

# A penalty-interior point method combined with MADS for equality and inequality constrained optimization

Charles Audet\*      Andrea Brilli<sup>†</sup>      Youssef Diouane\*  
Sébastien Le Digabel\*      Everton J. Silva<sup>‡</sup>      Christophe Tribes\*

January 29, 2026

## Abstract

This work introduces MADS-PIP, an efficient framework that integrates a penalty-interior point strategy into the mesh adaptive direct search (MADS) algorithm for solving nonsmooth blackbox optimization problems with general inequality and equality constraints. Inequality constraints are partitioned into two subsets: one treated via a logarithmic barrier applied to an aggregated interior constraint violation, and the other handled through an exterior quadratic penalty. All equality constraints are treated by the exterior penalty. A merit function defines a sequence of unconstrained subproblems, which are solved approximately using MADS, while a carefully designed update rule drives the penalty-barrier parameter to zero. In the nonsmooth setting, we establish convergence results ensuring feasibility for general constraints as well as Clarke stationarity for inequality-constrained problems. Computational experiments on both analytical test sets and challenging blackbox problems demonstrate that the proposed MADS-PIP algorithm is competitive with, and often outperforms, MADS with the progressive barrier strategy, particularly in the presence of equality constraints.

**Keywords:** Blackbox constrained optimization; Mesh adaptive direct search; Penalty-interior point method.

**AMS Classification:** 68Q25, 68R10, 68U05

---

\*GERAD and Department of Mathematics and Industrial Engineering, Polytechnique Montréal. 6079, Succ. Centre-ville Montréal, Québec H3C 3A7, Canada (charles.audet@gerad.ca, youssef.diouane@polymtl.ca, sebastien.le-digabel@polymtl.ca, christophe.tribes@polymtl.ca)

<sup>†</sup>“Sapienza” University of Rome, Department of Computer Control and Management Engineering “A. Ruberti”, Rome, Italy (brilli@diag.uniroma1.it). First and corresponding author.

<sup>‡</sup>NOVA University of Lisbon, Center for Mathematics and Applications (Nova Math), Campus da Caparica, 2829-516, Caparica, Portugal (ejo.silva@campus.fct.unl.pt).

ORCID: 0000-0002-3043-5393 (Audet), 0000-0002-8307-4106 (Brilli), 0000-0002-6609-7330 (Diouane), 0000-0003-3148-5090 (Le Digabel), 0000-0002-3627-3030 (Silva), 0000-0002-8740-6155 (Tribes).

# 1 Introduction

The presence of general constraints in optimization introduces challenges. For nonlinear or nonconvex constraints, classical optimization methods, such as gradient-based ones, may be inefficient [17]. The task is even harder when derivatives are unavailable. Algorithms that do not rely on derivative information are particularly relevant for problems where the objective and/or the constraint functions are defined by complex simulations, for which derivatives are not computable, unreliable, or simply nonexistent. Such nonsmooth problems are addressed by Derivative-Free Optimization (DFO) algorithms [10, 26]. This work targets constrained optimization problems of the form

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^n} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & g_\ell(\mathbf{x}) \leq 0, & \ell = 1, 2, \dots, m \\ & h_j(\mathbf{x}) = 0, & j = 1, 2, \dots, p \end{aligned} \tag{P_0}$$

where  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ ,  $g_\ell : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  for all  $\ell = 1, 2, \dots, m$ ,  $h_j : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  for all  $j = 1, 2, \dots, p$ , and  $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ .

## 1.1 Constraint handling in DFO

Over the years, several methods have been proposed to handle constraints, most of them employing penalty, filter, or barrier approaches [45].

In the context of direct search methods, early contributions to smooth constrained problems were proposed by Lewis and Torczon for bound-constrained [36] and linearly constrained [37] problems. They proposed an augmented Lagrangian approach [35] for general constraints, using a sequence of bound-constrained minimizations. Audet and Dennis [3] propose a filter approach [31] that accepts a trial point which improves either the objective function value or the constraint violation measure.

Mesh Adaptive Direct Search (MADS) [7] is a generalization of pattern search methods that handles nonsmooth constraints using an extreme barrier approach that only evaluates feasible points. To address the limitations of the extreme barrier, a progressive barrier approach was proposed in [4]. Similar to a filter approach, the progressive barrier method sets a threshold on the constraint violation accepted, progressively decreasing its value as the iterations progress. The approach was later extended to handle linear equality constraints [12], by reformulating the optimization problem, potentially reducing the number of original variables.

Merit-function techniques have been investigated within the derivative-free literature; for instance, [32] proposes a merit-based strategy for DFO methods, while [29] extends a class of evolution strategies to handle quantifiable relaxable constraints [34] through a merit function combined with a restoration procedure, and addresses unrelaxable constraints using either an extreme-barrier or projection approach.

Linesearch approaches are directional direct search methods. In this framework, a contribution to smooth bound-constrained problems was proposed by Lucidi and Sciandrone [40]. Slight modifications of the same approach have been recently proposed in [19], inspired by [21],

proving worst-case complexity bounds and a finite active-set identification property. In [38] the authors consider inequality constrained problems using a transformation into unconstrained or linearly constrained minimization of a smooth approximation of a nonsmooth exact penalty function. In [30] an exact penalty approach addressing nonsmooth inequalities is proposed. A sequential penalty approach was proposed in [39]. It solves the original problem through a sequence of approximate minimizations of a merit function, progressively increasing the penalization of constraint violations. In [22], using the same algorithmic framework, a merit function handles inequality constraints with a logarithmic barrier approach and equality constraints by including a penalization term. The latter approach has been recently integrated into direct search (LOG-DS) to address general nonlinear constrained optimization problems [20], assuming the continuous differentiability of the functions. The convergence to KKT-stationary points was established for LOG-DS under smooth assumptions.

