 4 shows the results for Problem ($\mathbf{P}_1^{\mathbf{test}}$). The discontinuity is easy to handle, so the `DSM` performs well while the `cDSM` requires 5% to 10% more function evaluations. For both algorithms, the contribution of the `search` step is important at the beginning of the optimization process but declines when the algorithm approaches the solution. Then, the `poll` step starts performing instead. The `covering` step behaves similarly to the `search` step.

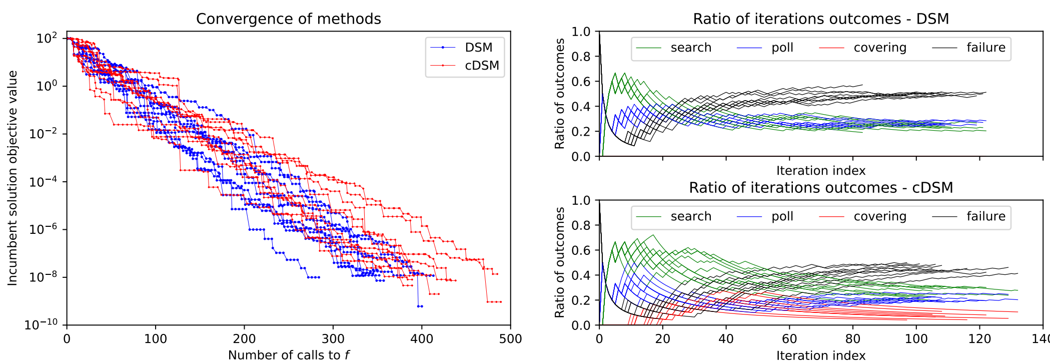

Figure 4: Results of our experiments on Problem ($\mathbf{P}_1^{\mathbf{test}}$).

The results on Problem ($\mathbf{P}_2^{\mathbf{test}}$) are shown in Figure 5. The `DSM` converges close to the solution, but it fails to reliably find the continuity set of f containing it. In 4 out of 10 cases, the `DSM` does not leave the continuity set it started from. In contrast, the `cDSM` systematically converges to the solution from the correct continuity set. Both algorithms trigger the stopping criterion $k \geq 300$, but other tests conducted by relaxing this criterion show a similar outcome. We suspect that this problem is difficult because the continuity set containing the solution does not have the ICP (from Definition 2).

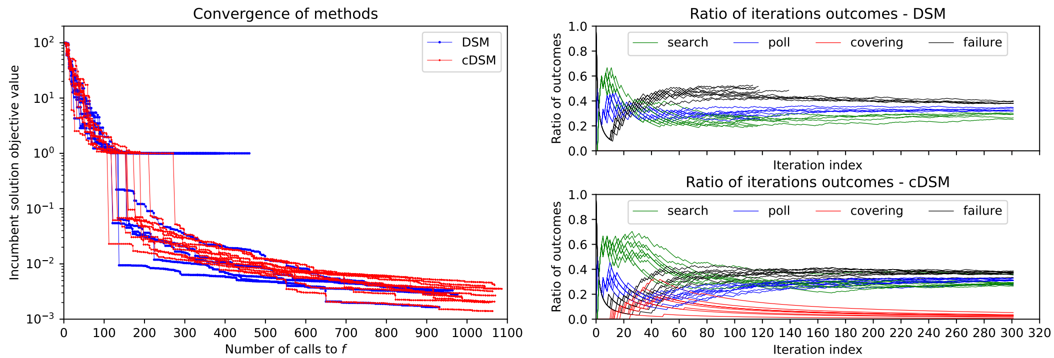

Figure 5: Results of our experiments on Problem $(\mathbf{P}_2^{\mathbf{test}})$.

B Proofs of auxiliary results

Proposition 7. For all $\mathcal{S} \subseteq \mathbb{R}^n$, $\mathcal{S}$ is locally thin if and only if it is not ample.

