**Les Cahiers du GERAD** ISSN: 0711–2440

# A derivative-free approach to partitioned optimization

C. Audet, P.-Y. Bouchet, L. Bourdin

G–2024–39

Juillet 2024



La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2024
– Bibliothèque et Archives Canada, 2024

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2024
– Library and Archives Canada, 2024

**GERAD** HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

**Tél. : 514 340-6053**
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

# A derivative-free approach to partitioned optimization

**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

**Juillet 2024**
**Les Cahiers du GERAD**
**G–2024–39**

**Abstract :** This work introduces a *partitioned optimization framework* (`POf`) to ease the solving process for optimization problems for which fixing some variables to a tunable value removes some difficulties. The variables space may be continuously partitioned into subsets where these variables are fixed to a given value, so that minimizing the objective function restricted to any of the partition sets is easier than minimizing the objective function over the whole variables space. Then the major difficulty is translated from solving the original problem to choosing the partition set minimizing the minimum of the objective function restricted to this set. Hence, a local solution to the original problem is given by computing this partition set and selecting the minimizer of the objective function restricted to this set. This work formalizes this framework, studies its theoretical guarantees, and provides a numerical method to seek for an optimal partition set using a derivative-free algorithm. Some illustrative problems are presented to show how to apply the `POf` and to highlight the gain in numerical performance it provides.

**Résumé :** Cet article propose un *cadre d'optimisation partitionnée* (*partitioned optimization framework*, `POf`) visant à simplifier le processus de résolution d'un problème d'optimisation tel que l'ajout de contraintes fixant certaines variables supprime des difficultés. L'espace des variables est continûment partitionné en sous-ensembles sur lesquels ces variables sont fixées à des valeurs données, de sorte que la minimisation de la fonction objectif à n'importe quel des sous-espaces de partition est plus simple que sur l'espace des variable entier. La difficulté principale du problème est alors transférée de la résolution du problème originel vers le choix de l'ensemble de partition pour lequel le minimum de la fonction objectif restreinte à cet ensemble est le plus petit possible. Le minimiseur de la fonction objectif restreinte à cet ensemble de partition est alors une solution locale au problème originel. Cet article formalise cette approche, étudie ses propriétés théoriques, et fournit une méthode numérique pour identifier un ensemble de partition optimal via un algorithme d'optimisation sans dérivées. Plusieurs exemples sont proposées pour illustrer comment ce cadre théorique s'emploit en pratique et quels gains de performance il apporte.

**Acknowledgements:** This work was supported by the first NSERC discovery grant RGPIN–2020–04448 of Audet.

# 1 Introduction

This work introduces a *partitioned optimization framework* (POf) to handle low-dimensional difficulties in an optimization problem, such as an objective function discontinuous with respect to a few variables. These difficulties are addressed as a problem from *derivative-free optimization* (DFO), while specialized methods handle all others aspects of the problem. First, the POf partitions the variables space according to these difficulties, so that the problem restricted to any partition set is simple enough to be solved with a dedicated method. Then, the POf identifies an optimal partition set, that is, a partition set for which the solution to the associated restricted problem has the lowest objective function value. Last, the POf returns the solution to the problem restricted to this optimal partition set, as this solution also solves the unrestricted problem. Section 1.1 discusses what motivated this work. Section 1.2 formalizes the POf. Section 1.3 reviews related concepts in the literature. Section 1.4 outlines this work.

**Notation:** We denote by $\overline{\mathbb{R}} \triangleq \mathbb{R}\cup\{\pm\infty\}$, $\overline{\mathbb{N}} \triangleq \mathbb{N}\cup\{+\infty\}$, $\mathbb{R}^* \triangleq \mathbb{R}\setminus\{0\}$ and $\overline{\mathbb{N}}^* \triangleq \overline{\mathbb{N}}\setminus\{0\}$. Consider two normed spaces $\mathbb{X}$ and $\mathbb{Y}$. A set $X \subseteq \mathbb{X}$ is said to be *ample* if $X \subseteq \text{cl}(\text{int}(X))$ (where cl and int denote respectively the closure and interior operators), and to be a *continuity set* of a function $\Phi : \mathbb{X} \to \mathbb{R}$ if $\Phi_{|X}$ is continuous. For all collections $(X_i)_{i=1}^N$ of subsets of $\mathbb{X}$, $N \in \overline{\mathbb{N}}^*$, their union is denoted by $\sqcup_{i=1}^N X_i$ when the sets are pairwise disjoint. For all sequences $(x^k)_{k\in\mathbb{N}} \in \mathbb{X}^\mathbb{N}$, we denote by $\text{acc}((x^k)_{k\in\mathbb{N}})$ the set of all accumulation points of $(x^k)_{k\in\mathbb{N}}$. For all functions $\gamma : \mathbb{X} \to \mathbb{Y}$ and all points $x \in \mathbb{X}$, we denote by $\mathcal{ACC}(\gamma; x)$ the union of $\text{acc}((\gamma(x^k))_{k\in\mathbb{N}})$ over all sequences $(x^k)_{k\in\mathbb{N}} \in \mathbb{X}^\mathbb{N}$ converging to $x$.

## 1.1 Motivation and motivating example

Consider for example the function $\varphi : \mathbb{R}^2 \to \mathbb{R}$ shown in Figure 1 (see details in Section 3.1). Although $\varphi$ is discontinuous, observe that, for all $y_1 \in \mathbb{R}$, the restricted function $\varphi_{|\{y_1\}\times\mathbb{R}}$ is quadratic, so its unique global minimizer, denoted by $\gamma(y_1)$, is easily tractable. Then, the POf computes the minimizer of $\varphi$ over $\mathbb{R}^2$ by minimizing over $y_1$, provided that $y_2$ follows accordingly so that $(y_1, y_2) = \gamma(y_1)$. In others words, the POf partitions $\mathbb{R}^2$ as $\mathbb{R}^2 = \sqcup_{y_1\in\mathbb{R}}\{y_1\}\times\mathbb{R}$ and seeks for the value $y_1^* \triangleq 0$ minimizing $\varphi(\gamma(y_1))$ (where $\gamma(y_1)$ minimizes $\varphi_{|\{y_1\}\times\mathbb{R}}$). The value $\gamma(y_1^*) = (0, 0)$ minimizes either $\varphi_{|\{y_1^*\}\times\mathbb{R}}$ and $\varphi$.

Several problems may be solved similarly via the POf, such as problems -i- nonsmooth with respect to a few variables; or -ii- made convex by fixing a few variables; or -iii- partially separable (as they become separated by fixing the variables linking the separated subproblems); or -iv- made smooth by introducing some constraints linking the variables. These properties may notably appear in large-dimensional and challenging problems. Our motivation to formalize the POf is to solve such problems by using a DFO algorithm only to address the low-dimensional difficulties.

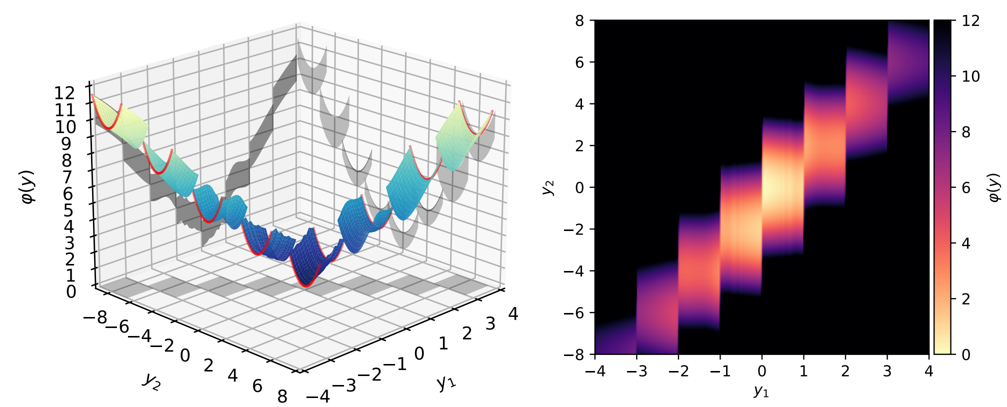


**Figure 1: $\varphi$ is a discontinuous alteration of $(y_1, y_2) \mapsto y_2^2$ via a perturbation depending on $y_1$ only, minimized efficiently via the POf. By fixing $y_1$, the resulting restricted function $\varphi_{|\{y_1\}\times\mathbb{R}}$ is a quadratic of $y_2$ easy to minimize. Minimizing over $y_1$, with $y_2$ always fixed accordingly at the minimizer of the associated restricted function $\varphi_{|\{y_1\}\times\mathbb{R}}$, is easier than minimizing over $(y_1, y_2)$ directly. For readability, at each $y_1 \in \mathbb{R}$ the two graphs actually plot only the low values of $\varphi_{|\{y_1\}\times\mathbb{R}}$.**

### 1.2 Contributions

Consider the generic optimization problem

$$\underset{y \in \mathbb{Y}}{\text{minimize}} \quad \varphi(y) \quad \text{subject to} \quad y \in \Omega, \tag{$\mathbf{P}_{\texttt{ini}}$}$$

where $\mathbb{Y}$ is a (possibly infinite-dimensional) normed space, $\varphi : \mathbb{Y} \to \overline{\mathbb{R}}$ is the (possibly discontinuous) objective function and $\Omega \subseteq \mathbb{Y}$ is a closed set intersecting $Y \triangleq \{y \in \mathbb{Y} : \varphi(y) \neq \pm\infty\} \neq \emptyset$. Our main contribution consists in a strategy to solve Problem ($\mathbf{P}_{\texttt{ini}}$) via the `POf` given in Framework 1.

**Framework 1** (partitioned optimization framework (`POf`))**.** Partition $\mathbb{Y}$ as $\mathbb{Y} = \sqcup_{x \in \mathbb{X}} \mathbb{Y}(x)$, where the normed space $\mathbb{X}$ has a finite dimension, such that the *subproblem associated to Problem* ($\mathbf{P}_{\texttt{ini}}$) *at* $x$,

$$\underset{y \in \mathbb{Y}(x)}{\text{minimize}} \quad \varphi(y) \quad \text{subject to} \quad y \in \Omega, \tag{$\mathbf{P}_{\texttt{sub}}(x)$}$$

admits a nonempty set of global solutions, denoted by $\Gamma(x)$, for all $x \in X \triangleq \{x \in \mathbb{X} : \mathbb{Y}(x) \cap Y \cap \Omega \neq \emptyset\}$. Define an *oracle function* $\gamma : X \to Y$, satisfying $\gamma(x) \in \Gamma(x)$ for all $x \in X$ and assumed accessible. Define also the *index function* $\chi : \mathbb{Y} \to \mathbb{X}$ which, for all $y \in \mathbb{Y}$, maps $y$ to the unique element $x \in \mathbb{X}$ for which $y$ belongs to $\mathbb{Y}(x)$. Then consider the *reformulation of Problem* ($\mathbf{P}_{\texttt{ini}}$),

$$\underset{x \in \mathbb{X}}{\text{minimize}} \quad \Phi(x), \qquad \text{where} \qquad \Phi : \left\{ \begin{array}{ccl} \mathbb{X} & \to & \overline{\mathbb{R}} \\ x & \mapsto & \left\{ \begin{array}{ll} \varphi(\gamma(x)) & \text{if } x \in X, \\ +\infty & \text{otherwise.} \end{array} \right. \end{array} \right. \tag{$\mathbf{P}_{\texttt{ref}}$}$$

We now introduce two assumptions, required for the algorithmic use of the `POf`.

**Assumption 1.** The function $\chi$ is continuous and, for all $\mathcal{Y} \subseteq \mathbb{Y}$ bounded, $\chi(\mathcal{Y})$ is bounded.

**Assumption 2.** The function $\varphi_{|\gamma(X)}$ is bounded below and has its sublevel sets bounded; and there exists a partition $X = \sqcup_{i=1}^{N} X_i$, where $N \in \overline{\mathbb{N}}^*$, such that, for each $i \in [\![1, N]\!]$, $X_i$ is ample, $\gamma_{|X_i}$ is uniformly continuous and $\varphi_{|\gamma(X_i)}$ is continuous. Moreover, for all $x \in X$, there exists a neighbourhood $\mathcal{X} \subseteq \mathbb{X}$ of $x$ such that $X_i \cap \mathcal{X} \neq \emptyset$ only for a finite number of indices $i \in [\![1, N]\!]$.

We consider the `POf` under these two assumptions and the `cDSM` [3] applied to Reformulation ($\mathbf{P}_{\texttt{ref}}$). The `cDSM` is discussed in Section 2.1. Theorem 1 ensures that the `cDSM` generates a sequence $(x^k)_{k \in \mathbb{N}}$ such that $(\gamma(x^k))_{k \in \mathbb{N}}$ has accumulation points solving Problem ($\mathbf{P}_{\texttt{ini}}$) locally, in a generalized sense.

**Theorem 1.** Consider that Framework 1 is applicable and Assumptions 1 and 2 hold. The `cDSM` solving Reformulation ($\mathbf{P}_{\texttt{ref}}$) generates a sequence $(x^k)_{k \in \mathbb{N}}$ such that, for all $x^* \in \mathrm{acc}((x^k)_{k \in \mathbb{N}}) \neq \emptyset$, there exists a neighbourhood $\mathcal{X}^* \subseteq \mathbb{X}$ of $x^*$ such that, for all subsequences $(x^k)_{k \in K^*}$ converging to $x^*$,

$$\emptyset \neq \mathrm{acc}\left((\gamma(x^k))_{k \in K^*}\right) \subseteq \mathcal{Y}^* \cap \Omega \quad \text{and} \quad \lim_{k \in K^*} \varphi(\gamma(x^k)) = \inf \varphi\left(\mathcal{Y}^* \cap \Omega\right),$$

where $\mathcal{Y}^* \triangleq \mathbb{Y}(\mathcal{X}^*)$ is a neighbourhood of all points of $\mathrm{acc}((\gamma(x^k))_{k \in K^*})$. If moreover $x^* \in X$ and $\varphi$ is lower-semicontinuous at all points of $\mathcal{ACC}(\gamma; x^*)$, then $\gamma(x^*)$ is a local solution to Problem ($\mathbf{P}_{\texttt{ini}}$).

Our auxiliary contribution is to show that Assumptions 1 and 2 are required only to analyze the `cDSM`. Precisely, even if Assumptions 1 and 2 do not hold, there exist some connections between solutions to Reformulation ($\mathbf{P}_{\texttt{ref}}$) and those to Problem ($\mathbf{P}_{\texttt{ini}}$). Theorem 2 formalizes these connections.

**Theorem 2.** Consider that Framework 1 is applicable. Then,

a) for each $x^*$ a global solution to Reformulation ($\mathbf{P}_{\texttt{ref}}$), $\gamma(x^*)$ is a global solution to Problem ($\mathbf{P}_{\texttt{ini}}$);
b) for each $x^*$ a local solution to Reformulation ($\mathbf{P}_{\texttt{ref}}$), if $\chi$ is continuous at $\gamma(x^*)$, then $\gamma(x^*)$ is a local solution to Problem ($\mathbf{P}_{\texttt{ini}}$).

Let us stress that ensuring the accessibility of the oracle function $\gamma$ is not within our scope. We assume that, for all $x \in X$, we possess an idealized optimization method able to solve Subproblem ($\mathbf{P}_{\texttt{sub}}(x)$) globally. A future work (that we announce in Section 5.2) will relax this requirement.

### 1.3 Literature review

This work originates from the corresponding author's PhD thesis [9, Chapters 5 and 6] (in French). The thesis also provides an application of the POf to a class of discontinuous optimal control problems [9, Chapter 7]. In this context, in all cases considered, $\mathbb{Y}$ is an infinite-dimensional functional space while the POf allows for $\mathbb{X} = \mathbb{R}^{d_{\mathbb{X}}}$ with $d_{\mathbb{X}} \leq 4$. An early version of [9, Chapter 7] appears in a technical report [2], without the formal theory supporting the POf.