## 1.2 Contributions and structure of the paper

DFO algorithms, particularly direct search methods, are well known for their ability to handle nonsmooth problems. This work extends the MADS algorithm to equality constrained optimization problems. The constraints are partitioned into two sets. One set is treated by a logarithmic barrier, and the other is handled by an exterior penalty. In the nonsmooth setting, we derive convergence guarantees that lead to feasibility under general constraints and to Clarke stationarity in the case of inequality constraints. Computational results obtained on both analytical benchmarks and realistic blackbox problems show that the proposed method is competitive with, and often outperforms, the MADS progressive barrier approach [4], especially in the presence of equality constraints.

Section 2 describes the proposed algorithm, with emphasis on the partition of inequality constraints. Section 3 develops the convergence analysis, addressing both feasibility for general constraints and stationarity for inequality constraints. Section 4 describes practical implementation aspects, and Section 5 reports computational experiments on analytical test problems and difficult blackbox problems. Section 6 concludes with final remarks.

# 2 A penalty-interior point method for MADS

This work proposes to solve  $(P_0)$  by means of a merit function approach that transforms the constrained problem into a sequence of unconstrained ones.

## 2.1 Partitioning the constraints

The approach partitions the inequality constraints indices into two sets, namely  $\mathcal{G}^{\text{int}}$  and  $\mathcal{G}^{\text{ext}}$ , so that  $\mathcal{G}^{\text{int}} \cap \mathcal{G}^{\text{ext}} = \emptyset$  and  $\mathcal{G}^{\text{int}} \cup \mathcal{G}^{\text{ext}} = \{1, 2, \dots, m\}$ . This partition depends on the starting point, and further details are provided in Section 4. The set of feasible points for  $(P_0)$  is the intersection of

$$\begin{aligned} \Omega^{\text{int}} &= \{\mathbf{x} \in \mathbb{R}^n \mid g_\ell(\mathbf{x}) \leq 0, \forall \ell \in \mathcal{G}^{\text{int}}\} \\ \text{and } \Omega^{\text{ext}} &= \{\mathbf{x} \in \mathbb{R}^n \mid g_\ell(\mathbf{x}) \leq 0, \forall \ell \in \mathcal{G}^{\text{ext}}; h_j(\mathbf{x}) = 0, \forall j = 1, 2, \dots, p\}. \end{aligned}$$

Therefore, a point  $\mathbf{x} \in \mathbb{R}^n$  is feasible for  $(P_0)$  if and only if  $\mathbf{x} \in \Omega \equiv \Omega^{\text{int}} \cap \Omega^{\text{ext}}$ .

We briefly recall the logarithmic-barrier/exterior-penalty strategy used in [20]. The approach analyzed in [22] employs a merit function that treats the inequality constraints via a logarithmic barrier and the equality constraints via an exterior penalty term. A well-known drawback of a pure barrier strategy is the requirement of a strictly feasible starting point; this is addressed in [20] by combining the logarithmic term with an exterior penalty that handles the constraints violated by the initial point. Another practical issue is that the use of the logarithmic term can potentially cause the merit function to be unbounded below. Since the method is based on minimizing such a function, this is a clear issue.

This paper exploits the nonsmooth settings to address this difficulty by imposing a threshold on the constraint values, namely by considering  $\min\{t, -g_\ell(\mathbf{x})\}$ , where  $t > 0$  is a fixed scalar. Although the choice of  $t$  lies outside the scope of this work, we note that setting  $t = 1$  ensures that the logarithm of  $\min\{t, -g_\ell(\mathbf{x})\}$  is nonnegative for any  $\mathbf{x} \in \mathbb{R}^n$  that strictly satisfies the constraint  $g_\ell(\mathbf{x}) < 0$ . The same effect of changing  $t$  is achieved by scaling the constraint functions. Therefore, without loss of generality, we fix  $t = 1$  for the remainder of the paper. With this choice, the constraints indexed by  $\mathcal{G}^{\text{int}}$  are aggregated into the *interior constraint violation* function

$$c^{\text{int}}(\mathbf{x}) = \begin{cases} -\prod_{\ell \in \mathcal{G}^{\text{int}}} \min\{1, -g_\ell(\mathbf{x})\} & \text{if } g_\ell(\mathbf{x}) \leq 0 \text{ for all } \ell \in \mathcal{G}^{\text{int}}, \\ \phi^{\text{prox}}(\mathbf{x}) & \text{otherwise,} \end{cases}$$