Proof. Let $\mathcal{S} \subseteq \mathbb{R}^n$ be locally thin and let us prove by contradiction that it is not ample. Assume that $\mathcal{S}$ is ample. Define $\mathcal{N} \subseteq \mathbb{R}^n$ open such that $\mathcal{S} \cap \mathcal{N} \neq \emptyset = \mathrm{int}(\mathcal{S}) \cap \mathcal{N}$, and take $x \in \mathcal{S} \cap \mathcal{N}$. Then $x \in \mathrm{cl}(\mathrm{int}(\mathcal{S})) \cap \mathcal{N}$ and thus there exists $\varepsilon > 0$ such that $\mathcal{B}_\varepsilon(x) \subseteq \mathcal{N}$ and $\mathcal{B}_\varepsilon(x) \cap \mathrm{int}(\mathcal{S}) \neq \emptyset$. Hence $\mathrm{int}(\mathcal{S}) \cap \mathcal{N} \neq \emptyset$ which raises a contradiction. Reciprocally, let $\mathcal{S} \subseteq \mathbb{R}^n$ be not ample. Then, $\mathcal{N} \triangleq \mathbb{R}^n \setminus \mathrm{cl}(\mathrm{int}(\mathcal{S}))$ is open, and $\mathcal{S} \cap \mathcal{N} \neq \emptyset = \mathrm{int}(\mathcal{S}) \cap \mathcal{N}$ by construction. Hence $\mathcal{S}$ is locally thin. $\square$

Proposition 8. We say that $\mathcal{S}_1 \subseteq \mathbb{R}^n$ *has a dense intersection with* $\mathcal{S}_2 \subseteq \mathbb{R}^n$ if $\mathcal{S}_2 \subseteq \mathrm{cl}(\mathcal{S}_1 \cap \mathcal{S}_2)$. The following properties hold for all $\mathcal{S}_1 \subseteq \mathbb{R}^n$, $\mathcal{S}_2 \subseteq \mathbb{R}^n$ and $\mathcal{S}_3 \subseteq \mathbb{R}^n$:

a) if $\mathcal{S}_1$ is ample and $\mathcal{S}_2$ is open, then $\mathcal{S}_1 \cap \mathcal{S}_2$ is ample;

b) if $\mathcal{S}_1$ has a dense intersection with $\mathcal{S}_2$ and $\mathcal{S}_3$ is an ample subset of $\mathcal{S}_2$, then $\mathcal{S}_1$ has a dense intersection with $\mathcal{S}_3$;

c) if $\mathcal{S}_1$ has a dense intersection with $\mathcal{S}_2$ and $\mathcal{S}_2$ is ample, then, for all $x \in \mathcal{S}_2$, $\mathcal{S}_1 \setminus \{x\}$ has a dense intersection with $\mathcal{S}_2$;

d) if $\mathcal{S}_1$ is open, $\mathcal{S}_2$ is ample and $\mathcal{S}_3$ has a dense intersection with $\mathcal{S}_1$, then for all $x \in \mathcal{S}_1$, the set $\mathcal{S}_3 \setminus \{x\}$ has a dense intersection with $\mathcal{S}_1 \cap \mathcal{S}_2$.

Proof. Let us prove the first statement. Let $\mathcal{S}_1 \subseteq \mathbb{R}^n$ be ample and $\mathcal{S}_2 \subseteq \mathbb{R}^n$ be open, and let $x \in \mathcal{S}_1 \cap \mathcal{S}_2$. We have $x \in \mathcal{S}_1 \subseteq \mathrm{cl}(\mathrm{int}(\mathcal{S}_1))$, so there exists $(x^k)_{k\in\mathbb{N}}$ converging to x with $x^k \in \mathrm{int}(\mathcal{S}_1)$ for all $k \in \mathbb{N}$. Since $x^k \to x \in \mathcal{S}_2$ with $\mathcal{S}_2$ open, it holds that $x^k \in \mathcal{S}_2$ for all k large enough. Finally, we have $x^k \in \mathrm{int}(\mathcal{S}_1) \cap \mathcal{S}_2 = \mathrm{int}(\mathcal{S}_1) \cap \mathrm{int}(\mathcal{S}_2) = \mathrm{int}(\mathcal{S}_1 \cap \mathcal{S}_2)$ for all k large enough. We deduce that $x \in \mathrm{cl}(\mathrm{int}(\mathcal{S}_1 \cap \mathcal{S}_2))$.