Solving Problem ($\mathbf{P}_{\mathtt{ini}}$) directly with a DFO algorithm is likely inefficient, since the dimension $d_{\mathbb{Y}}$ of $\mathbb{Y}$ may be high while DFO algorithms usually handle up to a few dozens of variables [6, Section 1.4]. The **NOMAD** [7] and **PRIMA** [23] solvers both state a critical value of 50 variables. Some dimension-selection techniques [14] handle some large-scale problems. They rely on *active subspaces* [11, 22], and they likely outperform the POf when the objective function has a small sensitivity with respect to most of the variables. However, active subspaces fail to describe cases where the objective function has a small sensitivity with respect to each variables individually but has a high sensitivity with respect to all variables jointly. For example, [9, Chapter 7] shows that this happens in some optimal control problems: an alteration of any single state variable has a small impact, but an alteration of all state variables together significantly alters the objective value. The POf is suited for cases without active subspaces, where it is easier to solve Reformulation ($\mathbf{P}_{\mathtt{ref}}$) with a DFO algorithm than Problem ($\mathbf{P}_{\mathtt{ini}}$).

The POf is not compatible with the subcase of DFO named *blackbox optimization*, but is related to *greybox optimization* [1]. Indeed, we require that Subproblem ($\mathbf{P}_{\mathtt{sub}}(x)$) may be explicitly constructed for all $x \in \mathbb{X}$, since we must attempt to solve it to determine whether or not $x \in X$ and what $\gamma(x)$ is. Nevertheless, we solve Reformulation ($\mathbf{P}_{\mathtt{ref}}$) with the *covering Direct Search Method* (cDSM) [3], suited for blackbox optimization. The cDSM extends the *Direct Search Method* (DSM) [6, Part 3] to return a local solution in a possibly discontinuous context under a mild assumption. Most DFO algorithms are designed for blackbox contexts, since adapting them to a greybox case likely requires a problem-dependent strategy. We provide ideas in Remark 2 to specialize the cDSM in the context of the POf.

The POf is also related to *parametric optimization* [8, 21], which studies problems where the objective function and the feasible set depend on a tunable parameter. Any problem fitting in any of the two frameworks may fit in the other. The local continuity of $\gamma$ in a discontinuous context is studied in [15], but, to the best of our knowledge, no reference in parametric optimization claims results similar to Theorem 1. Moreover, the terminology *partitioned optimization* highlights that Problem ($\mathbf{P}_{\mathtt{ini}}$) is non-parametric and that the nature of the so-called parameter of Subproblem ($\mathbf{P}_{\mathtt{sub}}(x)$) (the index of the associated partition set of $\mathbb{Y}$, which results from the chosen partition of $\mathbb{Y}$) is left to the user.

The POf also shares similarities with *bilevel optimization* [12], which describes problems in which some of the variables are fixed as a solution to a nested subproblem. Any problem expressed in any of the frameworks may be transferred to the other. However, the difference in terminology highlights a contextual difference since Problem ($\mathbf{P}_{\mathtt{ini}}$) has only one level. Moreover, either Reformulation ($\mathbf{P}_{\mathtt{ref}}$) and Subproblem ($\mathbf{P}_{\mathtt{sub}}(x)$) are minimization problems, while a bilevel problem usually deals with a conflictual structure. Some references in bilevel optimization study nonsmooth problems [19] and apply DSM [13] but, to the best of our knowledge, none provides results similar to Theorem 1.

The terminology *partitioned optimization* sometimes describes the DFO methods [17, 18]. However, in these methods, the partition is *discrete* and the partition sets do not simplify the restricted problem, as they partition an hyperrectangle into smaller ones and recursively partition the promising ones. Accordingly, we believe that our *continuous* and structure-simplifying partition may fit under the same terminology of partitioned optimization without confusion.

The POf may also be related to *optimization on manifold* [10, 16], which considers problems admitting a constraint inducing an explicit manifold on the variables space. When each partition set of $\mathbb{Y}$ is a manifold, Subproblem ($\mathbf{P}_{\mathtt{sub}}(x)$) is an optimization on manifold problem for all $x \in \mathbb{X}$. In practice, when $\mathbb{Y}$ has a finite dimension, it is likely that $\mathbb{Y}$ is indeed partitioned into a continuum of manifolds.

### 1.4 Outline of this work

This work is organized as follows. Section 2 discusses the `cDSM` [3] and proves Theorems 1 and 2. Section 3 proposes some illustrative problems that fit into the `POf`, and shows that the claims formalized by Theorem 1 hold numerically. Section 4 alters the problems from Section 3 into more challenging versions with a larger dimension, and compares two `DFO` solvers from the literature (solving the non-reformulated problems) to a naive implementation of the `cDSM` relying on the `POf`. We observe that, thanks to the `POf`, the reformulated problem has a much lower dimension than the non-reformulated problem since the dimension of $\mathbb{X}$ remains at most 10 while we test problems with $\mathbb{Y}$ having a dimension up to 100. Then, even a naive solver solving the reformulation outperforms powerful solvers solving the non-reformulated problem. Hence the `POf`, when applicable, allows a significant decrease in computation time to solve problems with large number of variables. Finally Section 5 discusses about the `POf` and its possible extensions.

## 2 The `cDSM` and the proofs of Theorems 1 and 2

First, Section 2.1 discusses the `cDSM` [3]. Its convergence analysis, adapted in the notation of the current work, is recalled as Theorem 3. Then Section 2.2 proves Theorems 1 and 2.

### 2.1 The `cDSM`

Algorithm 1 stands as a `cDSM` expressed in its simplest form given in [3, Algorithm 2], and [3, Theorem 2] (reported in the context of this work as Theorem 3) formalizes its convergence. We provide the simplest algorithm for ease of presentation only. The detailed algorithmic framework for the `cDSM` is given in [3, Algorithm 1], and Theorem 3 remains valid for this framework.

**Algorithm 1 Simplified `cDSM` [3, Algorithm 2] solving Reformulation ($\mathbf{P}_{\texttt{ref}}$).**

**Initialization**:
set $(x^0, \delta^0) \in X \times \mathbb{R}^*_+$ the initial (incumbent solution, poll radius) couple;
set $(\lambda, \upsilon) \in ]0, 1[ \times [1, +\infty[$ the poll radius shrinking and expanding parameters;
set $\mathbb{O}$ a *covering oracle* [3, Definition 1], set $\mathcal{H}^0 \triangleq \emptyset$ the initial trial points history;

**for** $k \in \mathbb{N}$ **do**:
**`covering` step**:
set $\mathcal{D}^k_{\texttt{C}} \triangleq \mathbb{O}(x^k, \mathcal{H}^k)$; set $\mathcal{T}^k_{\texttt{C}} \triangleq \{x^k\} + \mathcal{D}^k_{\texttt{C}}$ and $t^k_{\texttt{C}} \in \operatorname{argmin} \Phi(\mathcal{T}^k_{\texttt{C}})$;
if $\Phi(t^k_{\texttt{C}}) < \Phi(x^k)$, then set $t^k \triangleq t^k_{\texttt{C}}$ and $\mathcal{T}^k_{\texttt{S}} = \mathcal{T}^k_{\texttt{P}} \triangleq \emptyset$ and go to the `update` step;
**`search` step**:
set $\mathcal{D}^k_{\texttt{S}} \subseteq \mathbb{X}$ empty or finite; if $\mathcal{T}^k_{\texttt{S}} \triangleq \{x^k\} + \mathcal{D}^k_{\texttt{S}}$ is nonempty, then set $t^k_{\texttt{S}} \in \operatorname{argmin} \Phi(\mathcal{T}^k_{\texttt{S}})$;
if also $\Phi(t^k_{\texttt{S}}) < \Phi(x^k)$, then set $t^k \triangleq t^k_{\texttt{S}}$ and $\mathcal{T}^k_{\texttt{P}} \triangleq \emptyset$ and go to the `update` step;
**`poll` step**:
set $\mathcal{D}^k_{\texttt{P}} \subseteq \mathbb{X}$ a positive basis of length $\delta^k$; set $\mathcal{T}^k_{\texttt{P}} \triangleq \{x^k\} + \mathcal{D}^k_{\texttt{P}}$; set $t^k_{\texttt{P}} \in \operatorname{argmin} \Phi(\mathcal{T}^k_{\texttt{P}})$;
if $\Phi(t^k_{\texttt{P}}) < \Phi(x^k)$, then set $t^k \triangleq t^k_{\texttt{P}}$, otherwise set $t^k \triangleq x^k$;
**`update` step**:
set $(x^{k+1}, \delta^{k+1}) \triangleq (t^k, \upsilon\delta^k)$ if $t^k \neq x^k$ or $(x^{k+1}, \delta^{k+1}) \triangleq (x^k, \lambda\delta^k)$ if $t^k = x^k$;
set $\mathcal{H}^{k+1} \triangleq \mathcal{H}^k \cup \mathcal{T}^k_{\texttt{C}} \cup \mathcal{T}^k_{\texttt{S}} \cup \mathcal{T}^k_{\texttt{P}}$.

**Theorem 3** (convergence analysis of Algorithm 1)**.** Assume that $\Phi$ is bounded below with bounded sublevel sets and that $X$ admits a partition into $N \in \overline{\mathbb{N}}^*$ ample continuity sets of $\Phi$. Algorithm 1 solving Reformulation ($\mathbf{P}_{\texttt{ref}}$) generates a sequence $(x^k)_{k\in\mathbb{N}}$ such that, for all $x^* \in \operatorname{acc}((x^k)_{k\in\mathbb{N}}) \neq \emptyset$, there exists a neighbourhood $\mathcal{X}^* \subseteq \mathbb{X}$ of $x^*$ such that, for all subsequences $(x^k)_{k\in K^*}$ converging to $x^*$, the sequence $(\Phi(x^k))_{k\in K^*}$ converges to $\inf \Phi(\mathcal{X}^*)$, and moreover $\inf \Phi(\mathcal{X}^*) = \min \Phi(\mathcal{X}^*) = \Phi(x^*)$ holds if $\Phi$ is lower semicontinuous at $x^*$.

**Remark 1.** In Sections 3 and 4, we consider instances of Algorithm 1 with the following parameters. The point $x^0$ depends on the instance. We state $\delta^0 \triangleq 1$ in all cases, while $(\lambda, \upsilon) \triangleq (\frac{1}{2}, 1)$ when $\mathbb{X} = \mathbb{R}$ in easy cases (Sections 3.1, 3.2, 3.3, 4.1 and 4.2) and $(\lambda, \upsilon) \triangleq (\frac{1}{2}, 2)$ when a large interval of $\mathbb{X} = \mathbb{R}$ is explored (in Section 4.3), and $(\lambda, \upsilon) \triangleq (\frac{3}{4}, 2)$ otherwise (Sections 3.4 and 4.4). The `covering` step relies on $r \triangleq \frac{1}{10}$ and defines, for all $k \in \mathbb{N}$, $\mathbb{O}(x^k, \mathcal{H}^k)$ as a random point on the closed ball of radius $r$ centered at $x^k$. In all cases, the `search` step is skipped ($\mathcal{D}^k_{\mathtt{S}} \triangleq \emptyset$ for all $k \in \mathbb{N}$) and the `poll` step computes random orthogonal positive bases. The algorithm stops at the first $k \in \mathbb{N}$ such that $\delta^k < 10^{-10}$.

**Remark 2.** A sensitivity analysis of Subproblem ($\mathbf{P}_{\mathtt{sub}}(x)$) with respect to $x$ via a generalized implicit function theorem [20] locally describes the oracle function $\gamma$. This may provide a descent direction for $\Phi$ emanating from $x$. Also, a *surrogate* of the problem exists when $\gamma(x)$ is obtained numerically for all $x \in X$, by approximating $\gamma(x)$ as the output of a method solving ($\mathbf{P}_{\mathtt{sub}}(x)$) on low precision.

## 2.2 Proofs of Theorems 1 and 2

We first prove our auxiliary contribution, Theorem 2, and then our main contribution, Theorem 1. The proof of the latter is based on some lemmas and on the application of Theorems 2 and 3. Theorem 2 is proved in Section 2.2.1. The lemmas are stated in Section 2.2.2. The proof that Theorem 3 is applicable in our context follows in Section 2.2.3. Finally, Theorem 1 is proved in Section 2.2.4.

### 2.2.1 Proof of Theorem 2

Let $x^*$ be a global solution to Reformulation ($\mathbf{P}_{\mathtt{ref}}$). Then, $\gamma(x^*) \in \Omega$ and, for all $y \in Y \cap \Omega$, it holds that $\varphi(y) \geq \varphi(\gamma(\chi(y))) = \Phi(\chi(y)) \geq \Phi(x^*) = \varphi(\gamma(x^*))$. This proves Theorem 2.a).

Now let $x^*$ be a local solution to Reformulation ($\mathbf{P}_{\mathtt{ref}}$) (so $\gamma(x^*) \in \Omega$) and assume that $\chi$ is continuous at $\gamma(x^*)$. Thus, there exists a neighbourhood $\mathcal{X} \subseteq \mathbb{X}$ of $x^*$ such that $\Phi(x) \geq \Phi(x^*)$ for all $x \in \mathcal{X}$, and a neighbourhood $\mathcal{Y} \subseteq \mathbb{Y}$ of $\gamma(x^*)$ such that $\chi(\mathcal{Y}) \subseteq \mathcal{X}$. Then, for all $y \in \mathcal{Y} \cap \Omega$, it holds that $\varphi(y) \geq \varphi(\gamma(\chi(y))) = \Phi(\chi(y)) \geq \Phi(x^*) = \varphi(\gamma(x^*))$. This proves Theorem 2.b). □

### 2.2.2 Lemmas on the oracle and index functions

**Lemma 1.** If $\chi$ is continuous, then, for all $\mathcal{X} \subseteq \mathbb{X}$ open, $\mathbb{Y}(\mathcal{X}) \subseteq \mathbb{Y}$ is open.

**Proof.** Assume that $\chi$ is continuous. Let $\emptyset \neq \mathcal{X} \subseteq \mathbb{X}$ open and let $y \in \mathbb{Y}(\mathcal{X})$. Then $\chi(y) \in \mathcal{X}$, and there exists $\mathcal{Y} \subseteq \mathbb{Y}$ open containing $y$ and such that $\chi(\mathcal{Y}) \subseteq \mathcal{X}$. Thus, $y \in \mathcal{Y} \subseteq \mathbb{Y}(\mathcal{X})$. □

**Lemma 2.** The oracle function $\gamma : X \to Y$ is a bijection from $X$ to $\gamma(X)$, and its reciprocal is the restriction of the index function $\chi : \mathbb{Y} \to \mathbb{X}$ to $\gamma(X)$.

**Proof.** For all couples $(x_1, x_2) \in X^2$ with $x_1 \neq x_2$, it holds that $Y(x_1) \cap Y(x_2) = \emptyset$, and $\gamma(x_1) \in \mathbb{Y}(x_1)$ and $\gamma(x_2) \in \mathbb{Y}(x_2)$. It follows that $\gamma(x_1) \neq \gamma(x_2)$. Hence $\gamma$ is an injection from $X$ to $\gamma(X)$, and thus a bijection. Moreover, first, for all $x \in X$, the equality $\chi(\gamma(x)) = x$ holds by definition of $\chi$ applied to $\gamma(x) \in \mathbb{Y}(x)$. Second, for all $y \in \gamma(X)$, the equality $\gamma(\chi(y)) = y$ holds since there exists a unique $x \in X$ satisfying $y = \gamma(x) \in \mathbb{Y}(x)$, and then $\chi(y) = x$ so $\gamma(\chi(y)) = \gamma(x) = y$. □

**Lemma 3.** Under Assumption 2, for all $x \in \mathrm{cl}(X)$ and all sequences $(x^k)_{k\in\mathbb{N}}$ of elements of $X$ converging to $x$, there exists $i \in [\![1, N]\!]$ such that $K_i \triangleq \{k \in \mathbb{N} : x^k \in X_i\}$ is infinite. Moreover, $(\gamma(x^k))_{k\in K_i}$ has a limit for each $i \in [\![1, N]\!]$ such that $K_i$ is infinite.