where  $\phi^{\text{prox}}(\mathbf{x}) = \max_{\ell \in \mathcal{G}^{\text{int}}} g_\ell(\mathbf{x})$ . The function  $\phi^{\text{prox}}(\cdot)$  represents a measure of proximity to the boundary of  $\Omega^{\text{int}}$ , assuming negative values for all points  $\mathbf{x}$  that strictly satisfy the inequality constraints in  $\mathcal{G}^{\text{int}}$ , and positive values if at least one of such constraints is violated. The interior constraint violation function satisfies  $c^{\text{int}}(\mathbf{x}) \in [-1, +\infty[$  for all  $\mathbf{x} \in \mathbb{R}^n$ , and  $c^{\text{int}}(\mathbf{x}) \leq 0$  if and only if  $\mathbf{x} \in \Omega^{\text{int}}$ . Therefore, under this transformation, the set  $\Omega^{\text{int}}$  can be equivalently defined as  $\Omega^{\text{int}} = \{\mathbf{x} \in \mathbb{R}^n \mid c^{\text{int}}(\mathbf{x}) \leq 0\}$ . Furthermore, if for all  $\ell \in \mathcal{G}^{\text{int}}$ , the functions  $g_\ell(\mathbf{x})$  are (Lipschitz) continuous near a point  $\mathbf{x} \in \mathbb{R}^n$ , then both  $\phi^{\text{prox}}(\cdot)$  and  $c^{\text{int}}(\cdot)$  are also (Lipschitz) continuous around  $\mathbf{x}$ .

The proposed method deals with the constraint  $c^{\text{int}}(\mathbf{x}) \leq 0$  using the transformation  $\log(-c^{\text{int}}(\mathbf{x}))$ , thus treating it with a strictly feasible approach. Conceptually, this strategy is similar to the formulations in [20, 22], where all inequality constraints assigned to the logarithmic barrier appear individually in the merit function through the sum  $\sum_\ell \log(-g_\ell(\mathbf{x}))$ . In fact, the properties of the logarithm ensure that, for all  $\mathbf{x} \in \Omega^{\text{int}}$

$$\log(-c^{\text{int}}(\mathbf{x})) = \log\left(\prod_{\ell \in \mathcal{G}^{\text{int}}} \min\{1, -g_\ell(\mathbf{x})\}\right) = \sum_{\ell \in \mathcal{G}^{\text{int}}} \log(\min\{1, -g_\ell(\mathbf{x})\}) \in [0, +\infty[,$$

which differs only in the threshold described in the previous paragraph.

The remaining constraints are addressed by penalizing their violation, using the *exterior constraint violation* function

$$c^{\text{ext}}(\mathbf{x}) = \sum_{\ell \in \mathcal{G}^{\text{ext}}} (\max\{0, g_\ell(\mathbf{x})\})^2 + \sum_{j=1}^p h_j(\mathbf{x})^2 \geq 0 \quad (1)$$

that satisfies  $c^{\text{ext}}(\mathbf{x}) = 0$  if and only if  $\mathbf{x} \in \Omega^{\text{ext}}$ . Given these two constraint violation functions, we introduce a merit function that solves  $(P_0)$  through a sequence of unconstrained problems.

**Definition 2.1.** *The merit function  $z : \mathbb{R}^n \times \mathbb{R}_+ \rightarrow \bar{\mathbb{R}}$  associated to problem  $(P_0)$  is defined as*

$$z(\mathbf{x}; \rho) = \begin{cases} f(\mathbf{x}) - \rho \log(-c^{\text{int}}(\mathbf{x})) + \frac{1}{\rho} c^{\text{ext}}(\mathbf{x}) & \text{if } c^{\text{int}}(\mathbf{x}) < 0, \\ +\infty & \text{otherwise,} \end{cases}$$

where  $c^{\text{int}}$  and  $c^{\text{ext}}$  are the interior and exterior constraint violation functions.

Note that the penalty parameter  $\rho$  is the same for both terms of  $z(\cdot; \rho)$ . The key property of this merit function is that for any  $\mathbf{x} \in \mathbb{R}^n$ ,

$$\lim_{\rho \searrow 0} z(\mathbf{x}; \rho) = \begin{cases} f(\mathbf{x}) & \text{if } \mathbf{x} \in \Omega_{<}^{\text{int}} \cap \Omega^{\text{ext}}, \\ +\infty & \text{otherwise,} \end{cases}$$

where  $\Omega_{<}^{\text{int}} = \{\mathbf{x} \in \mathbb{R}^n \mid c^{\text{int}}(\mathbf{x}) < 0\}$ . The merit function is close, but not identical to the extreme barrier function [7], as the latter is equal to  $f$  on the entire feasible set, including when  $\mathbf{x}$  satisfies  $c^{\text{int}}(\mathbf{x}) = 0$ .

## 2.2 The algorithm: MADS-PIP

The Penalty-Interior Point (PIP) method consists of generating a sequence of penalty-barrier parameters  $\{\rho_k\}$ , and finding a sequence of inexact solutions  $\{\mathbf{x}_k\} \in \mathbb{R}^n$  to subproblems of the form

$$\min_{\mathbf{x} \in \mathbb{R}^n} z(\mathbf{x}; \rho_k). \tag{P_{\rho_k}}$$

The paper provides conditions ensuring that as  $k$  goes to  $\infty$ , the sequence of solutions  $\{\mathbf{x}_k\}$  converges to a point satisfying nonsmooth necessary optimality conditions. In order for that to happen, the merit function  $z(\cdot; \rho_k)$  must converge to the objective function  $f$ . This occurs when  $\rho_k$  converges to 0. Algorithm 1 uses MADS to approximately solve the unconstrained subproblems, with an additional update criterion for updating the penalty-barrier parameter  $\rho_k$ . MADS stops solving subproblem  $(P_{\rho_k})$  when  $\Delta_{k+1}$  is sufficiently small. When the penalty-barrier parameter is reduced (i.e.,  $\rho_{k+1} < \rho_k$ ), the algorithm proceeds to the next subproblem  $(P_{\rho_{k+1}})$ . The strategy is described in the next paragraphs.