Next, let us prove the second statement. Let $\mathcal{S}_1 \subseteq \mathbb{R}^n$ having a dense intersection with $\mathcal{S}_2 \subseteq \mathbb{R}^n$ and let $\mathcal{S}_3 \subseteq \mathcal{S}_2$ be ample, and let $x \in \mathcal{S}_3$. We have $x \in \mathrm{cl}(\mathrm{int}(\mathcal{S}_3))$ so there exists $(x^k)_{k\in\mathbb{N}}$ converging to x with $x^k \in \mathrm{int}(\mathcal{S}_3) \subseteq \mathcal{S}_2 \subseteq \mathrm{cl}(\mathcal{S}_1 \cap \mathcal{S}_2)$ for all $k \in \mathbb{N}$. Then, for all $k \in \mathbb{N}$, there exists $(x^k_\ell)_{\ell\in\mathbb{N}}$ converging to x^k with $x^k_\ell \in \mathcal{S}_1 \cap \mathcal{S}_2 \cap \mathrm{int}(\mathcal{S}_3)$ for all $\ell \in \mathbb{N}$. For all $k \in \mathbb{N}$, let $\ell(k) \in \mathbb{N}$ be such that $\|x^k_{\ell(k)} - x^k\| \leq 2^{-k}$. It follows that $(x^k_{\ell(k)})_{k\in\mathbb{N}}$ converges to x and $x^k_{\ell(k)} \in \mathcal{S}_1 \cap \mathcal{S}_2 \cap \mathrm{int}(\mathcal{S}_3) \subseteq \mathcal{S}_1 \cap \mathcal{S}_3$ for all $k \in \mathbb{N}$. Hence, we get that $x \in \mathrm{cl}(\mathcal{S}_1 \cap \mathcal{S}_3)$.

Now, let us prove the third statement. Let $\mathcal{S}_1 \subseteq \mathbb{R}^n$ and let $\mathcal{S}_2 \subseteq \mathbb{R}^n$ be ample. Assume that $\mathcal{S}_1$ has a dense intersection with $\mathcal{S}_2$ and let $x \in \mathcal{S}_2$. Let $y \in \mathcal{S}_2$. Then $y \in \mathrm{cl}(\mathrm{int}(\mathcal{S}_2))$, so there exists $(y^k)_{k\in\mathbb{N}}$ converging to y with $y \neq y^k \in \mathrm{int}(\mathcal{S}_2) \subseteq \mathcal{S}_2$ for all $k \in \mathbb{N}$. Let $\varepsilon^k \triangleq \|y^k - y\| > 0$ for all $k \in \mathbb{N}$. Then $(\varepsilon^k)_{k\in\mathbb{N}}$ converges to 0. Now, remark that by dense intersection of $\mathcal{S}_1$ with $\mathcal{S}_2$, for all $k \in \mathbb{N}$ there exists $z^k \in \mathcal{S}_1 \cap \mathcal{S}_2 \cap \mathcal{B}_{\varepsilon^k}(y^k)$. Finally, let $\kappa \triangleq \min\{p \in \mathbb{N} : \max_{k\geq p}\|y^k - y\| \leq \frac{1}{2}\|y - x\|\}$ if $y \neq x$ and $\kappa \triangleq 0$ if $y = x$. Hence, for all $k \geq \kappa$ we have $x \notin \mathcal{B}_{\varepsilon^k}(y^k)$, and thus $z^k \in \mathcal{S}_1 \cap \mathcal{S}_2 \cap \mathcal{B}_{\varepsilon^k}(y^k) \setminus \{x\}$. Since $(z^k)_{k\in\mathbb{N}}$ converges to y, it follows that $y \in \mathrm{cl}((\mathcal{S}_1 \setminus \{x\}) \cap \mathcal{S}_2)$ as desired.