**Proof.** Let $x \in \mathrm{cl}(X)$ and $(x^k)_{k\in\mathbb{N}}$ converging to $x$ with $x^k \in X$ for all $k \in \mathbb{N}$. Denote by $\mathcal{X} \subseteq \mathbb{X}$ the neighbourhood of $x$ provided by Assumption 2. Without loss of generality, assume that $x^k \in \mathcal{X}$ for all $k \in \mathbb{N}$. Define $K_i \triangleq \{k \in \mathbb{N} : x^k \in X_i\}$ for all $i \in [\![1, N]\!]$, and by $\mathcal{I} \triangleq \{i \in [\![1, N]\!] : X_i \cap \mathcal{X} \neq \emptyset\}$, so that $i \notin \mathcal{I}$ implies that $K_i = \emptyset$. Then, $\mathbb{N} = \sqcup_{i\in[\![1,N]\!]} K_i = \sqcup_{i\in\mathcal{I}} K_i$ by construction, and $\mathcal{I}$ is finite by Assumption 2. It follows that at least one element $i \in \mathcal{I}$ satisfies $K_i$ infinite. Moreover, for all $i \in [\![1, N]\!]$ such that $K_i$ is infinite, $(\gamma(x^k))_{k\in K_i}$ has a limit by uniform continuity of $\gamma_{|X_i}$. □

#### 2.2.3 Verification that the requirements of Theorem 3 hold under Assumptions 1 and 2

**Proposition 1.** Consider Framework 1. If Assumptions 1 and 2 hold, then the function $\Phi$ is bounded below and has its sublevel sets bounded, and for each $i \in [\![1, N]\!]$, $X_i$ is an ample continuity set of $\Phi$.

**Proof.** Let $\alpha \in \mathbb{R}$. For all $x \in \mathbb{X}$, the inequality $\Phi(x) \leq \alpha$ is equivalent to $\varphi(\gamma(x)) \leq \alpha$, and then to $\gamma(x) \in \varphi^{-1}_{|\gamma(X)}(]-\infty, \alpha])$. By Lemma 2, the latter is equivalent to $x \in \chi(\varphi^{-1}_{|\gamma(X)}(]-\infty, \alpha]))$. Moreover, the set $\varphi^{-1}_{|\gamma(X)}(]-\infty, \alpha])$ is bounded by Assumption 2, so the set $\chi(\varphi^{-1}_{|\gamma(X)}(]-\infty, \alpha]))$ is bounded by Assumption 1. Hence, $\Phi(x) \leq \alpha$ if and only if $x$ lies in the bounded set $\chi(\varphi^{-1}_{|\gamma(X)}(]-\infty, \alpha]))$. Moreover, $\inf \Phi(\mathbb{X}) = \inf \Phi(X) = \inf \varphi(\gamma(X)) > -\infty$ by Assumption 2, so $\Phi$ is bounded below. This proves the first claim. Now consider the partition of $X$ provided by Assumption 2. For all $i \in [\![1, N]\!]$, $X_i$ is ample by assumption and $\Phi_{|X_i} = (\varphi_{|\gamma(X_i)}) \circ (\gamma_{|X_i})$ is continuous as the composition of two continuous functions. This proves the second claim. □

**Proposition 2.** Consider Framework 1 and let $x \in X$. If $\varphi$ is lower semicontinuous at all points of $\mathcal{ACC}(\gamma; x)$, then $\Phi$ is lower semicontinuous at $x$.

**Proof.** Let $x \in X$ and assume that $\varphi$ is lower semicontinuous at all points of $\mathcal{ACC}(\gamma; x)$. Let $\mathcal{X} \subseteq \mathbb{X}$ be a neighbourhood of $x$ and let $(z^k)_{k \in \mathbb{N}}$ converging to $x$ with all elements in $\mathcal{X} \cap X$. By Lemma 3, we have $\emptyset \neq \mathrm{acc}((\gamma(z^k))_{k \in \mathbb{N}}) \subseteq \mathcal{ACC}(\gamma; x)$. Let $y \in \mathrm{acc}((\gamma(z^k))_{k \in \mathbb{N}})$ and $K \subseteq \mathbb{N}$ such that $(\gamma(z^k))_{k \in K}$ converges to $y$. So, $\liminf_{k \in K} \Phi(z^k) = \liminf_{k \in K} \varphi(\gamma(z^k)) \geq \varphi(y)$ by lower semicontinuity of $\varphi$ at $y$, and $\varphi(y) \geq \varphi(\gamma(x)) = \Phi(x)$ since $\chi(y) = x \in X$. Hence $\Phi$ is lower semicontinuous at $x$. □

#### 2.2.4 Proof of Theorem 1

Consider Framework 1 under Assumptions 1 and 2. This proof relies on Theorems 2.b) and 3. Proposition 1 ensures that Theorem 3 is applicable in our context. Then, consider a sequence $(x^k)_{k \in \mathbb{N}}$ generated by `cDSM` solving Reformulation ($\mathbf{P}_{\texttt{ref}}$). According to Theorem 3, for all $x^* \in \mathrm{acc}((x^k)_{k \in \mathbb{N}}) \neq \emptyset$, there exists an open neighbourhood $\mathcal{X}^* \subseteq \mathbb{X}$ of $x^*$ satisfying, for all subsequences $(x^k)_{k \in K^*}$ converging to $x^*$,

$$
\begin{aligned}
\lim_{k \in K^*} \Phi(x^k) \;&=\; \inf \Phi(\mathcal{X}^*), \\
\text{moreover} \quad \inf \Phi(\mathcal{X}^*) \;&=\; \min \Phi(\mathcal{X}^*) \;=\; \Phi(x^*) \quad \text{if } \Phi \text{ is lower semicontinuous at } x^*.
\end{aligned} \tag{1}
$$

Let $x^* \in \mathrm{acc}((x^k)_{k \in \mathbb{N}})$ and $K^* \subseteq \mathbb{N}$ such that $(x^k)_{k \in K^*}$ converges to $x^*$. Denote by $\mathcal{Y}^* \triangleq \mathbb{Y}(\mathcal{X}^*)$, and note that it is an open subset of $\mathbb{Y}$ according to Lemma 1.

**Accumulation points of $(\gamma(x^k))_{k \in K^*}$.** Lemma 3 proves that $\mathrm{acc}((\gamma(x^k))_{k \in K^*}) \neq \emptyset$. It remains to prove that $\mathrm{acc}((\gamma(x^k))_{k \in K^*}) \subseteq Y^* \cap \Omega$. Let $y \in \mathrm{acc}((\gamma(x^k))_{k \in K^*})$ and $K \subseteq K^*$ such that $(\gamma(x^k))_{k \in K}$ converges to $y$. It holds that $\chi(y) = \lim_{k \in K} \chi(\gamma(x^k)) = \lim_{k \in K} x^k$ by continuity of $\chi$ and by Lemma 2. Then the equality $\chi(y) = x^*$ holds, so $y \in \mathbb{Y}(x^*) \subset \mathcal{Y}^*$. Finally, $y \in \Omega$ since $\gamma(x^k) \in \Omega$ for all $k \in K^*$ and $\Omega$ is closed.

**Convergence of $(\varphi(\gamma(x^k)))_{k \in K^*}$.** For all $k \in K^*$, $\varphi(\gamma(x^k)) = \Phi(x^k)$ holds by construction, and $\lim_{k \in K^*} \Phi(x^k) = \inf \Phi(\mathcal{X}^*)$ holds according to (1). Then, $\lim_{k \in K^*} \varphi(\gamma(x^k)) = \inf \Phi(\mathcal{X}^*)$. Moreover, the equality $\inf \Phi(\mathcal{X}^*) = \inf \varphi(\mathcal{Y}^* \cap \Omega)$ holds by construction. Indeed, for all $y \in \mathcal{Y}^* \cap \Omega$, it holds that $\varphi(y) \geq \varphi(\gamma(\chi(y))) = \Phi(\chi(y)) \geq \inf \Phi(\mathcal{X}^*)$ since $\chi(y) \in \mathcal{X}^*$, and reciprocally, for all $x \in \mathcal{X}^*$, there exists $x' \in \mathcal{X}^*$ such that $\Phi(x) \geq \Phi(x') = \varphi(\gamma(x')) \geq \inf \varphi(\mathcal{Y}^* \cap \Omega)$ since $\gamma(x') \in \mathcal{Y}^* \cap \Omega$. It follows that $\lim_{k \in K^*} \varphi(\gamma(x^k)) = \inf \varphi(\mathcal{Y}^* \cap \Omega)$.

**Optimality of $\gamma(x^*)$ if $x^* \in X$ and $\varphi$ is lower semicontinuous at all points of $\mathcal{ACC}(\gamma; x^*)$.** Assume that $x^* \in X$ and $\varphi$ is lower semicontinuous at all points of $\mathcal{ACC}(\gamma; x^*)$. Then $\Phi$ is lower semicontinuous at $x^*$ by Proposition 2. Thus (1) implies that $x^*$ is a local solution to Reformulation ($\mathbf{P}_{\texttt{ref}}$), so Theorem 2.b) ensures that $\gamma(x^*)$ is a local solution to Problem ($\mathbf{P}_{\texttt{ini}}$) since $\chi$ is continuous at $\gamma(x^*)$ thanks to Assumption 1. □

# 3 Illustrative examples on problems with a noisy objective function

In this section, four illustrative problems are solved by application of Theorem 1. Their scope focuses to the class of problems where $\Omega$ is described by smooth constraints and $\varphi$ is the sum of a smooth function $\widetilde{\varphi} : \mathbb{Y} \to \mathbb{R}$ that is easy to minimize and of a noise function $\varepsilon : \mathbb{R}^d \to \mathbb{R}$ depending on $d$ smooth combinations of $y$. That is, $\varphi(y) = \varepsilon(f(y)) + \widetilde{\varphi}(y)$ for all $y \in \mathbb{Y}$, where $\widetilde{\varphi} : \mathbb{Y} \to \mathbb{R}$ is smooth, $\varepsilon : \mathbb{R}^d \to \mathbb{R}$ is nonsmooth, and $f : \mathbb{Y} \to \mathbb{R}^d$ is a $d$-dimensional explicit smooth function mapping the variables $y$ to the inputs of the noise function $\varepsilon$. In this context, Problem ($\mathbf{P}_{\texttt{ini}}$) is hard to solve because of the noise $\varepsilon$, while minimizing $\widetilde{\varphi}$ would be easier. However, $\varepsilon$ becomes constant if the problem is restricted by the addition of an equality constraint $f(y) = x$ with $x \in \mathbb{R}^d$ fixed. Then, the `POf` is applicable. A natural partition of $\mathbb{Y}$ is to define $\mathbb{X} \triangleq \mathbb{R}^d$ and set $\mathbb{Y}(x) \triangleq \{y \in \mathbb{Y} : f(y) = x\}$ for all $x \in \mathbb{X}$. Indeed, for all $x \in \mathbb{X}$, the restriction of $\varepsilon$ to $\mathbb{Y}(x)$ is constant (equal to the value $\varepsilon(x)$) and the constraint $y \in \mathbb{Y}(x)$ is an usual smooth nonlinear equality constraint. Hence, for all $x \in X$, Subproblem ($\mathbf{P}_{\texttt{sub}}(x)$) is a smooth problem with an (assumed accessible) global solution $\gamma(x)$. Consequently Reformulation ($\mathbf{P}_{\texttt{ref}}$) is a $d$-dimensional nonsmooth problem that handles the noise $\varepsilon$, since $\Phi = \widetilde{\varphi} \circ \gamma + \varepsilon$.

In Section 3.1, $\varphi$ is the motivational example from Section 1.1, where $\varepsilon$ depends only on the first variable and $\gamma$ is directly tractable. In Section 3.2, $\varepsilon$ depends only on the radius of the polar coordinates, and again $\gamma$ is tractable analytically. In Section 3.3, $\varepsilon$ depends on all the variables, but more accurately it depends on a single nonlinear combination of them. In this case, $\gamma$ requires some analytical additional calculation to be found. Finally, in Section 3.4, $\varepsilon$ depends on two nonlinear combinations involving all the variables. In this last problem, for all $x \in X$, $\gamma(x)$ is not analytically tractable and Subproblem ($\mathbf{P}_{\texttt{sub}}(x)$) is actually not a smooth optimization problem, but $\gamma(x)$ may be solved globally by an algorithmic method. On each example, eight instances of the `cDSM` are tested. Their algorithmic parameters and stopping criterion are described in Remark 1.

These four problems highlight that Theorem 1 may claim various results. In Sections 3.1 and 3.3, the claim about the local optimality of $\gamma(x^*)$ is not applicable since its requirements are not satisfied. Nevertheless this claim is actually true in Section 3.1, which shows that Theorem 1 is not tight. In Sections 3.2 and 3.4, either the requirements and the conclusion of all claims from Theorem 1 hold.

## 3.1 Mono-variable noise

Consider the function

$$\varphi : \left\{ \begin{array}{ccc} \mathbb{Y} = Y \triangleq \mathbb{R}^2 & \to & \mathbb{R} \\ y = (y_1, y_2) & \mapsto & (y_2 - \sigma(y_1))^2 + \varepsilon(y_1), \end{array} \right.$$

where $\varepsilon(z) \triangleq |z| \left(1 + \sin(\frac{2\pi}{z})^2\right)^{\frac{1}{2}} + |\lfloor z \rceil|$ if $z \in \mathbb{R}^*$ and $\varepsilon(0) \triangleq 0$, where $\sigma(z) \triangleq 2\lfloor z \rceil$, and where $\lfloor z \rceil \triangleq \lfloor z \rfloor$ if $z \in \mathbb{R}_-$ and $\lfloor z \rceil \triangleq \lceil z \rceil - 1$ if $z \in \mathbb{R}_+^*$. The unique global minimizer of $\varphi$ is $y^* \triangleq (0, 0)$, with $\varphi(y^*) = 0$. In this context, we define $\mathbb{Y}(x) \triangleq \{x\} \times \mathbb{R}$ for all $x \in \mathbb{X} = X \triangleq \mathbb{R}$. For all $x \in \mathbb{X}$, $\varphi_{|\mathbb{Y}(x)}$ is a quadratic function of $y_2$ while $y_1 = x$ is fixed. Its unique minimizer is therefore

$$\gamma(x) \triangleq (x, \sigma(x)), \quad \text{with} \quad \varphi(\gamma(x)) = \varepsilon(x).$$

Figure 2 shows $\varphi$, the partition of $\mathbb{Y}$ and the locus of $\gamma$, and Figure 3 (left) shows $\varepsilon$. Then we consider

$$\Phi : \left\{ \begin{array}{ccc} \mathbb{X} & \to & \mathbb{R} \\ x & \mapsto & \varepsilon(x), \end{array} \right.$$

shown on Figure 3 (right). Its global minimizer is $x^* \triangleq 0$, with $\Phi(x^*) = 0$, and $\Phi$ is left-discontinuous at $x^*$. Assumptions 1 and 2 hold. We test eight instances of the `cDSM` minimizing $\Phi$ from eight starting points. Table 1 shows that they all approach $x^*$ closely and from the side avoiding the discontinuity. They all return $\widetilde{x}^* \in [0, 2E^{-10}]$. Hence they all provide $\widetilde{y}^* \triangleq \gamma(\widetilde{x}^*)$ such that $\|\widetilde{y}^* - y^*\| \leq 2E^{-10}$ and $|\varphi(\widetilde{y}^*) - \varphi(y^*)| \leq 2E^{-10}$. This empirically confirms that Theorem 1 holds. We also remark that $\gamma(x^*)$ is a local solution to Problem ($\mathbf{P}_{\texttt{ini}}$), even if the requirement in Theorem 1 that $\varphi$ is lower-semicontinuous at all points in $\mathcal{ACC}(\gamma; x^*)$ does not hold in this context (as $\varphi$ is not lower-semicontinuous at $(0, -1) \in \mathcal{ACC}(\gamma; x^*)$).