Algorithm 1 builds upon the MADS algorithm [7, 10]. Thanks to the use of the mesh, MADS accepts any strict decrease in the objective function value to declare an iteration successful. More recently, the ADS algorithm [9] proposes a strategy that does not require a mesh nor any sufficient decrease.

Starting from an initial guess  $\mathbf{x}_0$  and an initial *frame size*  $\Delta_0$ , Algorithm 1 generates a sequence of iterates  $\{\mathbf{x}_k\}$ . The best known solution at the start of iteration  $k$  is denoted  $\mathbf{x}_k$  and is called the *incumbent*. At each iteration  $k$ , the frame size is used to define the mesh  $\mathcal{M}_k$  (**Step 0**), as with OrthoMADS [1]. Then, the optional *search step* (**Step 1**) is performed, wherein various heuristics, e.g., quadratic models [25], surrogates [14], Latin

---

**Algorithm 1 Penalty-Interior Point with MADS (MADS-PIP)**


---

**Require:** Starting point/incumbent  $\mathbf{x}_0 \in \mathbb{R}^n$ , partition sets  $\mathcal{G}^{\text{int}}$  and  $\mathcal{G}^{\text{ext}}$  such that  $g_\ell(\mathbf{x}_0) < 0$  for all  $\ell \in \mathcal{G}^{\text{int}}$ , penalty contraction parameter  $\theta_\rho \in (0, 1)$ , initial frame size  $\Delta_0 > 0$ , frame size contraction/expansion parameter  $\theta_\Delta \in (0, 1) \cap \mathbb{Q}$ ,  $\beta > 1$ , set  $k = 0$ .

► **Step 0: Mesh definition** ◀

Set  $\mathcal{M}_k = \{\mathbf{x}_k + \delta_k \mathbf{z} : \mathbf{z} \in \mathbb{Z}^n\}$  where  $\delta_k = \min \left\{ \Delta_k, \frac{\Delta_k^2}{\Delta_0} \right\}$

► **Step 1: Search step (optional)** ◀

**if**  $z(\mathbf{s}; \rho_k) < z(\mathbf{x}_k; \rho_k)$  holds for some  $\mathbf{s}$  in a finite set  $\mathcal{S}_k \subset \mathcal{M}_k$  **then**

┌  $\mathbf{x}_{k+1} \leftarrow \mathbf{s}$ ,  $\Delta_{k+1} \leftarrow \theta_\Delta^{-1} \Delta_k$ ,  $\rho_{k+1} \leftarrow \rho_k$   
└ declare the iteration as **successful** and go to **Step 4**

► **Step 2: Poll step** ◀

Select a finite set of poll directions  $\mathcal{D}_k$  such that  $\{\mathbf{x}_k + \mathbf{d}_k \mid \mathbf{d}_k \in \mathcal{D}_k\} \subseteq \mathcal{F}_k$   
with  $\mathcal{F}_k = \{\mathbf{x} \in \mathcal{M}_k \mid \|\mathbf{x}_k - \mathbf{x}\| \leq \Delta_k\}$

**if**  $z(\mathbf{x}_k + \mathbf{d}_k; \rho_k) < z(\mathbf{x}_k; \rho_k)$  holds for some  $\mathbf{d}_k \in \mathcal{D}_k$  **then**

┌  $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \mathbf{d}_k$ ,  $\Delta_{k+1} \leftarrow \theta_\Delta^{-1} \Delta_k$ ,  $\rho_{k+1} \leftarrow \rho_k$   
└ declare the iteration as **successful** and go to **Step 4**

**else**

┌  $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k$ ,  $\Delta_{k+1} \leftarrow \theta_\Delta \Delta_k$   
└ declare the iteration as **unsuccessful**

► **Step 3: Penalty parameter update (at unsuccessful iterations)** ◀

**if**  $\Delta_{k+1} \leq \min \left\{ \rho_k^\beta, [\phi^{\text{prox}}(\mathbf{x}_k)]^2 \right\}$  **then**

┌  $\rho_{k+1} \leftarrow \theta_\rho \rho_k$       ►  $\mathbf{x}_{k+1} = \mathbf{x}_k$  is an approximate solution to subproblem  $(P_{\rho_k})$

**else**

┌  $\rho_{k+1} \leftarrow \rho_k$       ► The subproblem  $(P_{\rho_k})$  is not satisfactorily solved

► **Step 4: Stopping criterion check** ◀

**if** some stopping criterion is satisfied **then**

┌ **Stop**

**else**

┌  $k \leftarrow k + 1$  and go to **Step 0**

---

Hypercube sampling [43] and Nelder-Mead [16], are exploited in order to improve the solution, i.e., the search attempts to find a point  $\mathbf{s} \in \mathcal{M}_k$  satisfying  $z(\mathbf{s}; \rho_k) < z(\mathbf{x}_k; \rho_k)$ , which would become the incumbent of the next iteration. If the search step succeeds, the algorithm proceeds directly to the iteration  $k + 1$ , repeating the search step. If it fails, **Step 2**, i.e. the *poll step*, is invoked.