Finally, let us prove the fourth statement. Let $\mathcal{S}_1 \subseteq \mathbb{R}^n$ open, $\mathcal{S}_2$ ample and $\mathcal{S}_3$ having a dense intersection with $\mathcal{S}_1$. Then, $\mathcal{S}_1 \cap \mathcal{S}_2$ is an ample subset of $\mathcal{S}_1$ via Proposition 8.a). Thus, $\mathcal{S}_3$ has a dense intersection with $\mathcal{S}_1 \cap \mathcal{S}_2$, from Proposition 8.b) applied to $\mathcal{S}_1 \triangleq \mathcal{S}_3$ and $\mathcal{S}_2 \triangleq \mathcal{S}_1$ and $\mathcal{S}_3 \triangleq \mathcal{S}_1 \cap \mathcal{S}_2$. The claim follows from Proposition 8.c) applied to $\mathcal{S}_1 \triangleq \mathcal{S}_3$ and $\mathcal{S}_2 \triangleq \mathcal{S}_1 \cap \mathcal{S}_2$. □

Proposition 9. Under Assumption 2, consider the family $(X_i)_{i=1}^N$ it provides. For all $i \in [\![1, N]\!]$, denote by $\mathrm{cl}_f(X_i) \triangleq \{x \in \mathrm{cl}(X_i) : f_{|X_i \cup \{x\}} \text{ is continuous}\}$. Define $I(x) \triangleq \min\{i \in [\![1, N]\!] : x \in \mathrm{cl}_f(X_i)\}$ for all $x \in X$. Define $Y_i \triangleq \{x \in X : I(x) = i\}$ for all $i \in [\![1, N]\!]$. Then, $X = \sqcup_{i=1}^N Y_i$ and Y_i is an ample continuity set of f for all $i \in [\![1, N]\!]$.

Proof. Consider the notation from Proposition 9. By design, the sets $(Y_i)_{i=1}^N$ are pairwise disjoint and their union covers X, so $X = \sqcup_{i=1}^N Y_i$. Then we prove that for all $i \in [\![1, N]\!]$, Y_i is ample and $f_{|Y_i}$ is continuous. Let $i \in [\![1, N]\!]$. First, the properties of the sets $(X_i)_{i=1}^N$ and the construction of I lead to

$$\mathrm{int}(X_i) \underbrace{=}_{X_i \text{ open}} X_i \underbrace{\subseteq}_{I(X_i)=\{i\}} Y_i \underbrace{\subseteq}_{\substack{x \notin \mathrm{cl}_f(X_i) \implies I(x) \neq i \\ x \in \mathrm{cl}_f(X_i) \implies I(x) \leq i}} \mathrm{cl}_f(X_i) \underbrace{\subseteq}_{\text{by construction}} \mathrm{cl}(X_i).$$