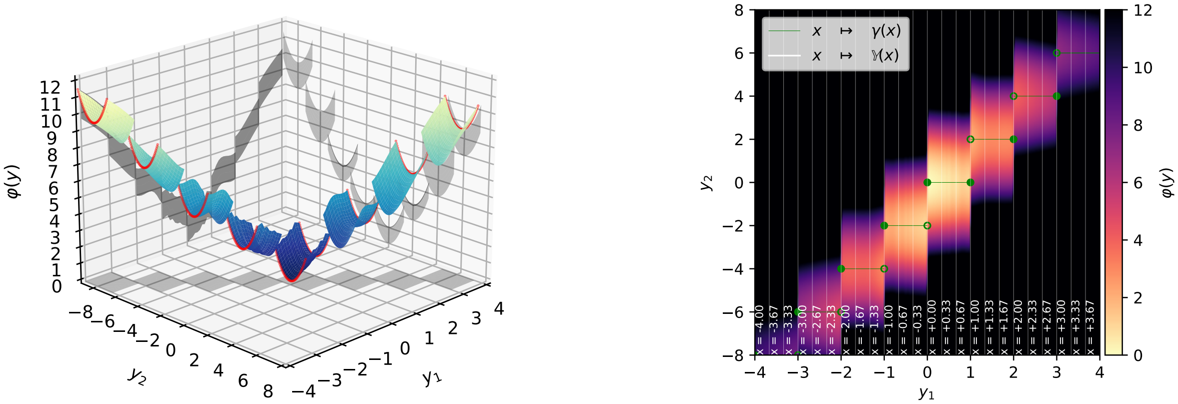


**Figure 2: function $\varphi$ in Section 3.1, (left) 3D view, (right) 2D view, partition of $\mathbb{Y}$ and locus of $\gamma$. For each $y_1 \in \mathbb{R}$, we plot $\varphi_{|\{y_1\}\times\mathbb{R}}$ for the values $y_2 \in [\sigma(y_1) - 1.3,\ \sigma(y_1) + 1.3]$ only.**

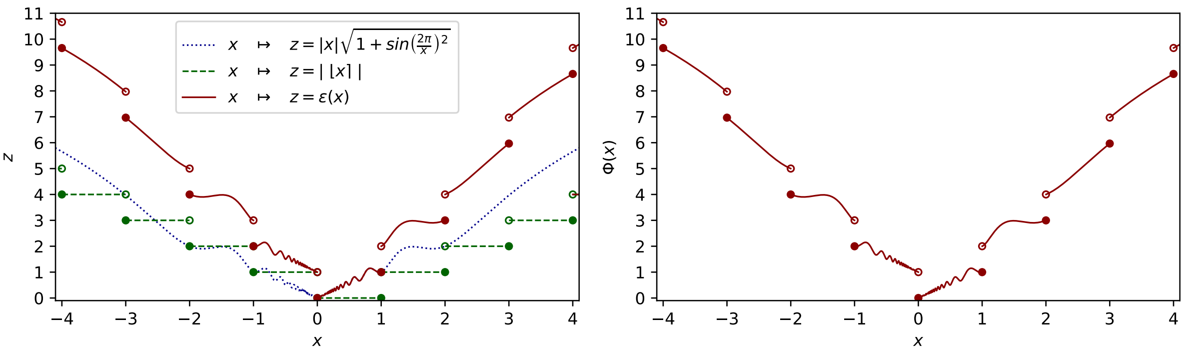


**Figure 3: functions $\varepsilon$ and $\Phi$ in Section 3.1, (left) function $\varepsilon$ and its components, (right) function $\Phi$. In this section, $\Phi = \varepsilon$.**

**Table 1: results for eight instances of the `cDSM` in the context of Section 3.1. In this table, $\hat{x}^k \triangleq x^k$.**

| $x^0$ | $1^{st}\ \hat{x}^k \in \left[\pm 5E^{-03}\right]$ | $1^{st}\ \hat{x}^k \in \left[\pm 5E^{-06}\right]$ | $1^{st}\ \hat{x}^k \in \left[\pm 5E^{-09}\right]$ | returned $\hat{x}^k$ |
|---|---|---|---|---|
| $+9.753$ | $\hat{x}^{23} = +3.37E^{-03}$ | $\hat{x}^{37} = +8.86E^{-07}$ | $\hat{x}^{55} = +3.15E^{-09}$ | $\hat{x}^{64} = +7.41E^{-12}$ |
| $+\pi$ | $\hat{x}^{18} = +3.12E^{-03}$ | $\hat{x}^{31} = +3.93E^{-06}$ | $\hat{x}^{49} = +1.00E^{-09}$ | $\hat{x}^{56} = +6.89E^{-11}$ |
| $+\sqrt{2}$ | $\hat{x}^{13} = +4.65E^{-03}$ | $\hat{x}^{29} = +2.79E^{-06}$ | $\hat{x}^{44} = +3.08E^{-09}$ | $\hat{x}^{54} = +5.12E^{-11}$ |
| $+e+1$ | $\hat{x}^{17} = +2.08E^{-03}$ | $\hat{x}^{24} = +3.01E^{-06}$ | $\hat{x}^{45} = +3.29E^{-09}$ | $\hat{x}^{53} = +3.45E^{-11}$ |
| $-9.753$ | $\hat{x}^{17} = +4.61E^{-04}$ | $\hat{x}^{33} = +3.33E^{-06}$ | $\hat{x}^{51} = +3.96E^{-09}$ | $\hat{x}^{60} = +4.05E^{-12}$ |
| $-\pi$ | $\hat{x}^{20} = +2.94E^{-03}$ | $\hat{x}^{33} = +3.46E^{-06}$ | $\hat{x}^{43} = +2.33E^{-10}$ | $\hat{x}^{54} = +1.16E^{-10}$ |
| $-\sqrt{2}$ | $\hat{x}^{06} = +3.45E^{-03}$ | $\hat{x}^{27} = +1.13E^{-06}$ | $\hat{x}^{43} = +1.51E^{-09}$ | $\hat{x}^{52} = +7.67E^{-13}$ |
| $-e-1$ | $\hat{x}^{21} = +2.74E^{-03}$ | $\hat{x}^{36} = +2.40E^{-06}$ | $\hat{x}^{49} = +4.02E^{-09}$ | $\hat{x}^{57} = +5.70E^{-11}$ |

## 3.2 Radial noise

Consider the function expressed in polar coordinates by

$$\varphi : \left\{ \begin{array}{rcl} \mathbb{Y} = Y \triangleq \mathbb{R}_+ \times [0, 2\pi[ & \to & \mathbb{R} \\ y = (r, \theta) & \mapsto & \sqrt{r} \sin\left(\dfrac{\theta - \pi - 2\pi \log_2(r)}{2}\right)^2 + \varepsilon(r) \text{ if } r > 0 \text{ and } \varepsilon(0) \text{ otherwise,} \end{array} \right.$$

where $\varepsilon(z) \triangleq \sqrt{|z^2 - 2|} + \frac{\sin(10\pi(z - \sqrt{2}))^2}{10}$ for all $z \in \mathbb{R}_+$. The global minimizer of $\varphi$ is $y^* \triangleq (\sqrt{2}, 0)$, with $\varphi(y^*) = 0$. We define $\mathbb{Y}(x) \triangleq \{x\} \times [0, 2\pi[$ for all $x \in \mathbb{X} = X \triangleq \mathbb{R}_+$. Then, for all $x \in X$, $\varphi_{|\mathbb{Y}(x)}$ is a $2\pi$-periodic smooth function of $\theta$ while $r = x$ is fixed. Its unique global minimizer for $\theta \in [0, 2\pi[$ is

$$\gamma(x) \triangleq (x, (\pi + 2\pi \log_2(x)) \sim 2\pi) \quad \text{with} \quad \varphi(\gamma(x)) = \varepsilon(x),$$

where $z \sim 2\pi$ denotes the residual of $z \in \mathbb{R}$ modulo $2\pi$. Figure 4 shows $\varphi$, the partition of $\mathbb{Y}$ and the locus of $\gamma$, and Figure 5 (left) shows $\varepsilon$. Then the function $\Phi$ equals

$$\Phi : \left\{ \begin{array}{ccc} \mathbb{X} & \to & \mathbb{R} \\ x & \mapsto & \varepsilon(x), \end{array} \right.$$

shown on Figure 5 (right). Its minimizer is $x^* \triangleq \sqrt{2}$, with $\Phi(x^*) = 0$. Assumptions 1 and 2 hold. Table 2 shows the convergence of the cDSM towards $x^*$. All instances return $\widetilde{x}^* \in [x^* \pm 6E^{-11}]$, and then $\widetilde{y}^* \triangleq \gamma(\widetilde{x}^*)$ satisfies $\|\widetilde{y}^* - y^*\| \leq 6E^{-11}$ and $|\varphi(\widetilde{y}^*) - \varphi(y^*)| \leq 3E^{-5}$. This agrees with Theorem 1.

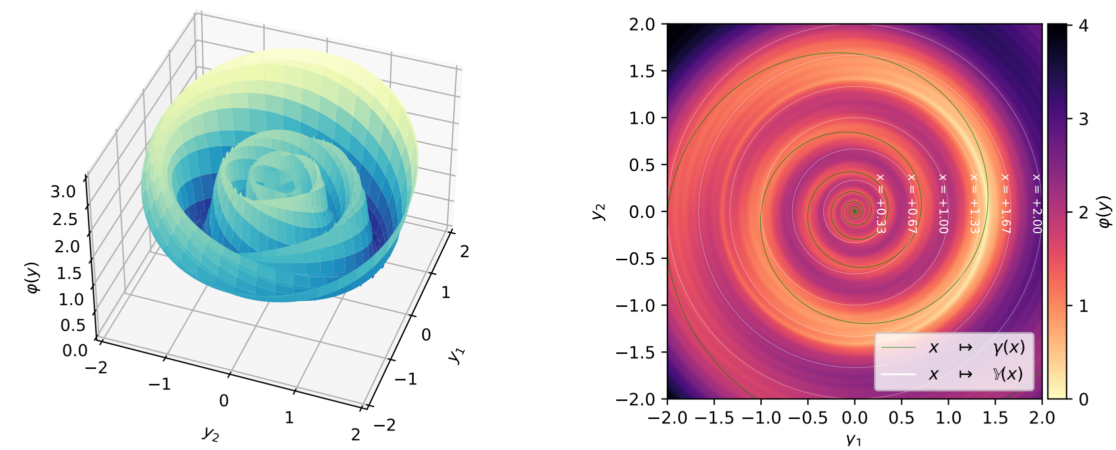


**Figure 4: function $\varphi$ in Section 3.2, (left) 3D view, (right) 2D view, partition of $\mathbb{Y}$ and locus of $\gamma$.**

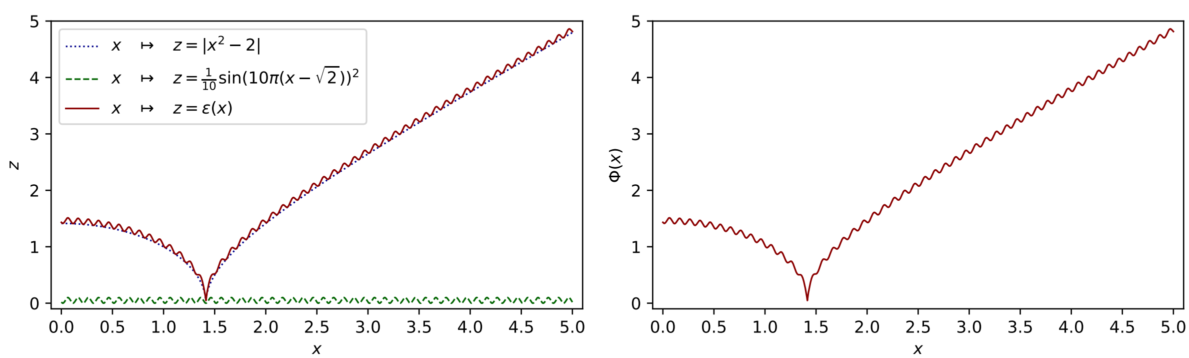


**Figure 5: functions $\varepsilon$ and $\Phi$ in Section 3.2, (left) function $\varepsilon$ and its components, (right) function $\Phi$. In this section, $\Phi = \varepsilon$.**

**Table 2: results for eight instances of the cDSM in the context of Section 3.2. In this table, $\hat{x}^k \triangleq x^k - \sqrt{2}$.**

| $x^0$ | $1^{st}$ $\hat{x}^k \in \left[\pm 5E^{-03}\right]$ | $1^{st}$ $\hat{x}^k \in \left[\pm 5E^{-06}\right]$ | $1^{st}$ $\hat{x}^k \in \left[\pm 5E^{-09}\right]$ | returned $\hat{x}^k$ |
|---|---|---|---|---|
| $0$ | $\hat{x}^{09} = +4.39E^{-03}$ | $\hat{x}^{29} = -8.47E^{-07}$ | $\hat{x}^{41} = +2.73E^{-09}$ | $\hat{x}^{52} = +5.39E^{-11}$ |
| $2^{-5}$ | $\hat{x}^{09} = +5.08E^{-04}$ | $\hat{x}^{27} = +4.42E^{-06}$ | $\hat{x}^{44} = -3.57E^{-09}$ | $\hat{x}^{54} = +3.83E^{-11}$ |
| $3\sqrt{2}$ | $\hat{x}^{17} = -1.14E^{-03}$ | $\hat{x}^{35} = -1.51E^{-06}$ | $\hat{x}^{51} = -1.95E^{-09}$ | $\hat{x}^{60} = +2.94E^{-11}$ |
| $4\pi$ | $\hat{x}^{25} = +1.37E^{-03}$ | $\hat{x}^{39} = +1.50E^{-06}$ | $\hat{x}^{57} = +2.05E^{-09}$ | $\hat{x}^{66} = -4.99E^{-11}$ |
| $5$ | $\hat{x}^{08} = +6.72E^{-05}$ | $\hat{x}^{27} = -1.45E^{-06}$ | $\hat{x}^{41} = -1.81E^{-09}$ | $\hat{x}^{48} = +5.71E^{-11}$ |
| $e$ | $\hat{x}^{08} = -1.75E^{-03}$ | $\hat{x}^{28} = +9.13E^{-07}$ | $\hat{x}^{41} = +4.04E^{-09}$ | $\hat{x}^{53} = -3.28E^{-11}$ |
| $e^2$ | $\hat{x}^{11} = -5.45E^{-04}$ | $\hat{x}^{25} = +4.01E^{-06}$ | $\hat{x}^{43} = +5.27E^{-10}$ | $\hat{x}^{53} = -5.47E^{-11}$ |
| $e^3$ | $\hat{x}^{30} = +3.83E^{-03}$ | $\hat{x}^{44} = +4.66E^{-06}$ | $\hat{x}^{59} = -4.19E^{-09}$ | $\hat{x}^{69} = +3.72E^{-12}$ |

### 3.3 Noise affected by a nonlinear combination of all variables

Consider the function

$$\varphi : \left\{ \begin{array}{rcl} \mathbb{Y} \triangleq \mathbb{R}^2 & \to & \mathbb{R} \\ y = (y_1, y_2) & \mapsto & \ln\left(1 + \left(\dfrac{y_1^2}{y_2^2+1} - 1\right)^2\right) + \varepsilon(y_1 y_2) \text{ if } y \in \Omega \triangleq \mathbb{R}_+ \times \mathbb{R}, \text{ else } +\infty, \end{array} \right.$$

where $\varepsilon(z) \triangleq \exp(\frac{1}{z-4}) + \frac{\sqrt{|z-4|}}{5}$ if $z \neq 4$ and $\varepsilon(4) \triangleq +\infty$. Here $\varphi$ has no global minimizer, but $\inf \varphi(\Omega) = 0$ and $\varphi(y) \to 0$ when $y \to y^* \approx (2.1287, 1.8791)$ from some directions. However, approaching $y^*$ from some others directions raises $\varphi(y) \to +\infty = \varphi(y^*)$. The exact value of $y^*$ and the directions are given below. We define $\mathbb{Y}(x) \triangleq \{y = (y_1, y_2) \in \mathbb{Y} : y_1 y_2 = x\}$ for all $x \in \mathbb{X} = X \triangleq \mathbb{R}$. For all $x \in \mathbb{X}$, $\varphi_{|\mathbb{Y}(x)}$ is smooth (since $y_1 y_2 = x$ so $\varepsilon(y_1 y_2)$ is constant) and its minimizer is computable. First, $y_1^2/(y_2^2+1) = 1$ provides $y_2^2 = y_1^2 - 1$ (and thus $y_1^2 \geq 1$ so $y_1 \neq 0$), and $(y_1, y_2) \in \mathbb{Y}(x) \cap \Omega$ raises $y_2 = x/y_1$ and $y_1 \geq 0$. Second, equalling the two expressions of $y_2^2$ provides $(y_1^2)^2 - y_1^2 - x^2 = 0$, with one admissible value for $y_1^2$ as $y_1^2 = \frac{1}{2}(1 + \sqrt{1+4x^2})$ since $y_1^2 \geq 1$. Hence $y_1$ follows (since $y \in \Omega$ forces $y_1 \geq 0$), and then $y_2$ follows as well. Then, for all $x \in \mathbb{X}$,

$$\gamma(x) \triangleq \left( \sqrt{\frac{1+\sqrt{1+4x^2}}{2}}, x\sqrt{\frac{2}{1+\sqrt{1+4x^2}}} \right), \quad \text{with} \quad \varphi(\gamma(x)) = \varepsilon(x).$$

Figure 6 shows $\varphi$, the partition of $\mathbb{Y}$ and the locus of $\gamma$, and Figure 7 (left) shows $\varepsilon$. Remark that $\gamma$ is continuous, even if $\varphi$ is not. In this context, $\Phi$ equals

$$\Phi : \left\{ \begin{array}{rcl} \mathbb{X} & \to & \mathbb{R} \\ x & \mapsto & \varepsilon(x), \end{array} \right.$$

shown on Figure 7 (right). Then $\Phi$ has no minimum but its infimum is 0, and $x^* \triangleq 4$ is such that $\lim_{x \nearrow x^*} \Phi(x) = 0$ despite $\lim_{x \searrow x^*} \Phi(x) = +\infty$. Thus, $y^* \triangleq \gamma(x^*)$ and $\lim_{y=\gamma(x \nearrow x^*)} \varphi(y) = 0$ while $\lim_{y=\gamma(x \searrow x^*)} \varphi(y) = \varphi(y^*) = +\infty$. Assumptions 1 and 2 hold.