The poll step is restricted to the frame  $\mathcal{F}_k = \{\mathbf{x} \in \mathcal{M}_k \mid \|\mathbf{x}_k - \mathbf{x}\| \leq \Delta_k\}$  around the incumbent  $\mathbf{x}_k$ . A set of poll directions  $\mathcal{D}_k$  is generated, so that the set of trial points  $\{\mathbf{x}_k + \mathbf{d}_k \mid \mathbf{d}_k \in \mathcal{D}_k\}$  is a subset of  $\mathcal{F}_k$ , and thus  $\|\mathbf{d}_k\| \leq \Delta_k$  for all  $\mathbf{d}_k \in \mathcal{D}_k$ . If there exists a poll direction  $\mathbf{d}_k \in \mathcal{D}_k$  such that  $\mathbf{x}_k + \mathbf{d}_k$  improves the value of the function  $z(\cdot; \rho_k)$ , then the iteration is declared successful and the algorithm proceeds to iteration  $k + 1$ , otherwise the iteration is declared unsuccessful and the frame size is reduced by a factor  $\theta_\Delta \in (0, 1)$ . Whenever an iteration is unsuccessful,  $\mathcal{F}_k$  is called a *minimal frame*, and the incumbent point  $\mathbf{x}_k$  is called a *minimal frame center*. The novelty of the mixed penalty-barrier approach comes into play when iterations are unsuccessful. In such case, **Step 3** is invoked.

In **Step 3**, the algorithm verifies whether this measure is sufficiently small by balancing two quantities. The parameter  $\rho_k$  represents how well the problem  $(P_{\rho_k})$  represents  $(P_0)$ . The closer  $\rho_k$  is to zero, the better the representation. Therefore, we are not interested in providing a high precision for the stationarity measure if  $\rho_k$  is large. Furthermore, it is important to avoid trial points outside of  $\Omega^{\text{int}}$ , where the function  $z(\cdot; \rho_k)$  is infinite. To achieve this, the method takes into account the value of  $\phi^{\text{prox}}(\mathbf{x})$ , which is a measure of proximity to the boundary of  $\Omega^{\text{int}}$ . The criterion to determine if the frame size parameter is sufficiently small is

$$\Delta_{k+1} \leq \min\{\rho_k^\beta, [\phi^{\text{prox}}(\mathbf{x}_k)]^2\}. \quad (2)$$

If the latter is satisfied, the barrier-penalty parameter is decreased by a factor  $\theta_\rho \in (0, 1)$ , otherwise it remains the same and the algorithm proceeds to the next iteration on the same subproblem  $(P_{\rho_k})$ . The criterion (2) ensures that if the parameter  $\rho_k$  tends to zero, so does the frame size  $\Delta_k$ , at least on the subsequence where the criterion is satisfied. A trivial consequence of the updating rule of  $\rho_k$  is that, if the criterion is satisfied for an infinite subset of iterations, then  $\rho_k \rightarrow 0$ . Furthermore, we prove in Theorem 3.6 that (2) also ensures that, for  $k$  large enough, the entire minimal frame belongs to the interior of  $\Omega^{\text{int}}$ , i.e.,  $c^{\text{int}}(\mathbf{s}) < 0$  for all  $\mathbf{s} \in \mathcal{F}_k$ .

We conclude the algorithm description with brief comments on the initialization of the sets of indices  $\mathcal{G}^{\text{int}}$  and  $\mathcal{G}^{\text{ext}}$ . It is necessary that  $g_\ell(\mathbf{x}_0) < 0$  for all  $\ell \in \mathcal{G}^{\text{int}}$ , otherwise one could have  $c^{\text{int}}(\mathbf{x}_0) \geq 0$  which may result in  $z(\mathbf{x}_0; \rho_0) = +\infty$ . This requirement can be satisfied by simply setting  $\ell \in \mathcal{G}^{\text{int}}$  if  $g_\ell(\mathbf{x}_0) < 0$ , and  $\ell \in \mathcal{G}^{\text{ext}}$  otherwise. This strategy produces  $\mathcal{G}^{\text{int}} = \emptyset$  when  $g_\ell(\mathbf{x}_0) \geq 0$  for all  $\ell = 1, 2, \dots, m$ .

Finally, observe that the partition sets  $\mathcal{G}^{\text{int}}$  and  $\mathcal{G}^{\text{ext}}$  may undergo a finite number of updates during the execution of the algorithm, as detailed in Section 4.2. The convergence analysis is then carried out on the sequence of iterations following the last change to the partition. To simplify the notation and without loss of generality, we omitted the iteration subscript  $k$  in the definition of the sets  $\mathcal{G}^{\text{int}}$  and  $\mathcal{G}^{\text{ext}}$  in Algorithm 1.

### 3 Convergence analysis

This section is devoted to the convergence properties of the sequence of iterates  $\{\mathbf{x}_k\}$  produced by Algorithm 1. The first part introduces the framework of Clarke’s directional derivatives, which is a suitable concept for nonsmooth functions. Additionally, the *path-refining* directions are defined. For such directions, it is possible to derive properties that can be exploited to understand the behaviour at the limit of the sequences generated by the Algorithm. Then, a subsection studies conditions under which the feasibility of end-path points can be ensured. Finally, the last subsection provides further conditions ensuring stationarity results.

#### 3.1 Preliminary results

This section presents the key results required for the convergence analysis of Algorithm 1. We begin with a standard assumption ensuring that MADS behaves properly on the associated unconstrained subproblem.