Then, $\mathrm{int}(Y_i) \supseteq \mathrm{int}(X_i)$ and $Y_i \subseteq \mathrm{cl}(\mathrm{int}(X_i))$ so Y_i is ample. Moreover, let $x \in Y_i$ and let $(x^k)_{k\in\mathbb{N}}$ converging to x with $x^k \in Y_i$ for all $k \in \mathbb{N}$. For all $k \in \mathbb{N}$, $f_{|X_i \cup \{x^k\}}$ is continuous so there exists $(x_\ell^k)_{\ell\in\mathbb{N}}$ converging to x^k such that $x_\ell^k \in X_i$ for all $\ell \in \mathbb{N}$ and $(f(x_\ell^k))_{\ell\in\mathbb{N}}$ converges to $f(x^k)$. Let $\ell(k) \in \mathbb{N}$ such that $|f(x_{\ell(k)}^k) - f(x^k)| \leq 2^{-k}$ and $\|x_{\ell(k)}^k - x^k\| \leq 2^{-k}$. Then $(x_{\ell(k)}^k)_{k\in\mathbb{N}}$ converges to x and $x_{\ell(k)}^k \in X_i$ for all $k \in \mathbb{N}$, so $(f(x_{\ell(k)}^k))_{k\in\mathbb{N}}$ converges to $f(x)$ by continuity of $f_{|X_i \cup \{x\}}$. Hence $(f(x^k))_{k\in\mathbb{N}}$ converges to $f(x)$. Thus $f_{|Y_i}$ is continuous at x, as desired. □

Proposition 10. A set has the interior cone property from Definition 2 if and only if its complement has the exterior cone property from [20, Definition 4.1] (quoted below).

> **[20, Definition 4.1]** A set $\mathcal{S} \subseteq \mathbb{R}^n$ is said to *have the exterior cone property* if at all points $x \in \partial\mathcal{S}$ there exists a cone $\mathcal{K}_x \triangleq \{x\} + \mathbb{R}_+^*\mathcal{U}$ (with $\emptyset \neq \mathcal{U} \subseteq \mathbb{S}^n$ open in induced topology) emanating from x, a neighborhood $\mathcal{O}_x$ of x and an angle $\theta > 0$ such that $\mathcal{E}_x \subseteq \mathcal{S}^c$ and $\Theta(e-x, a-x) \geq \theta$ for all $(e, a) \in \mathcal{E}_x \times \mathcal{S}_x$, where $\Theta(\cdot,\cdot)$ computes the unsigned internal angle between two vectors and $\mathcal{S}^c \triangleq (\mathbb{R}^n \setminus \mathcal{S})$ and $\mathcal{E}_x \triangleq (\mathcal{K}_x \cap \mathcal{O}_x)$ and $\mathcal{S}_x \triangleq (\mathcal{S} \cap \mathcal{O}_x \setminus \{x\})$.

Proof. Denote by (ICP) and (ECP) the interior and exterior cone properties stated in Definition 2, and by [ECP] the exterior cone property stated in [20, Definition 4.1]. A set has the (ICP) if and only if its complement has the (ECP), so we only need to prove that the (ECP) is equivalent to the [ECP]. Denote by $[x, y] \triangleq \{x + t(y-x) : t \in [0, 1]\}$, for all $(x, y) \in (\mathbb{R}^n)^2$. Let $\mathcal{S} \subseteq \mathbb{R}^n$ and $\mathcal{S}^c \triangleq (\mathbb{R}^n \setminus \mathcal{S})$.

Assume that $\mathcal{S}$ has the [ECP]. For all $x \in \partial\mathcal{S}^c = \partial\mathcal{S}$, the [ECP] of $\mathcal{S}$ applied to x provides the cone $\mathcal{K}_x$ and the neighborhood $\mathcal{O}_x$ such that $(\mathcal{K}_x \cap \mathcal{O}_x) \subseteq \mathcal{S}^c$. Then $\mathcal{S}^c$ has the (ICP), so $\mathcal{S}$ has the (ECP). Thus the [ECP] implies the (ECP).