We test eight instances of the `cDSM` minimizing $\Phi$. Table 3 shows their convergence towards $x^*$ up to three thresholds. All the instances starting with $x^0 < x^*$ return $\widetilde{x}^* \in [x^* - 2E^{-10}, x^*]$, which approaches $x^*$ closely and from the direction minimizing $\Phi$. Hence they provide $\widetilde{y}^* \triangleq \gamma(\widetilde{x}^*) \approx y^*$ with $\varphi(\widetilde{y}^*) \approx \inf \varphi(Y)$, which agrees with Theorem 1. The instance starting with $x^0 \triangleq 3\sqrt{2} \approx 4.24 > x^*$ behaves similarly. The last two instances, who start with $x^0 \gg 4$, converge instead to $\widetilde{x}^* \approx 9.27$, but this is also consistent with Theorem 1 since $\Phi$ indeed has a local minimizer around 9.27.

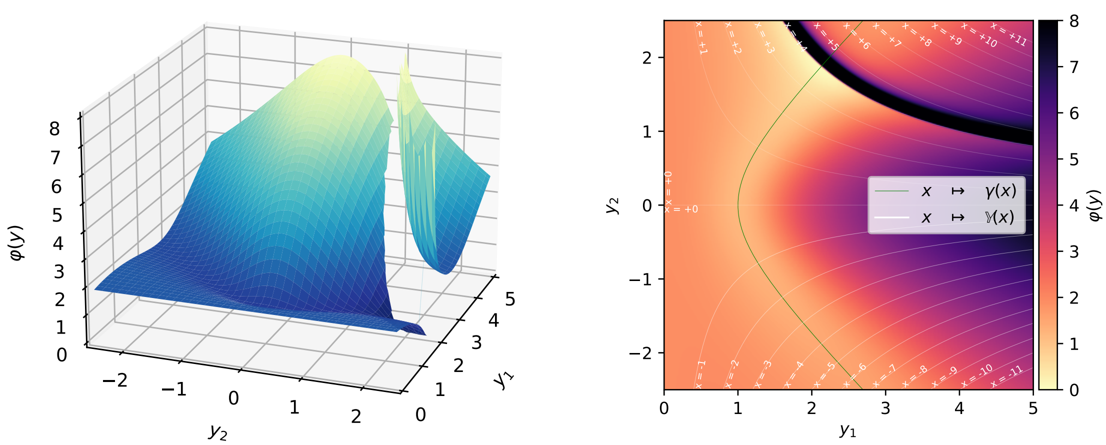


**Figure 6: function $\varphi$ in Section 3.3, (left) 3D view, (right) 2D view, partition of $\mathbb{Y}$ and the locus of $\gamma$.**

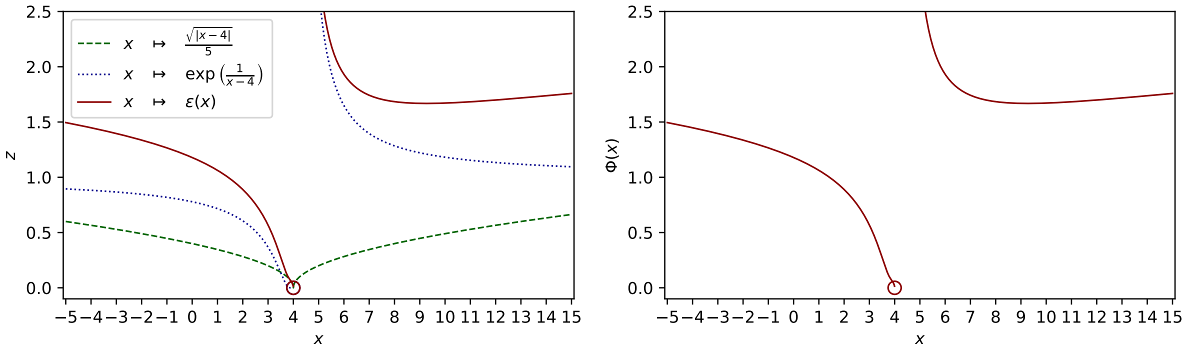


**Figure 7: functions $\varepsilon$ and $\Phi$ in Section 3.3, (left) function $\varepsilon$ and its components, (right) function $\Phi$. In this section, $\Phi = \varepsilon$. Note that neither $\varepsilon$ nor $\Phi$ have a global minimum.**

**Table 3: results for eight instances of Algorithm 1 in the context of Section 3.3. In this table, $\hat{x}^k \triangleq x^k - 4$. For each of the last three instances, we arbitrarily return $x^{201} \approx 9.26779505$ since $(x^k)_{k\geq 40}$ has its first 9 digits constant in all cases.**

| $x^0$ | $1^{st}$ $\hat{x}^k \in \left[\pm 5E^{-03}\right]$ | $1^{st}$ $\hat{x}^k \in \left[\pm 5E^{-06}\right]$ | $1^{st}$ $\hat{x}^k \in \left[\pm 5E^{-09}\right]$ | returned $\hat{x}^k$ |
|---|---|---|---|---|
| $-e^2$ | $\hat{x}^{107} = -4.17E^{-03}$ | $\hat{x}^{120} = -2.20E^{-06}$ | $\hat{x}^{138} = -1.68E^{-09}$ | $\hat{x}^{146} = -4.65E^{-11}$ |
| $-\pi$ | $\hat{x}^{026} = -1.76E^{-03}$ | $\hat{x}^{043} = -1.70E^{-06}$ | $\hat{x}^{057} = -3.88E^{-09}$ | $\hat{x}^{065} = -4.14E^{-11}$ |
| $-\sqrt{2}$ | $\hat{x}^{024} = -4.25E^{-04}$ | $\hat{x}^{040} = -4.95E^{-06}$ | $\hat{x}^{051} = -1.40E^{-09}$ | $\hat{x}^{062} = -8.01E^{-12}$ |
| $+e$ | $\hat{x}^{009} = -4.68E^{-04}$ | $\hat{x}^{026} = -2.78E^{-06}$ | $\hat{x}^{043} = -3.13E^{-09}$ | $\hat{x}^{051} = -1.06E^{-10}$ |
| $+3\sqrt{2}$ | $\hat{x}^{017} = -2.66E^{-03}$ | $\hat{x}^{030} = -1.33E^{-07}$ | $\hat{x}^{046} = -2.41E^{-09}$ | $\hat{x}^{053} = -8.40E^{-11}$ |
| $+2e^2$ | / | / | / | $\hat{x}^{201} = +5.27E^{+00}$ |
| $+4\pi$ | / | / | / | $\hat{x}^{201} = +5.27E^{+00}$ |
| $+e^3$ | / | / | / | $\hat{x}^{201} = +5.27E^{+00}$ |

## 3.4 Bidimensional noise with non-analytical oracle

Consider the function

$$\varphi : \left\{ \begin{array}{rcl} \mathbb{Y} = Y \triangleq \mathbb{R}^3 & \to & \mathbb{R} \\ y = (y_1, y_2, y_3) & \mapsto & \|y\|_\infty + \varepsilon\left(y_2 - y_1^3, y_1 - y_3^3\right), \end{array} \right.$$

where $\varepsilon(z_1, z_2) \triangleq (\frac{\sin(10\pi(z_2 - z_1^3))}{5} + \frac{\sin(6\pi(z_2 - e^{-z_1} + 1))}{7} + \frac{\sin(12\pi(z_1^2 + z_2^2)^{\frac{1}{2}})}{11})^2$ for all $z \in \mathbb{R}^2$. The global minimizer of $\varphi$ is $y^* \triangleq (0,0,0)$, with $\varphi(y^*) = 0$. In this context, we define the partition sets of $\mathbb{Y}$ as $\mathbb{Y}(x) \triangleq \{(y_1, y_2, y_3) \in \mathbb{Y} : y_2 - y_1^3 = x_1 \text{ and } y_1 - y_3^3 = x_2\} = \{(t, t^3 + x_1, \sqrt[3]{t - x_2}) : t \in \mathbb{R}\}$ for all $x \in \mathbb{X} = X \triangleq \mathbb{R}^2$. Then, for all $x \in \mathbb{X}$, $\varepsilon_{|\mathbb{Y}(x)}$ is constant and the unique minimizer of $\varphi_{|\mathbb{Y}(x)}$ is the solution to the problem

$$\underset{y \in \mathbb{Y}(x)}{\text{minimize}} \quad \varepsilon(x) + \max\left\{|t_x(y)|, \left|t_x(y)^3 + x_1\right|, \left|\sqrt[3]{t_x(y) - x_2}\right|\right\},$$

where, for any $y \in \mathbb{Y}(x)$, $t_x(y)$ denotes the unique $t \in \mathbb{R}$ such that $s_x(t) \triangleq (t, t^3 + x_1, \sqrt[3]{t - x_2}) = y$. This solution is hardly tractable. However, the problem admits a smooth exact reformulation given by

$$\begin{array}{cl} \underset{M \geq 0,\ T \in \mathbb{R}}{\text{minimize}} & \varepsilon(x) + M \\ \text{subject to} & T \in I_x^1(M) \triangleq [-M, M], \\ & T \in I_x^2(M) \triangleq \left[\sqrt[3]{-M - x_1}, \sqrt[3]{M - x_1}\right], \\ & T \in I_x^3(M) \triangleq \left[-M^3 + x_2, M^3 + x_2\right]. \end{array}$$

Its unique solution, denoted by $(M(x), T(x))$, provides

$$\gamma(x) \triangleq s_x(T(x)), \quad \text{with} \quad \varphi(\gamma(x)) = \varepsilon(x) + M(x).$$

Figure 8 shows $\varphi$, Figure 9 shows the partition of $\mathbb{Y}$ and the locus of $\gamma$, and Figure 10 (left) and (center) represent respectively $\varepsilon$ and $M$. In this context, $\Phi$ equals

$$\Phi : \left\{ \begin{array}{ccc} \mathbb{X} & \to & \mathbb{R} \\ x & \mapsto & \varepsilon(x) + M(x), \end{array} \right.$$

shown on Figure 10 (right). Its minimizer is $x^* \triangleq (0,0)$, with $\Phi(x^*) = 0$. Assumptions 1 and 2 hold.

Note that $(M(x), T(x))$ is intractable analytically. Nevertheless, we observe that

$$\begin{array}{rcl} M(x) & = & \min\left\{ M \in \mathbb{R}_+ \ : \ I_x(M) \triangleq I^1_x(M) \cap I^2_x(M) \cap I^3_x(M) \neq \emptyset \right\}, \\ \{T(x)\} & = & I_x\left(M(x)\right), \end{array}$$

so $M(x)$ is obtainable via a dichotomic search and the singleton $\{T(x)\}$ follows. In practice, running sufficiently many iterations of the dichotomic search provides an accurate approximation of $M(x)$. We define $M^0_{\text{inf}} \triangleq \lfloor M(x) \rfloor$ and $M^0_{\text{sup}} \triangleq M^0_{\text{inf}} + 1$, and $\ell \triangleq 0$, and we iterate the dichotomic search on $[M^\ell_{\text{inf}}, M^\ell_{\text{sup}}]$ while $M^\ell_{\text{sup}} - M^\ell_{\text{inf}} > 2^{-30}$. This allows to define $M(x) \approx \hat{M}(x) \triangleq \frac{1}{2}(M^\ell_{\text{inf}} + M^\ell_{\text{sup}})$, and then $T(x) \approx \hat{T}(x) \triangleq \frac{1}{2}(\min I_x(\hat{M}(x)) + \max I_x(\hat{M}(x)))$, and finally $\gamma(x) \approx \hat{\gamma}(x) \triangleq s_x(\hat{T}(x))$.

A total of eight `cDSM` instances are tested, where $\hat{\Phi} \triangleq \varepsilon + \hat{M}$ is minimized since we approximate $M \approx \hat{M}$ for tractability. Table 4 shows the convergence of these instances towards $x^*$. They all return $\|\widetilde{x}^* - x^*\| \leq 9E^{-7}$, hence they provide the value $\widetilde{y}^* \triangleq \hat{\gamma}(\widetilde{x}^*)$ such that $\|\widetilde{y}^* - y^*\| \leq 9E^{-7}$ and $|\varphi(\widetilde{y}^*) - \varphi(y^*)| \leq 9E^{-7}$. Five instances even return $\|\widetilde{x}^* - x^*\| \leq 5E^{-10}$. These behaviours agree with Theorem 1, even if the problem we actually solve slightly differs from the true problem.