**Assumption 1.** *The lower level sets  $\mathcal{L}_f(\alpha) = \{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) \leq \alpha\}$  are bounded for all  $\alpha \in \mathbb{R}$ .*

Assumption 1 implies that the lower level sets  $\mathcal{L}_z(\alpha; \rho) = \{\mathbf{x} \in \mathbb{R}^n \mid z(\mathbf{x}; \rho) \leq \alpha\}$  are bounded for all  $\rho > 0$ . In fact, by definition of  $z(\cdot; \rho)$ , one has  $z(\mathbf{x}; \rho) \geq f(\mathbf{x})$  for all  $\mathbf{x} \in \mathbb{R}^n$  and all positive values of  $\rho$ , which implies that  $\mathcal{L}_z(\alpha; \rho) \subseteq \mathcal{L}_f(\alpha)$  for all  $\alpha \in \mathbb{R}$  and  $\rho \in (0, +\infty)$ .

The convergence analysis in this work does not establish properties for the entire sequence of iterates generated by the algorithm. As in the case of Generalized Pattern Search (GPS) and MADS [10], where specific *refining* subsequences are considered, we define an index set corresponding to the iterations at which the penalty parameter  $\rho_k$  is updated – these are the unsuccessful iterations for which convergence results are derived. They are named *path-following* subsequences, a term adopted from the literature on interior-point methods (see, e.g., [18]), to describe sequences of approximate solutions to the barrier subproblems. In addition, to guide the reader, we introduce a second index set that identifies the accumulation points associated with the sequence of incumbents.

**Definition 3.1.** *The sequence  $\{\mathbf{x}_k\}_{k \in \mathcal{K}_\rho}$  of iterates generated by Algorithm 1, corresponding to iterations where  $\rho_{k+1} < \rho_k$ , is said to be a path-following subsequence. Furthermore, a point  $\bar{\mathbf{x}}$  is said to be an end-path point if there exists  $\mathcal{K}_\rho^{\mathbf{x}} \subseteq \mathcal{K}_\rho$  such that  $\lim_{k \in \mathcal{K}_\rho^{\mathbf{x}}} \mathbf{x}_k = \bar{\mathbf{x}}$ . In such case,  $\{\mathbf{x}_k\}_{k \in \mathcal{K}_\rho^{\mathbf{x}}}$  is referred to as an end-path subsequence.*

**Step 3** of Algorithm 1 ensures that all iterations  $k \in \mathcal{K}_\rho$  of a path-following subsequence are unsuccessful. Thus, since  $\mathcal{K}_\rho^{\mathbf{x}} \subseteq \mathcal{K}_\rho$ , we have for all  $k \in \mathcal{K}_\rho^{\mathbf{x}}$

$$\begin{aligned} z(\mathbf{x}_k + \mathbf{d}_k; \rho_k) &\geq z(\mathbf{x}_k; \rho_k) && \text{for all } \mathbf{d}_k \in \mathcal{D}_k, \\ \Delta_{k+1} &\leq \min\{\rho_k^\beta, [\phi^{\text{prox}}(\mathbf{x}_k)]^2\} && \text{and } \Delta_{k+1} = \theta_\Delta \Delta_k. \end{aligned} \tag{3}$$

In the following, we refer to the set of successful iteration indexes as  $\mathcal{K}_s$ , and to the set of unsuccessful indexes as  $\mathcal{K}_u$ . Note that  $\mathcal{K}_\rho^{\mathbf{x}} \subseteq \mathcal{K}_\rho \subseteq \mathcal{K}_u$ . The first result to be shown is that the index set  $\mathcal{K}_\rho$  is infinite, which we prove to be equivalent to showing that the penalty parameter  $\rho_k$  goes to zero as  $k \rightarrow \infty$ . Therefore, we prove that the updating rule, i.e.

$\Delta_{k+1} \leq \min\{\rho_k^\beta, [\phi^{\text{prox}}(\mathbf{x}_k)]^2\}$  with  $k \in \mathcal{K}_u$ , in **Step 3** of Algorithm 1, is eventually satisfied for any value of  $\rho$ . Within the proof, we exploit the known results for MADS algorithms. In particular, let us state the following proposition. The proof is omitted as the result follows from [6, Proposition 3.4].

**Proposition 3.2.** *Let Assumption 1 be satisfied. If the sequence of penalty-barrier parameters  $\{\rho_k\}$  produced by Algorithm 1 satisfies  $\rho_k = \rho_0$  for all  $k \in \mathbb{N}$ , then  $\liminf_{k \rightarrow \infty} \Delta_k = 0$ .*

Note that, when the penalty-barrier parameter is fixed to  $\rho_0$  for all iterates, the algorithm is solving the unconstrained problem of minimizing  $z(\mathbf{x}; \rho_0)$ , and behaves as MADS. That is why it is not necessary to provide the proof of the latter proposition.

Let us now prove the following result on the convergence of the sequence of penalty-barrier parameters  $\{\rho_k\}$ .

**Theorem 3.3.** *Let Assumption 1 be satisfied. If  $\{\mathbf{x}_k\}_{k \in \mathcal{K}_\rho}$  is a path-following subsequence generated by Algorithm 1, then  $\mathcal{K}_\rho$  is infinite and*

$$\lim_{k \rightarrow \infty} \rho_k = 0. \quad (4)$$

*Proof.* By contradiction, assume that  $\mathcal{K}_\rho$  is finite. Without loss of generality, assume that  $\mathcal{K}_\rho = \emptyset$ . Therefore,  $\rho_k = \rho_0$  for all  $k \in \mathbb{N}$ , and Theorem 3.2 ensures that  $\liminf_{k \rightarrow \infty} \Delta_k = 0$ .