Assume that $\mathcal{S}$ has the (ECP). Let $x \in \partial\mathcal{S}$. The (ECP) of $\mathcal{S}$ applied to x provides $\emptyset \neq \mathcal{U} \subsetneq \mathbb{S}^n$ open in $\mathbb{S}^n$ and $\mathcal{K} \triangleq \mathbb{R}_+^*\mathcal{U}$ and the neighborhood $\mathcal{O}$ of 0 such that $(\mathcal{K}_x \cap \mathcal{O}_x) \subseteq \mathcal{S}^c$, where $\mathcal{K}_x \triangleq \{x\} + \mathcal{K}$ and $\mathcal{O}_x \triangleq \{x\} + \mathcal{O}$. Let $\theta > 0$ small enough so that the open set $\mathcal{U}' \triangleq \{u \in \mathcal{U} : \Theta(u, v) > \theta,\ \forall v \in \mathbb{S}^n \setminus \mathcal{U}\}$ is nonempty. Define $\mathcal{K}' \triangleq \mathbb{R}_+^*\mathcal{U}'$ and $\mathcal{K}'_x \triangleq \{x\} + \mathcal{K}' \subseteq \mathcal{K}_x$ and $\mathcal{E}_x \triangleq (\mathcal{K}'_x \cap \mathcal{O}_x)$ and $\mathcal{S}_x \triangleq (\mathcal{S} \cap \mathcal{O}_x \setminus \{x\})$. Then $\mathcal{E}_x \subseteq \mathcal{S}^c$ and θ satisfies the requirement in the [ECP]. Indeed, $\mathcal{E}_x \subseteq \mathcal{K}_x$ while $\mathcal{S}_x \cap \mathcal{K}_x = \emptyset$, thus for all $(e, a) \in \mathcal{E}_x \times \mathcal{S}_x$ there exists $y \in \partial\mathcal{K}_x \cap [e, a] \neq \emptyset$. Since e and y and a belong to the same line, we get $\Theta(e-x, a-x) = \Theta(e-x, y-x) + \Theta(y-x, a-x) = \Theta(\frac{e-x}{\|e-x\|}, \frac{y-x}{\|y-x\|}) + \Theta(y-x, a-x) \geq \theta + 0$. The first term is greater than θ since $\frac{e-x}{\|e-x\|} \in \mathcal{U}'$ while $\frac{y-x}{\|y-x\|} \in \mathbb{S}^n \setminus \mathcal{U}$, and the second term is positive by definition of Θ. Thus, $\mathcal{S}$ has the [ECP] at x. Hence the (ECP) implies the [ECP]. □

Proposition 11. If $\mathcal{S} \subseteq \mathbb{R}^n$ has the ICP from Definition 2, then $\mathcal{S}$ is ample. The reciprocal is not true.

Proof. Let $\mathcal{S} \subseteq \mathbb{R}^n$ having the ICP. Let $x \in \mathcal{S}$. The ICP provides $\mathcal{K}_x$ and $\mathcal{O}_x$ satisfying $(\mathcal{K}_x \cap \mathcal{O}_x) \subseteq \mathcal{S}$ and $x \in \mathrm{cl}(\mathcal{K}_x \cap \mathcal{O}_x)$. Moreover, $\mathcal{K}$ is open as the image of $\mathbb{R}_+^* \times \mathcal{U}$ (an open subset of $\mathbb{R}_+^* \times \mathbb{S}^n$) by the homeomorphism $(\lambda, u) \in \mathbb{R}_+^* \times \mathbb{S}^n \mapsto \lambda u \in \mathbb{R}^n \setminus \{0\}$. Thus, $\mathcal{K}_x \cap \mathcal{O}_x$ is also open, so $(\mathcal{K}_x \cap \mathcal{O}_x) \subseteq \mathrm{int}(\mathcal{S})$. Hence, $x \in \mathrm{cl}(\mathrm{int}(\mathcal{S}))$, which proves the direct implication. To observe that the reciprocal implication fails, consider $n \triangleq 2$ and $\mathcal{S} \triangleq \mathrm{epi}(\sqrt{|\cdot|})$. Then $\mathcal{S}$ is ample but the ICP fails at $x = (0, 0) \in \mathrm{cl}(\mathcal{S})$. $\square$

References

[1] M.A. Abramson, C. Audet, J.E. Dennis Jr, S. Le Digabel. OrthoMADS: A Deterministic MADS Instance with Orthogonal Directions. SIAM Journal on Optimization, 20(2): 948–966, 2009.