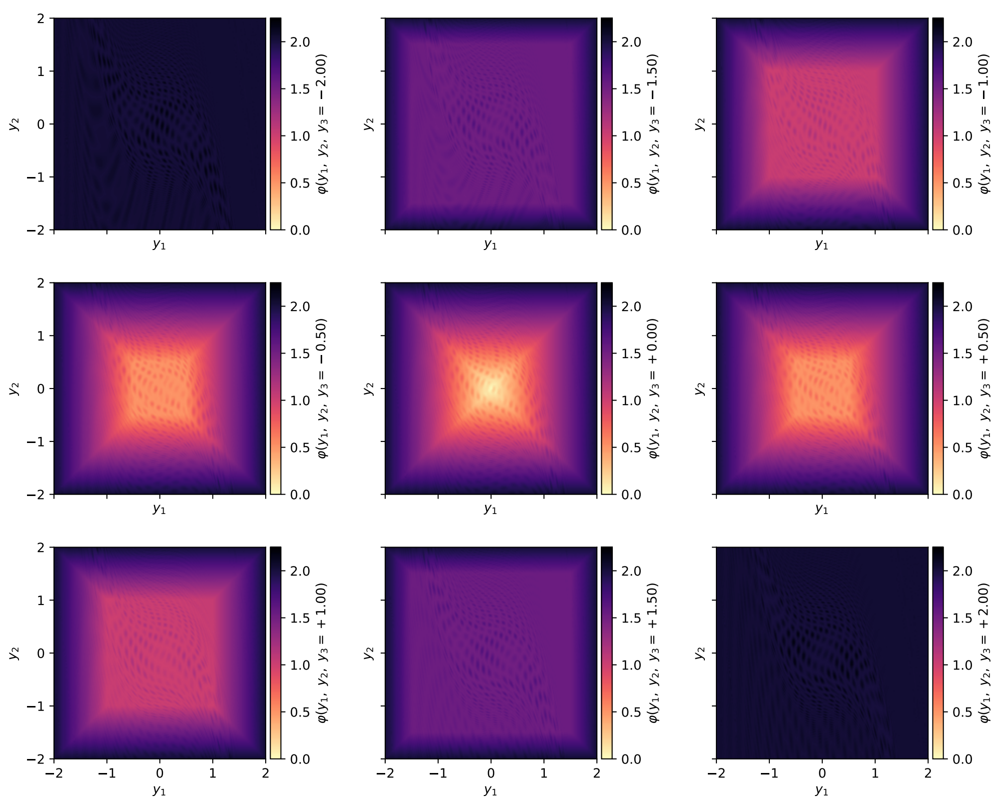


**Figure 8: function $\varphi$ in Section 3.4, restriction over planes with constant third variable.**

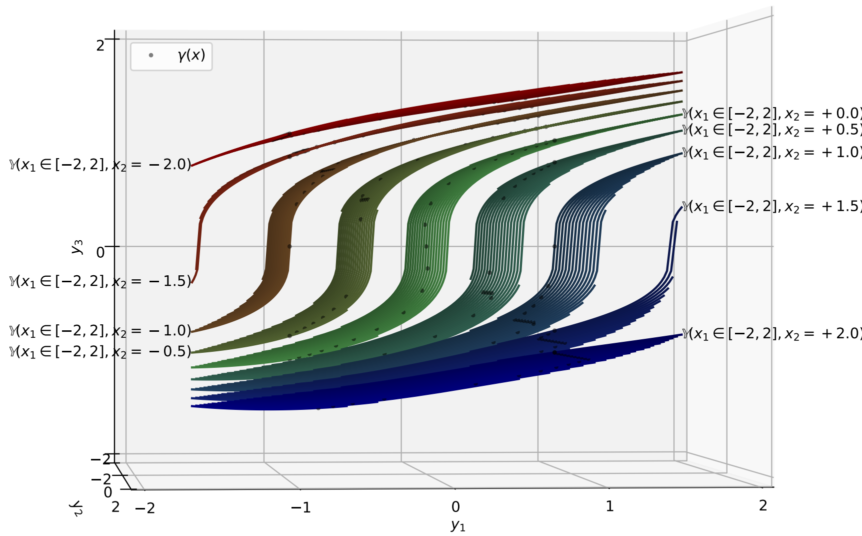


**Figure 9: Partition sets of $\mathbb{Y}(x)$ and locus of $\gamma$ in the context of Section 3.4.**

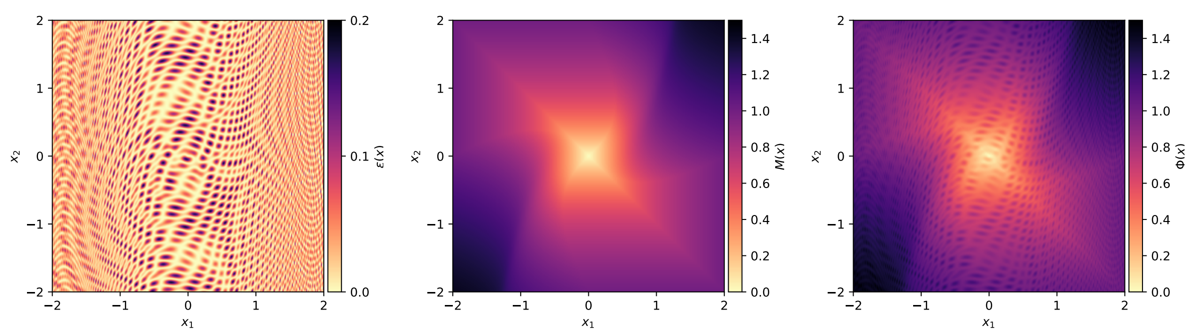


**Figure 10: functions (left) $\varepsilon$, (center) $M$ and (right) $\Phi$ in Section 3.4. In this section $\Phi = \varepsilon + M$.**

**Table 4: results for eight instances of Algorithm 1 in the context of Section 3.4. In this table, $\hat{x}^k \triangleq \|x^k\|_\infty$.**

| $x^0$ | $1^{st}\ \hat{x}^k \leq 5E^{-02}$ | $1^{st}\ \hat{x}^k \leq 5E^{-05}$ | $1^{st}\ \hat{x}^k \leq 5E^{-08}$ | returned $\hat{x}^k$ |
|---|---|---|---|---|
| $[-2, 2]$ | $\hat{x}^{032} = 3.52E^{-02}$ | $\hat{x}^{110} = 4.03E^{-05}$ | $\hat{x}^{194} = 4.79E^{-08}$ | $\hat{x}^{220} = 1.10E^{-08}$ |
| $[\frac{-1}{100}, e^2]$ | $\hat{x}^{037} = 2.02E^{-02}$ | $\hat{x}^{100} = 3.99E^{-05}$ | $\hat{x}^{176} = 1.10E^{-08}$ | $\hat{x}^{198} = 9.03E^{-10}$ |
| $[\frac{-\pi}{2}, \frac{7}{4}]$ | $\hat{x}^{027} = 6.34E^{-03}$ | $\hat{x}^{095} = 4.60E^{-05}$ | $\hat{x}^{177} = 1.23E^{-08}$ | $\hat{x}^{195} = 5.51E^{-09}$ |
| $[\frac{-\pi}{4}, e^{\frac{1}{2}}]$ | $\hat{x}^{040} = 3.66E^{-02}$ | $\hat{x}^{126} = 1.46E^{-05}$ | $\hat{x}^{235} = 3.45E^{-08}$ | $\hat{x}^{253} = 4.54E^{-09}$ |
| $[\frac{1}{4}, \frac{1}{4}]$ | $\hat{x}^{022} = 2.85E^{-02}$ | $\hat{x}^{119} = 2.90E^{-05}$ | $\hat{x}^{193} = 1.84E^{-08}$ | $\hat{x}^{212} = 2.88E^{-09}$ |
| $[\frac{3\pi}{2}, \frac{1}{\sqrt{8}}]$ | $\hat{x}^{039} = 3.29E^{-02}$ | $\hat{x}^{124} = 1.96E^{-05}$ | $\hat{x}^{186} = 4.86E^{-08}$ | $\hat{x}^{226} = 5.22E^{-09}$ |
| $[e^2, 2\pi]$ | $\hat{x}^{036} = 3.43E^{-02}$ | $\hat{x}^{109} = 3.00E^{-05}$ | $\hat{x}^{171} = 4.40E^{-08}$ | $\hat{x}^{211} = 8.19E^{-09}$ |
| $[e^2, \frac{-1}{11}]$ | $\hat{x}^{029} = 1.48E^{-02}$ | $\hat{x}^{093} = 4.77E^{-05}$ | $\hat{x}^{148} = 2.48E^{-08}$ | $\hat{x}^{172} = 1.48E^{-09}$ |

# 4 Numerical gain of performance provided by the POf

This section highlights that the POf, when applicable, may sensibly accelerate the solving process for large-scale DFO problems. We alter the four problems from Section 3 to rise their dimension around 100. For each problem, we compare the performance of the DFO solvers **NOMAD** [7] and **PRIMA** [23] solving Problem ($\mathbf{P}_{\texttt{ini}}$) to a naive implementation of the cDSM solving Reformulation ($\mathbf{P}_{\texttt{ref}}$).

In most of these examples, computing $\varphi(y)$, $\gamma(x)$ and $\Phi(x)$ is almost instantaneous, for all $y \in \mathbb{Y}$ and $x \in X$. In contrast, this would likely not be true in real cases, since either $\varphi(y)$ may be computed by an expensive computer program and $\gamma(x)$ may be obtained by solving an optimization subproblem. Thus, to establish a relevant comparison between the solvers, we proceed as follows. For each solver, we track the sequence of all evaluated points and their associated objective values. This consists in evaluations of $\varphi$ for **NOMAD** and **PRIMA**, and of $\Phi$ for the cDSM. Then, for each problem, we consider that, for all $y \in \mathbb{Y}$, the cost to compute $\varphi(y)$ is 1 unit (identical for all $y \in \mathbb{Y}$) and that, for all $x \in X$, the computation of $\Phi(x)$ costs $1+\tau$ units, where $\tau \geq 0$ is fixed. The value $\tau$ captures the relative cost to evaluate $\gamma$ versus whose to evaluate $\varphi$, so the computational cost of $\Phi(x)$ models the additional cost required to first compute $\gamma(x)$ before computing $\varphi(\gamma(x))$. Finally, we plot the graph representing $c \in \mathbb{R}_+$ versus the lowest objective value found by each solver within a budget of $c$ units.

Our graphs show that the POf sensibly eases the identification of a solution to Problem ($\mathbf{P}_{\texttt{ini}}$), and that, if $\tau$ is not excessive (that is, the computation time of $\gamma$ is not excessive compared to whose of $\varphi$), then the overall computation time is reduced. A naive instance of our cDSM (solving Problem ($\mathbf{P}_{\texttt{ini}}$) via Reformulation ($\mathbf{P}_{\texttt{ref}}$) and Theorem 2) performs better than competitive solvers (solving Problem ($\mathbf{P}_{\texttt{ini}}$) directly). Indeed, our cDSM instance works, while, even under a massive budget, both **NOMAD** and **PRIMA** fail to initiate a convergence towards a relevant solution. Let us stress that our goal is not to criticize **NOMAD** and **PRIMA** (it is expected that they do not perform well in this context). We only seek to prove that we outperform these two usual solvers thanks to, and only to, our reliance on the POf to sensibly reduce the dimension of the problem to address.

We use **NOMAD** with its default parameters, **PRIMA** with the BOBYQA algorithm (as it handles the box constraints), and our cDSM with the parameters provided in Remark 1. When additional constraints exist in the problem, we follow the *extreme barrier* strategy [4] that redefines $\varphi(y) \triangleq +\infty$ for all $y \notin \Omega$. For each problem, we give to **NOMAD** and **PRIMA** six starting points $(y^0_\ell)_{\ell\in[\![1,6]\!]}$ chosen randomly in $Y$, and we give to our cDSM the associated starting points $(\chi(y^0_\ell))_{\ell\in[\![1,6]\!]}$. In this section, we denote by $\mathbb{1} \triangleq (1,\dots,1)$, when the dimension of that vector is clear from the context.

## 4.1 101-dimensional problem with mono-variable noise

Consider the function

$$\varphi : \begin{cases} \mathbb{Y} = Y \triangleq \mathbb{R}^{101} & \to \quad \mathbb{R} \\ y = (y_i)_{i=0}^{100} & \mapsto \quad \sum\limits_{i=1}^{100} (y_i - \sigma_i(y_0))^2 + \varepsilon(y_0), \end{cases}$$

where $\varepsilon(z) \triangleq |z|\,(1+\sin(\frac{2\pi}{z})^2)^{\frac{1}{2}} + |\lfloor z \rceil|$ if $z \in \mathbb{R}^*$ and $\varepsilon(0) \triangleq 0$, and where

$$\sigma : \begin{cases} \mathbb{R} & \to \quad \mathbb{R}^{101} \\ z & \mapsto \quad \left[\, z\ ,\ \left(2\left(1+\frac{i-1}{5}\right)\lfloor \frac{z}{i} \rceil\right)_{i=1}^{25}\ ,\ \left(25\cos\left(\frac{i-25}{5}\pi z\right)\right)_{i=26}^{50}\ ,\ \left(z-\frac{10}{i}\right)_{i=51}^{75}\ ,\ \left(\frac{i}{10}\right)_{i=76}^{100}\ \right]. \end{cases}$$

The global minimizer of $\varphi$ is $y^* \triangleq \sigma(0)$, with $\varphi(y^*) = 0$. We define $\mathbb{Y}(x) \triangleq \{x\}\times\mathbb{R}^{100}$ for all $x \in \mathbb{X} \triangleq \mathbb{R}$, since $\varphi_{|\mathbb{Y}(x)}$ is therefore a quadratic function of $(y_i)_{i=1}^{100}$ while $y_0 = x$ is fixed. Its minimizer is

$$\gamma(x) \triangleq \sigma(x), \quad \text{with} \quad \varphi(\gamma(x)) = \varepsilon(y_0).$$

Thus, $\Phi = \varepsilon$ and Assumptions 1 and 2 hold. Figure 11 compares the different strategies.

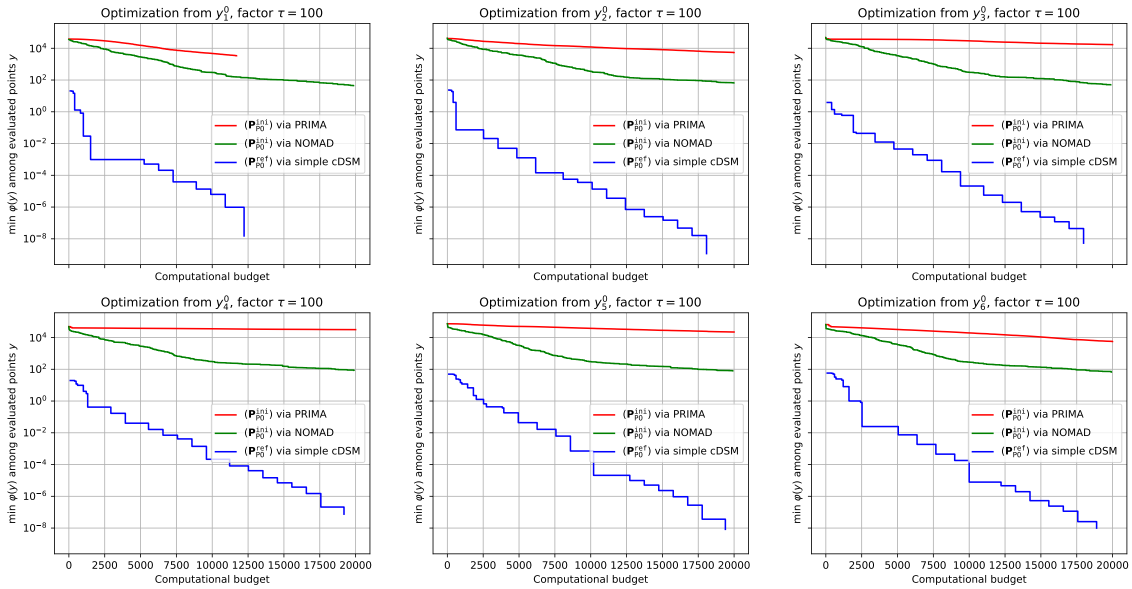


**Figure 11: Comparison of the best solution found by each solver depending on the computational cost spent. The six starting points $y_i^0$, $i \in [\![1, 6]\!]$, are chosen as 6 observations of the random uniform independent distribution over $Y$.**

Reformulation ($\mathbf{P_{ref}}$) is a 1-dimensional `DFO` problem, and the `POf` highlights how to handle the 100 others in the original problem. Even by considering $\tau = 100$, the `POf` provides a sensible gain of performance: our naive `cDSM` relying on the `POf` outperforms the solvers solving the original problem. Its returned solution has a significantly lower objective value; and the computational budget is comparable if the pointwise cost to evaluate $\gamma$ is at most 100 times greater than whose of $\varphi$. Our `cDSM` converges towards the global minimizer (the returned objective value is around $10^{-8}$ is all cases), while **NOMAD** and **PRIMA** solving the original problem remain both in an exploratory phase that returns poor results. The early stop of **PRIMA** in the first case even shows that **PRIMA** interrupted itself on a poor incumbent solution with half of its budget remaining.

## 4.2 Radial noise in a 101-dimensional problem

Consider the function expressed in polar coordinates

$$\varphi : \begin{cases} \mathbb{Y} = Y \triangleq \mathbb{R}_+ \times [0, 2\pi[^{100} & \to \quad \mathbb{R} \\ y = (r, \theta) & \mapsto \quad \dfrac{\sqrt{r}}{100} \displaystyle\sum_{i=1}^{100} \sin\left(\dfrac{\theta_i - 2\pi \log_{i+1}(r)}{2}\right)^2 + \varepsilon(r) \text{ if } r > 0, \text{ else } \varepsilon(0), \end{cases}$$

where $\varepsilon(z) \triangleq \sqrt{|z^2 - 2|} + \frac{\sin(10\pi(z - \sqrt{2}))^2}{10}$ for all $z \in \mathbb{R}_+$. The global minimizer of $\varphi$ is $y^* \triangleq (r^*, \theta^*)$ where $r^* \triangleq \sqrt{2}$ and $\theta^* \triangleq (2\pi \log_{i+1}(\sqrt{2}))_{i=1}^{100}$, with $\varphi(y^*) = 0$. We define $\mathbb{Y}(x) \triangleq \{x\} \times [0, 2\pi[^{100}$ for all $x \in \mathbb{X} = X \triangleq \mathbb{R}_+$. Then, $\varphi_{|\mathbb{Y}(x)}$ is a smooth function of $\theta$ while $r = x$ is fixed. It follows that

$$\gamma(x) \triangleq (x, (2\pi \log_{i+1}(x))_{i=1}^{100} \sim 2\pi), \quad \text{with} \quad \varphi(\gamma(x)) = \varepsilon(x),$$

where $z \sim 2\pi$ denotes the component-wise residual of $z \in \mathbb{R}^{100}$ modulo $2\pi$. Thus, $\Phi = \varepsilon$, and Assumptions 1 and 2 hold. The comparison between the different strategies is provided in Figure 12.