Since  $\Delta_{k+1} < \Delta_k$  if and only if  $k \in \mathcal{K}_u$ , it follows that there exists an infinite index set  $\mathcal{K}_u^\Delta \subseteq \mathcal{K}_u$  such that

$$\lim_{k \in \mathcal{K}_u^\Delta} \Delta_k = 0. \quad (5)$$

The condition of **Step 3** is never satisfied, and therefore  $\Delta_{k+1} > \min\{\rho_0^\beta, [\phi^{\text{prox}}(\mathbf{x}_k)]^2\} > 0$  for all  $k \in \mathcal{K}_u^\Delta$ . Taking the limit for  $k \rightarrow \infty$  and  $k \in \mathcal{K}_u^\Delta$  and using (5), it follows that

$$\lim_{k \in \mathcal{K}_u^\Delta} \phi^{\text{prox}}(\mathbf{x}_k) = 0, \quad \text{and} \quad \lim_{k \in \mathcal{K}_u^\Delta} c^{\text{int}}(\mathbf{x}_k) = 0, \quad \text{and} \quad \lim_{k \in \mathcal{K}_u^\Delta} z(\mathbf{x}_k, \rho_0) = +\infty.$$

This contradicts the fact that  $z(\mathbf{x}_{k+1}; \rho_0) \leq z(\mathbf{x}_k; \rho_0) \leq z(\mathbf{x}_0; \rho_0) < +\infty$ . Thus, the index set  $\mathcal{K}_\rho$  is infinite.

For any  $s \in \mathbb{N}$ , let  $k_s \in \mathcal{K}_\rho$  be the  $s$ -th iteration index in the set  $\mathcal{K}_\rho$ . Therefore, it follows that  $\rho_{k_s} = \theta_\rho \rho_{k_{s-1}} = \theta_\rho^s \rho_0$ , and  $\lim_{s \rightarrow \infty} \rho_{k_s} = \lim_{s \rightarrow \infty} \theta_\rho^s \rho_0 = 0$ , since  $\theta_\rho \in (0, 1)$ . The result follows by noticing that the sequence  $\{\rho_k\}$  is non-increasing.  $\square$

The next result shows that the limit of frame-size parameters is equal to zero for path-following subsequences.

**Theorem 3.4.** *Let Assumption 1 be satisfied. If  $\{\mathbf{x}_k\}_{k \in \mathcal{K}_\rho}$  is a path-following subsequence generated by Algorithm 1, and  $\{\Delta_k\}_{k \in \mathcal{K}_\rho}$  is the sequence of frame sizes, then*

$$\lim_{k \in \mathcal{K}_\rho} \Delta_k = 0. \quad (6)$$

*Proof.* Theorem 3.3 ensures that  $\mathcal{K}_\rho$  is an infinite index set. For every  $k \in \mathcal{K}_\rho$ ,

$$\Delta_{k+1} \leq \min\{\rho_k^\beta, [\phi^{\text{prox}}(\mathbf{x}_k)]^2\} \leq \rho_k^\beta, \quad (7)$$

the same theorem guarantees that (4) holds. Taking the limit as  $k \rightarrow \infty$  and  $k \in \mathcal{K}_\rho$  on (7) gives  $\lim_{k \in \mathcal{K}_\rho} \Delta_{k+1} = 0$ . The algorithm sets  $\Delta_{k+1} = \theta_\Delta \Delta_k$  for all  $k \in \mathcal{K}_\rho \subseteq \mathcal{K}_u$ , thus  $\lim_{k \in \mathcal{K}_\rho} \Delta_k = 0$ . □

As a consequence of Theorem 3.4, we deduce that end-path points exist.

**Corollary 3.5.** *Let Assumption 1 be satisfied. If  $\{\mathbf{x}_k\}_{k \in \mathcal{K}_\rho}$  is a path-following subsequence generated by Algorithm 1, then there exists at least one end-path point.*

The results of Theorem 3.3, Theorem 3.4, and Theorem 3.5 are of paramount importance and they are used throughout the analysis to derive further properties. In the following, Equation (6) is exploited to show additional preliminary results on the path-following sequence generated by Algorithm 1. In particular, the minimal frames are proved to belong to the interior of  $\Omega^{\text{int}}$  in point (i) of the proposition, which additionally allows to prove a relation between the different terms of  $z(\cdot; \rho)$ , facilitating the analysis of the generalized directional derivatives later in the paper.

**Proposition 3.6.** *Let Assumption 1 be satisfied,  $\{\mathbf{x}_k\}_{k \in \mathcal{K}_\rho}$  be a path-following subsequence, and  $\mathcal{K}_\rho^x \subseteq \mathcal{K}_\rho$  be the index set of an end-path subsequence with end-path point  $\bar{\mathbf{x}}$ . If  $c^{\text{int}}(\cdot)$  is Lipschitz continuous near  $\bar{\mathbf{x}}$ , then for all  $k \in \mathcal{K}_\rho^x$  sufficiently large, the following holds*

$$\begin{aligned} (i) \quad & c^{\text{int}}(\mathbf{x}_k + \mathbf{d}) < 0 \text{ for any direction } \mathbf{d} \in \mathbb{R}^n \text{ such that } \mathbf{x}_k + \mathbf{d} \in \mathcal{F}_k; \\ (ii) \quad & f(\mathbf{p}_k) - f(\mathbf{x}_k) + \frac{1}{\rho_k} (c^{\text{ext}}(\mathbf{p}_k) - c^{\text{ext}}(\mathbf{x}_k)) \geq \frac{\rho_k}{c^{\text{int}}(\mathbf{p}_k)} (c^{\text{int}}(\mathbf{p}_k) - c^{\text{int}}(\mathbf{x}_k)) \end{aligned} \quad (8)$$

where  $\mathbf{d}_k$  is any poll direction in  $\mathcal{D}_k$ , and  $\mathbf{p}_k = \mathbf{x}_k + \mathbf{d}_k$  is the corresponding poll point.