[2] C. Audet, A. Batailly, S. Kojtych. Escaping unknown discontinuous regions in blackbox optimization. SIAM Journal on Optimization, 32(3): 1843–1870, 2022.

[3] C. Audet, P.-Y. Bouchet, L. Bourdin. Counterexample and an additional revealing poll step for a result of "analysis of direct searches for discontinuous functions". Mathematical Programming, 2024.

[4] C. Audet, J.E. Dennis Jr. Analysis of generalized pattern searches. SIAM Journal on Optimization, 13(3): 889–903, 2003.

[5] C. Audet, J.E. Dennis Jr. Mesh Adaptive Direct Search Algorithms for Constrained Optimization. SIAM Journal on Optimization, 17(1): 188–217, 2006.

[6] C. Audet, J.E. Dennis Jr. A Progressive Barrier for Derivative-Free Nonlinear Programming. SIAM Journal on Optimization, 20(1): 445–472, 2009.

[7] C. Audet, W. Hare. Derivative-Free and Blackbox Optimization. Springer Series in Operations Research and Financial Engineering. Springer, Cham, Switzerland, 2017.

[8] C. Audet, W. Hare. Model-based methods in derivative-free nonsmooth optimization. In A.M. Bagirov, M. Gaudioso, N. Karmitsa, M.M. Mäkelä, S. Taheri, eds., Numerical nonsmooth optimization, chapter 15. Springer, 2020.

[9] A.S. Berahas, L. Cao, K. Scheinberg. Global convergence rate analysis of a generic line search algorithm with noise. SIAM Journal on Optimization, 31(2): 1489–1518, 2021.

[10] P.-Y. Bouchet. Théorie de l'optimisation sans dérivées dans le cas discontinu. Polytechnique Montréal, 2023.

[11] A.R. Conn, M. Mongeau. Discontinuous piecewise linear optimization. Mathematical Programming, 80(3): 315–380, 1998.

[12] A.R. Conn, K. Scheinberg, L.N. Vicente. Introduction to Derivative-Free Optimization. MOS-SIAM Series on Optimization, SIAM, Philadelphia, 2009.

[13] J.H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numerische Mathematik, 2(1): 84–90, 1960.

[14] J. Larson, M. Menickelly, S.M. Wild. Derivative-free optimization methods. Acta Numerica, 28: 287–404, 2019.

[15] S. Le Digabel, S.M. Wild. A Taxonomy of Constraints in black-box Simulation-Based Optimization. Optimization and Engineering, 25(2): 1125–1143, 2024.

[16] N. Levine. Semi-open sets and semi-continuity in topological spaces. American Mathematical Monthly, 70:36–41, 1963.

[17] A. Meijster, J.B.T.N. Roerdink, W.H. Hesselink. A General Algorithm for Computing Distance Transforms in Linear Time. In J. Goutsias, L. Vincent, D.S. Bloomberg, eds., Mathematical Morphology and its Applications to Image and Signal Processing, Springer, 2002.

[18] L. Roberts, C.W. Royer. Direct search based on probabilistic descent in reduced spaces. SIAM Journal on Optimization, 33(4): 3057–3082, 2023.

[19] R.T. Rockafellar. Generalized directional derivatives and subgradients of nonconvex functions. Canadian Journal of Mathematics, 32(2):257–280, 1980.

[20] L.N. Vicente, A.L. Custódio. Analysis of direct searches for discontinuous functions. Mathematical Programming, 133(1-2):299–325, 2012.

[21] A. Wechsung, P.I. Barton. Global optimization of bounded factorable functions with discotninuities. Journal of Global Optimization, 58: 1–30, 2014.

[22] D. Zhan, H. Xing. Expected improvement for expensive optimization: a review. Journal of Global Optimization, 78(3): 507–544, 2020.