The results are similar to those in Section 4.1. Reformulation ($\mathbf{P_{ref}}$) is a 1-dimensional `DFO` problem and the `POf` handles the 100 others in the original problem. Even by considering $\tau = 100$, the `POf` provides a gain of performance. Within a comparable budget, our `cDSM` converges towards the global minimizer, while **NOMAD** and **PRIMA** solving the original problem remain in an exploratory phase that returns irrelevant points. The very early stop of **NOMAD** in three cases even shows that **NOMAD** sometimes interrupts itself early on its poor incumbent solution.

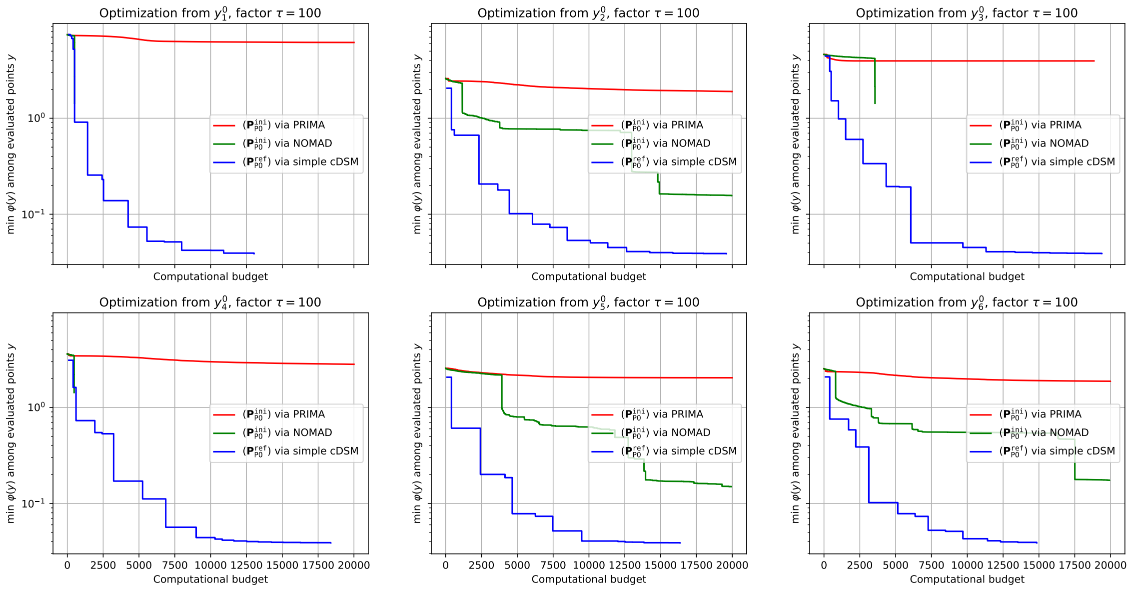


**Figure 12: Comparison of the best solution found by each solver depending on the computational cost spent. The six starting points $y_i^0$, $i \in [\![1,6]\!]$, are chosen as 6 observations of the random uniform independent distribution over $Y$.**

## 4.3 Noise affected by nonlinear combinations of all variables

Consider the function

$$\varphi : \left\{ \begin{array}{ccc} \mathbb{Y} \triangleq \mathbb{R}^{100} & \to & \mathbb{R} \\ y = (y_i)_{i=1}^{100} & \mapsto & \displaystyle\sum_{(\ell,p)\in[\![0,19]\!]^2} \ln\left(1 + \left(\frac{\pi_\ell(y)}{\pi_p(y)} - 1\right)^2\right) + \varepsilon\left(\sum_{\ell=0}^{19} \pi_\ell(y)\right), \end{array} \right.$$

where $\varepsilon(z) \triangleq \exp(\frac{1}{z-4}) + \frac{\sqrt{|z-4|}}{5}$ if $z \neq 4$ and $\varepsilon(4) \triangleq +\infty$; and $\pi_\ell(y) \triangleq \prod_{i=5\ell+1}^{5(\ell+1)} y_i$ for each $\ell \in [\![0,19]\!]$; and $\Omega \triangleq \{y \in (\mathbb{R}_+^*)^{100} : \forall i \in [\![1,100]\!],\ y_{i+1} \geq y_i\}$. Here $\varphi$ has no global minimizer, but $\inf \varphi(\Omega) = 0$ and $\varphi(y) \to 0$ when $y$ approaches $y^* \triangleq \mathbb{1}$ from some directions, but some others directions raise $\varphi(y) \to +\infty = \varphi(y^*)$. For all $x \in \mathbb{X} = X \triangleq \mathbb{R}$, we define $\mathbb{Y}(x) \triangleq \{y \in \mathbb{Y} : \Sigma_{\ell=0}^{19}\pi_\ell(y) = x\}$. Then, $\varphi_{|\mathbb{Y}(x)}$ is a smooth function since $\varepsilon(\Sigma_{\ell=0}^{19}\pi_\ell(y))$ is constant. Moreover its minimizer is analytically obtainable as follows. For all $x \in \mathbb{X}$, $\varphi_{|\mathbb{Y}(x)} \geq 0$ and equals 0 when all $(\pi_\ell)_{\ell\in[\![0,19]\!]}$ are equal. Under the constraint $\gamma(x) \in \Omega$, equalling all $(\pi_\ell)_{\ell\in[\![0,19]\!]}$ implies that $\gamma(x) = \alpha\mathbb{1}$ for some $\alpha \in \mathbb{R}$. Then, $\gamma(x) \in \mathbb{Y}(x)$ leads to $20\alpha^5 = x$. The value of $\alpha$ follows and we get that

$$\gamma(x) \triangleq \sqrt[5]{\frac{x}{20}}\mathbb{1}, \quad \text{with} \quad \varphi(\gamma(x)) = \varepsilon(x).$$

Thus, $\Phi = \varepsilon$, and Assumptions 1 and 2 hold. Figure 13 compares our strategies.

Similarly to Sections 4.1 and 4.2, Reformulation $(\mathbf{P}_{\texttt{ref}})$ is a `DFO` problem fixing the value of a single combination of all the variables, and the `POf` fixes the 100 variables of the original problem accordingly. Even with $\tau = 100$, the `POf` provides a gain of performance. **NOMAD** and **PRIMA** solving the original problem both fail to significantly improve their initial incumbent solution and, in the second case, both solvers behave as if their initial incumbent solution cannot be improved. Within a comparable budget, our `cDSM` converges towards the global minimizer on all but one case. Yet, in this case, first we observe that this solution remains 1000 times better than those returned by **NOMAD** and **PRIMA**, and second we claim that the poor performance results from our naive implementation. Even if Reformulation $(\mathbf{P}_{\texttt{ref}})$ is 1-dimensional, it is preferable to use a globalization strategy because the interval of possibly relevant values is large. Our instance lacks one since it lacks a `search` step.

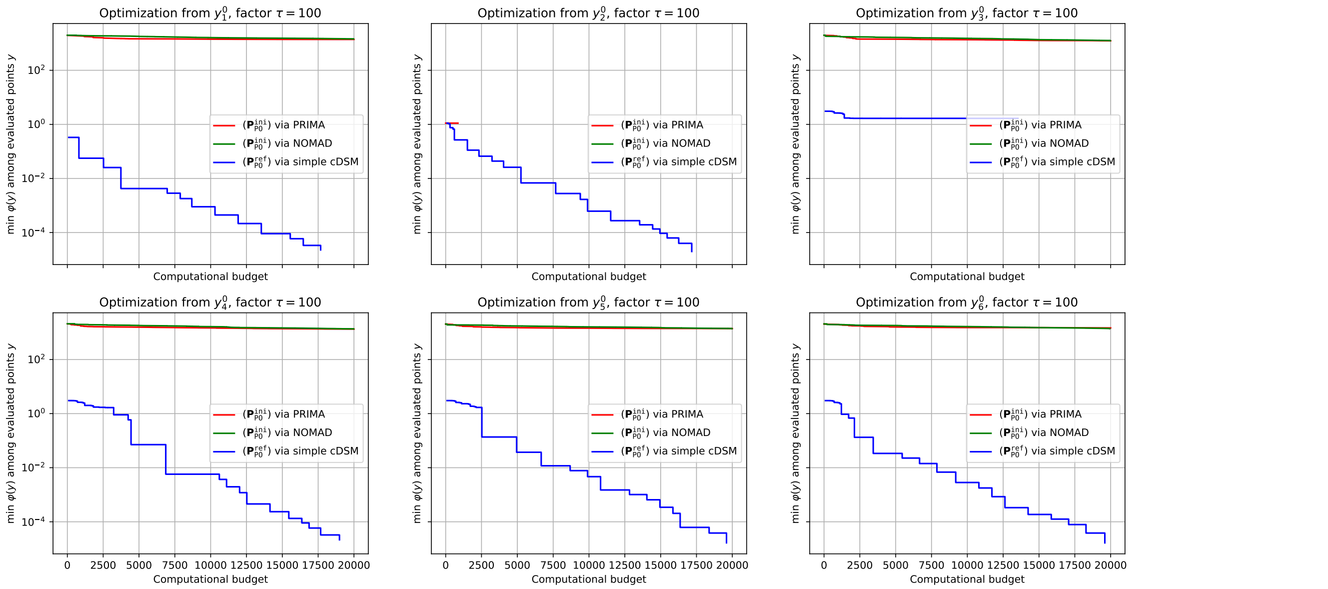


**Figure 13: Comparison of the best solution found by each solver depending on the computational cost spent. The starting points are chosen as $y_1^0 \triangleq (i/100)_{i=1}^{100}$, $y_2^0 \triangleq 0.5\mathbb{1}$, $y_3^0 \triangleq (\frac{2i}{100})_{i=1}^{100}$, and $y_i^0$, for each $i \in [\![4,6]\!]$, is chosen randomly but such that for each $j \in [\![1,100]\!]$, the $j^{th}$ component of $y_i^0$ lies in $[\frac{j-1}{100}, \frac{j}{100}]$.**

## 4.4 Ten-dimensional noise with non-analytical oracle and 100 variables

Consider the function

$$\varphi : \left\{ \begin{array}{lll} \mathbb{Y} = Y \triangleq \mathbb{R}^{100} & \to & \mathbb{R} \\ y = (y_i)_{i=1}^{100} & \mapsto & \|y\|_1 + \varepsilon\left(g_1(y_{10}) - \sum\limits_{i=1}^{9} y_i, \; \dots, \; g_{10}(y_{100}) - \sum\limits_{i=91}^{99} y_i\right), \end{array} \right.$$

where $g_j(z) \triangleq (z + (1+\frac{j}{10})^z - 1)$ for all $z \in \mathbb{R}$ and all $j \in \mathbb{N}$, and

$$\varepsilon : \left\{ \begin{array}{lll} \mathbb{X} = X \triangleq \mathbb{R}^{10} & \to & \mathbb{R} \\ z = (z_j)_{j=1}^{10} & \mapsto & \left(\frac{\sin 5\pi\left(z_2 - z_1^3\right)}{5} + \frac{\sin 6\pi\left(z_4 - e^{-z_2 - z_3} + 1\right)}{7} + \frac{\sin 7\pi\sqrt{z_5^2 + z_6^2 + z_7^2}}{11} + \frac{\sin 8\pi z_8 z_9 z_{10}}{13}\right)^2. \end{array} \right.$$

The global minimizer of $\varphi$ is $y^* \triangleq 0\mathbb{1}$, with $\varphi(y^*) = 0$. For all $x \in \mathbb{X}$, we define $\mathbb{Y}(x)$ such that the vector of the 10 inputs of $\varepsilon$ is fixed to $x$. That is, $\mathbb{Y}(x) \triangleq \{y \in \mathbb{Y} : g_j(y_{10j}) - \Sigma_{i=10(j-1)+1}^{10(j-1)+9} y_i = x_j, \; \forall j \in [\![1,10]\!]\}$. Then $\varepsilon_{|\mathbb{Y}(x)}$ is constant and the minimizer of $\varphi_{|\mathbb{Y}(x)}$ is the solution to the problem

$$\underset{y \in \mathbb{Y}(x)}{\text{minimize}} \quad \varepsilon(x) + \sum_{j=1}^{10} \left( \sum_{i=10(j-1)+1}^{10(j-1)+9} |y_i| + \left| g_j^{-1}\left( x_j + \sum_{i=10(j-1)+1}^{10(j-1)+9} y_i \right) \right| \right),$$

where the variables $(y_{10j})_{j=1}^{10}$ are fixed via the constraints inducing $\mathbb{Y}(x)$. This solution is[1] the vector

$$\gamma(x) = (\gamma_i(x))_{i=1}^{100} \quad \text{where} \quad \gamma_i(x) \triangleq \left\{ \begin{array}{ll} 0 & \text{if } i \notin 10\mathbb{N}, \\ g_j^{-1}(x_j) & \text{if } i = 10j, \; j \in [\![1,10]\!], \end{array} \right.$$

with

$$\varphi(\gamma(x)) = \varepsilon(x) + \sum_{j=1}^{10} \left| g_j^{-1}(x_j) \right|.$$

[1] We have $\frac{d}{dz} g_j(z) = (1 + (1+\frac{j}{10})^z \ln(1+\frac{j}{10})) > 1$ for all $(j,z) \in \mathbb{N}^* \times \mathbb{R}$, so $\frac{d}{dz} g_j^{-1}(z) = \frac{1}{\frac{d}{dz} g_j\left(g_j^{-1}(z)\right)} \in ]0,1[$. The claim follows by a sensitivity analysis of the objective function with respect to each sum $\sum_{i=10(j-1)+1}^{10(j-1)+9} |y_i|$.

As a result, $\Phi$ equals

$$\Phi : \left\{ \begin{array}{ccc} \mathbb{X} = X \triangleq \mathbb{R}^{10} & \to & \mathbb{R} \\ x = (x_j)_{j=1}^{10} & \mapsto & \varepsilon(x) + \sum_{j=1}^{10} \left| g_j^{-1}(x_j) \right| . \end{array} \right.$$

Its global minimizer is $x^* \triangleq 0\mathbb{1}$, with $\Phi(x^*) = 0$. Assumptions 1 and 2 hold.

Unfortunately, $\gamma$ admit no analytical expression since, for all $j \in [\![1, 10]\!]$, $g_j^{-1}$ cannot be expressed with elementary functions. However, since $g_j^{-1}(x_j)$ solves the equation $(z + (1 + \frac{j}{10})^z - 1) - x_j = 0$ (with variable $z \in \mathbb{R}$), we approximately solve this equation using a dichotomic search to obtain an approximate solution $\hat{g_j}^{-1}(x_j) \approx g_j^{-1}(x_j)$. As a result, for all $x \in X$, we approximate

$$\gamma(x) = (\gamma_i(x))_{i=1}^{100} \approx \hat{\gamma}(x) = (\hat{\gamma}_i(x))_{i=1}^{100} \quad \text{where} \quad \hat{\gamma}_i(x) \triangleq \left\{ \begin{array}{ll} 0 & \text{if } i \notin 10\mathbb{N}, \\ \hat{g_j}^{-1}(x_j) & \text{if } i = 10j, \ j \in [\![1, 10]\!], \end{array} \right.$$

and

$$\Phi \approx \hat{\Phi} \quad \text{where} \quad \hat{\Phi} : \left\{ \begin{array}{ccc} \mathbb{X} = X \triangleq \mathbb{R}^{10} & \to & \mathbb{R} \\ x = (x_j)_{j=1}^{10} & \mapsto & \varepsilon(x) + \sum_{j=1}^{10} \left| \hat{g_j}^{-1}(x_j) \right| . \end{array} \right.$$

Figure 14 shows a comparison between the different strategies, where for tractability our naive `cDSM` actually minimizes $\hat{\Phi}$ instead of $\Phi$. Even if the problem we actually solve slightly differs from the true problem, we observe results similar to those in Sections 4.1, 4.2 and 4.3. Reformulation ($\mathbf{P}_{\texttt{ref}}$) is a 10-dimensional `DFO` problem fixing the value of the important combinations of all the variables, and the `POf` highlights how to fix the 100 variables of the original problem accordingly. Using our `cDSM`, the `POf` provides a gain of performance if we allow $\tau \approx 10$ at most. With $\tau = 10$, the `POf` allows our `cDSM` to converge under a computational budget similar to whose required by **NOMAD** and **PRIMA**, but the solutions returned by our `cDSM` are $10^2$ times better than those returned by **NOMAD** and $10^7$ times better than those returned by **PRIMA**. The value $\tau = 10$ is quite low, but we could consider a higher value by using a better implementation of our algorithm. For example, a test with **NOMAD** solving the reformulated problem (instead of our naive instance of `cDSM`) provided similar graphs with $\tau = 30$. This last run is represented on Figure 15.

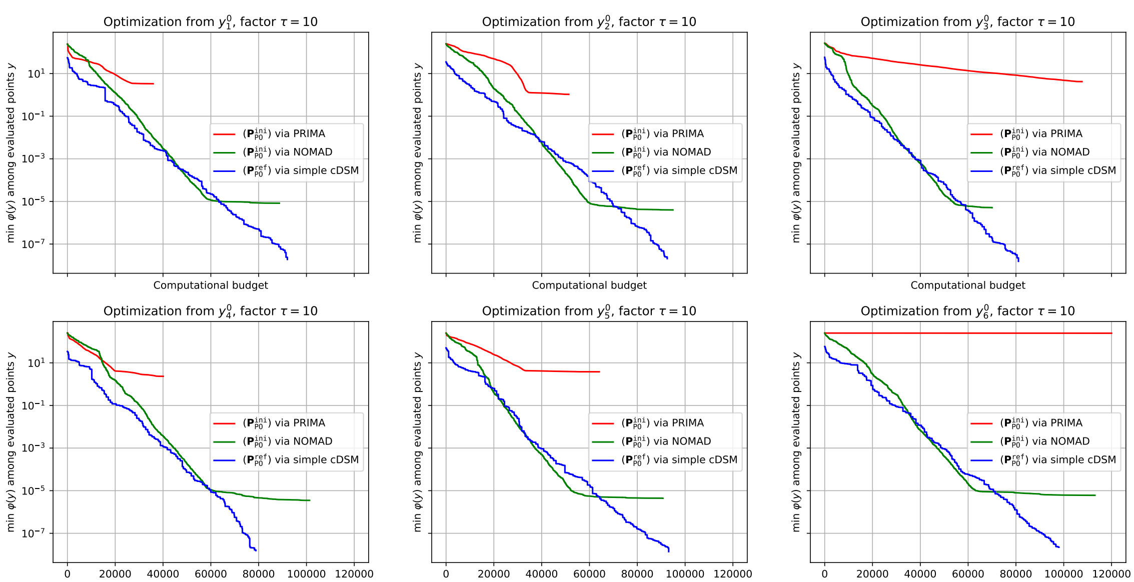


**Figure 14: Comparison of the best solution found by each solver depending on the computational cost spent. The six starting points $y_i^0$, $i \in [\![1, 6]\!]$, are chosen as 6 observations of the random uniform independent distribution over $Y$.**

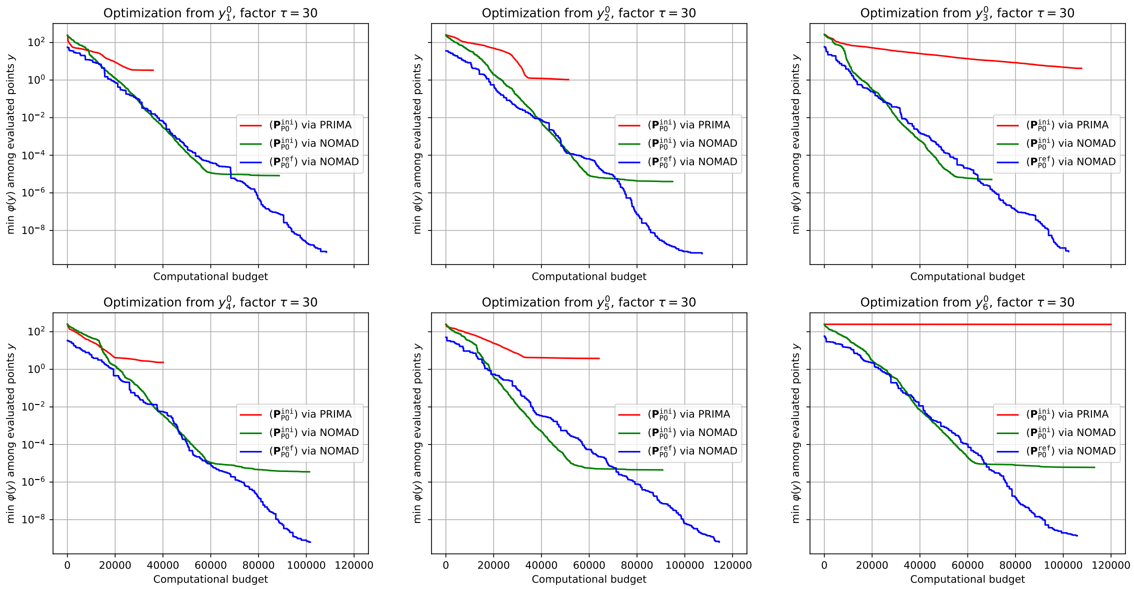


**Figure 15: Additional gain provided by solving the reformulated problem using a more efficient solver. Despite the higher value for $\tau$ compared to the results on Figure 14, the graphs are similar.**

# 5 General discussion

To conclude this work, Section 5.1 summarizes the most important aspects of the `POf` and Section 5.2 lists some routes for improvements.

## 5.1 Comments on the `POf`

As already highlighted, a stringent aspect of the `POf` is the oracle function $\gamma$, resulting from the chosen partition sets of $\mathbb{Y}$. For all $x \in X$, Subproblem ($\mathbf{P}_{\mathtt{sub}}(x)$) must have an explicitly accessible global solution. That is, on each partition set the objective function must have a tractable global minimizer. Ensuring this requirement is challenging in general and likely requires a problem-dependent strategy. Nevertheless, Sections 3.4 and 4.4 show that this theoretical requirement may be slightly relaxed in practice. The `POf` may be used when $\gamma$ is defined pointwise as, for all $x \in X$, the output of a numerical method solving Subproblem ($\mathbf{P}_{\mathtt{sub}}(x)$) with great confidence to approximate closely a global solution.

The usual case of application of the `POf` is when $\mathbb{Y}$ has a dimension $d_{\mathbb{Y}} \geq 2$ and $\mathbb{X} = \mathbb{R}^{d_{\mathbb{X}}}$ has a finite dimension $0 < d_{\mathbb{X}} \ll d_{\mathbb{Y}}$. In practice, it is preferable to ensure that $d_{\mathbb{X}} \leq 50$, since this usual threshold in `DFO` [6, Section 1.4] is related to the performance of most solvers. Although our numerical examples in Section 4 illustrate cases with $d_{\mathbb{Y}} = 101$ at most, we stress that the performance of the `POf` depends only on $d_{\mathbb{X}}$ and on the computation time related to $\gamma$. Hence, $d_{\mathbb{Y}}$ may be possibly infinite, as in [9, Chapter 7] where $\mathbb{Y}$ is a functional space. Moreover, the smaller $d_{\mathbb{X}}$ is and the more efficient the `DFO` solver solving Reformulation ($\mathbf{P}_{\mathtt{ref}}$) is, the larger the computation time related to $\gamma$ can be. As a rule of thumb, if one can estimate the number $\mathbf{N}^{\mathtt{ref}}$ of points in $\mathbb{X}$ required by the `DFO` solver to solve Reformulation ($\mathbf{P}_{\mathtt{ref}}$), as well as the number $\mathbf{N}^{\mathtt{ini}}$ of points in $\mathbb{Y}$ required by a dedicated solver to solve Problem ($\mathbf{P}_{\mathtt{ini}}$) directly, then the `POf` would be deemed useful when $\mathbf{N}^{\mathtt{ref}}\tau < \mathbf{N}^{\mathtt{ini}}$.

The choice of the partition to apply the `POf` is usually left to the user. There are many ways to partition $\mathbb{Y}$ to ensure that the `POf` is applicable and Assumptions 1 and 2 hold. Assumption 1 is presumably easy to satisfy since $\chi$ follows directly from the user-defined partition. Nevertheless, Assumption 2 is nontrivial. Indeed it involves $\gamma$, and thus it implicitly requires that $\varphi$ and the chosen partition of $\mathbb{Y}$ satisfy some additional properties. We are not aware of any easy and exhaustive rule about $\varphi$ and the partition of $\mathbb{Y}$ that ensures that Assumption 2 holds.

Theorem 1 relies on a sufficient but not necessary condition to claim that $\gamma(x^*)$ is a local solution to Problem ($\mathbf{P}_{\texttt{ini}}$). That is, $\gamma(x^*)$ may be a local solution even when the condition about the lower semicontinuity of $\varphi$ at all points of $\mathcal{ACC}(\gamma; x^*)$ is not met, as shown in Section 3.1. However, Theorem 1 always ensures that $\gamma(x^*)$ is a local solution to Problem ($\mathbf{P}_{\texttt{ini}}$) if $\varphi$ is continuous, since $\varphi$ is therefore lower semicontinuous at all points of $\mathcal{ACC}(\gamma; x^*)$. A necessary and sufficient condition for the local optimality of $\gamma(x^*)$ when $\varphi$ is discontinuous is the lower semicontinuity of $\varphi \circ \gamma = \Phi$ at $x^*$.

Theorem 1 is stated for the `cDSM` only, but it is actually compatible with many others `DFO` algorithms. We refer to [3, Algorithm 2] for the minimal framework for `DFO` algorithm to ensure Theorem 3. Any `DFO` algorithm satisfying Theorem 3 is compatible with Theorem 1.

The `POf` sometimes leads to an instance of Reformulation ($\mathbf{P}_{\texttt{ref}}$) that may be solved analytically. A nontrivial example from optimal control theory is given in [9, Chapter 7.3.1]. In such cases, we recover a solution to Problem ($\mathbf{P}_{\texttt{ini}}$) via Theorem 2 directly, so Assumptions 1 and 2 are not required.

## 5.2 Perspectives for future work

A future work will relax the requirement for an accessible global solution for Subproblem ($\mathbf{P}_{\texttt{sub}}(x)$) for all $x \in X$. We plan to alter the `POf` to allow for the access to only an approximation of a local solution is accessible. This would be representative of most practical cases, where all subproblems are solved numerically by an optimization method which only approximates a local solution. We conducted some tests following this idea on some problems from [9, Chapter 7] by defining, for all $x \in X$, $\gamma(x)$ as the output of a numerical method solving Subproblem ($\mathbf{P}_{\texttt{sub}}(x)$). This approach seems to work well, despite its possibly heuristic nature since the theory is not analyzed yet.

We will also study how to use a `DFO` algorithm differing from the `cDSM`, since a faster algorithm ensuring only some necessary optimality conditions may be preferable in some cases. This raises the question to determine if, given a point $x^*$ satisfying necessary optimality conditions for Reformulation ($\mathbf{P}_{\texttt{ref}}$), the point $\gamma(x^*)$ satisfies necessary optimality conditions for Problem ($\mathbf{P}_{\texttt{ini}}$).

Given a collection of partition sets of $\mathbb{Y}$, there is a flexibility in the way to index them. This choice of indexation impacts the shape of $\Phi$. Then, we may fine-tune the indexation of the partition sets of any partition of $\mathbb{Y}$ so that the resulting Reformulation ($\mathbf{P}_{\texttt{ref}}$) is solved as easily as possible.

Reformulation ($\mathbf{P}_{\texttt{ref}}$) is currently defined according to an *extreme barrier* [4], since $\Phi(x) \triangleq +\infty$ for all $x \in \mathbb{X}$ such that Subproblem ($\mathbf{P}_{\texttt{sub}}(x)$) is infeasible. We observed in [9, Chapter 7] that an extreme barrier may lead to an important number of evaluated points returning the value $+\infty$. A *progressive barrier* [5] may be more efficient. We have initiated some research to associate an *infeasibility metric* to all points $x \in \mathbb{X}$ and use it in the progressive barrier. For all $x \in \mathbb{X}$, our infeasibility associated to $x$ with respect to Reformulation ($\mathbf{P}_{\texttt{ref}}$) is the infimum over all $y \in \mathbb{Y}(x)$ of the infeasibility of $y$ with respect to Subproblem ($\mathbf{P}_{\texttt{sub}}(x)$). In all the problems we tested in [9, Chapter 7], solving Reformulation ($\mathbf{P}_{\texttt{ref}}$) is significantly easier with such a progressive barrier than than with an extreme barrier.

This framework may be seen as a prototypical *hybrid method* using either a `DFO` algorithm (to solve Reformulation ($\mathbf{P}_{\texttt{ref}}$)) and other classes of algorithms (to solve Subproblem ($\mathbf{P}_{\texttt{sub}}(x)$) for any $x \in X$). We plan to develop more advanced hybrid methods. Roughly speaking, we may solve Problem ($\mathbf{P}_{\texttt{ini}}$) directly by optimizing jointly the partition set to consider and the point to consider into the set.

Finally, the class of problems illustrated in Section 3 (that is, noisy problems where the noise depends on explicit combinations of the variables) is not the only class where the `POf` might be considered. A noticeable example lies in partially separable problems, where fixing some key variables makes the subproblem fully separable. Another example consists in problems where each variable has little individual influence on the objective value but their joint influence is much greater, similarly to the state variables in optimal control theory. As the theory is now stated, applying the `POf` requires only to follow the content of Section 1.2. We plan to reword our former work [2] accordingly.

# References

[1] R. Astudillo and P.I. Frazier. Thinking inside the box: A tutorial on grey-box bayesian optimization. In 2021 Winter Simulation Conference (WSC), pages 1–15. IEEE, 2021.

[2] C. Audet, P.-Y. Bouchet, and L. Bourdin. A derivative-free approach to optimal control problems with a piecewise constant mayer cost function. Technical Report G–2021–63, Les Cahiers du GERAD, 2021.

[3] C. Audet, P.-Y. Bouchet, and L. Bourdin. Convergence towards a local minimum by direct search methods with a covering step. in second round of revision in optimization letters, ArXiV, 2024.

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

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

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

[7] C. Audet, S. Le Digabel, V. Rochon Montplaisir, and C. Tribes. Algorithm 1027: NOMAD version 4: Nonlinear optimization with the MADS algorithm. ACM Transactions on Mathematical Software, 48(3):35:1–35:22, 2022.

[8] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer. Nonlinear Parametric Optimization. Birkhäuser Basel, 1983.

[9] P.-Y. Bouchet. Théorie de l'optimisation sans dérivées dans le cas discontinu. PhD thesis, Polytechnique Montréal, 12 2023.

[10] N. Boumal. An introduction to optimization on smooth manifolds. Cambridge University Press, 2023.

[11] P.G. Constantine. Active Subspaces. SIAM, Philadelphia, PA, 2015.

[12] S. Dempe and A. Zemkoho. Bilevel Optimization. Number 978–3–030–52119–6 in Springer Optimization and Its Applications. Springer, 09 2020.

[13] Y. Diouane, V. Kungurtsev, F. Rinaldi, and D. Zeffiro. Inexact direct-search methods for bilevel optimization problems, 2023.

[14] K.J. Dzahini and S.M. Wild. Stochastic trust-region algorithm in random subspaces with convergence and expected complexity analyses, 2022.

[15] E.A. Feinberg, P.O. Kasyanov, and D.N. Kraemer. Continuity of parametric optima for possibly discontinuous functions and noncompact decision sets, 2021.

[16] J. Hu, X. Liu, Z. W. Wen, and Y. X. Yuan. A brief introduction to manifold optimization. Journal of the Operations Research Society of China, 8:199–248, 2020.

[17] D.R. Jones and J.R.R.A. Martins. The DIRECT algorithm: 25 years Later. Journal of Global Optimization, 79:521–566, 2021.

[18] G. Liuzzi, S. Lucidi, and V. Piccialli. A partition-based global optimization algorithm. Journal of Global Optimization, 48:113–128, 2010.

[19] B.S. Mordukhovich. Bilevel Optimization and Variational Analysis, pages 197–226. Springer International Publishing, Cham, 2020.

[20] S.M. Robinson. An implicit-function theorem for a class of nonsmooth functions. Mathematics of Operations Research, 16(2):292–309, 1991.

[21] G. Still. Lectures on parametric optimization: An introduction. Optimization Online, 2018.

[22] N. Wycoff, M. Binois, and S.M. Wild. Sequential learning of active subspaces. Journal of Computational and Graphical Statistics, 30(4):1224–1237, 2021.

[23] Z. Zhang. PRIMA: Reference Implementation for Powell's Methods with Modernization and Amelioration. available at http://www.libprima.net, DOI: 10.5281/zenodo.8052654, 2023.