*Proof.* (i) We first prove the result for  $\phi^{\text{prox}}(\mathbf{x}_k + \mathbf{d})$  and then deduce that it also holds for  $c^{\text{int}}(\mathbf{x}_k + \mathbf{d})$ . Indeed, for any  $\mathbf{y} \in \mathbb{R}^n$ ,  $\phi^{\text{prox}}(\mathbf{y}) < 0$  if and only if  $c^{\text{int}}(\mathbf{y}) < 0$ , and  $\phi^{\text{prox}}(\mathbf{y}) = 0$  if and only if  $c^{\text{int}}(\mathbf{y}) = 0$ . Let  $\mathbf{d} \in \mathbb{R}^n$  be such that  $\mathbf{x}_k + \mathbf{d} \in \mathcal{F}_k$ . By the definition of  $\mathcal{F}_k$ ,  $\|\mathbf{d}\| \leq \Delta_k$ . Since  $\lim_{k \in \mathcal{K}_\rho^x} \mathbf{x}_k = \bar{\mathbf{x}}$ , and  $\lim_{k \in \mathcal{K}_\rho} \Delta_k = 0$  (from Theorem 3.4), it follows that

$$\lim_{k \in \mathcal{K}_\rho^x} \phi^{\text{prox}}(\mathbf{x}_k) = \lim_{k \in \mathcal{K}_\rho^x} \phi^{\text{prox}}(\mathbf{x}_k + \mathbf{d}) = \phi^{\text{prox}}(\bar{\mathbf{x}}) \leq 0,$$

where the last inequality follows from the fact that  $c^{\text{int}}(\mathbf{x}_k) < 0$  for all  $k$ , and, therefore,  $\phi^{\text{prox}}(\mathbf{x}_k) < 0$  for all  $k$ . Equation (i) follows trivially when  $c^{\text{int}}(\bar{\mathbf{x}}) < 0$ , so we consider the situation where  $\phi^{\text{prox}}(\bar{\mathbf{x}}) = c^{\text{int}}(\bar{\mathbf{x}}) = 0$ . All iterations  $k \in \mathcal{K}_\rho^x$  are unsuccessful, and so  $\mathbf{x}_{k+1} = \mathbf{x}_k$ ,  $\Delta_{k+1} = \theta_\Delta \Delta_k$ , with  $\theta_\Delta \in (0, 1)$ , and (2) ensures that  $\Delta_{k+1} \leq \phi^{\text{prox}}(\mathbf{x}_k)^2$ . Since  $c^{\text{int}}(\cdot)$  is Lipschitz near  $\bar{\mathbf{x}}$ , we deduce that  $\phi^{\text{prox}}(\cdot)$  is also Lipschitz near  $\bar{\mathbf{x}}$ . Thus, there exists a constant  $L_\phi > 0$  such that for all  $k \in \mathcal{K}_\rho^x$  sufficiently large

$$|\phi^{\text{prox}}(\mathbf{x}_k) - \phi^{\text{prox}}(\mathbf{x}_k + \mathbf{d})| \leq L_\phi \|\mathbf{d}\| \leq L_\phi \Delta_k = \frac{L_\phi \Delta_{k+1}}{\theta_\Delta} \leq \frac{L_\phi [\phi^{\text{prox}}(\mathbf{x}_k)]^2}{\theta_\Delta}.$$

It follows that

$$\phi^{\text{prox}}(\mathbf{x}_k + \mathbf{d}) \leq \frac{L_\phi}{\theta_\Delta} \phi^{\text{prox}}(\mathbf{x}_k)^2 + \phi^{\text{prox}}(\mathbf{x}_k) = \phi^{\text{prox}}(\mathbf{x}_k) \left( \frac{L_\phi}{\theta_\Delta} \phi^{\text{prox}}(\mathbf{x}_k) + 1 \right).$$

The right-hand-side is strictly negative when  $k \in \mathcal{K}_\rho^{\mathbf{x}}$  is large, since  $\phi^{\text{prox}}(\mathbf{x}_k) < 0 = \phi^{\text{prox}}(\bar{\mathbf{x}})$ ,  $L_\phi > 0$ , and  $\theta_\Delta > 0$ . Thus, the entire frame is strictly feasible with respect to the constraints treated by the logarithmic barrier, which proves (i).

(ii) Consider  $k \in \mathcal{K}_\rho^{\mathbf{x}}$ ,  $\mathbf{d}_k \in \mathcal{D}_k$  and set  $\mathbf{p}_k = \mathbf{x}_k + \mathbf{d}_k$ . Equation (3) can be rewritten as

$$f(\mathbf{p}_k) - f(\mathbf{x}_k) + \frac{1}{\rho_k} (c^{\text{ext}}(\mathbf{p}_k) - c^{\text{ext}}(\mathbf{x}_k)) + \rho_k \log \left( \frac{-c^{\text{int}}(\mathbf{x}_k)}{-c^{\text{int}}(\mathbf{p}_k)} \right) \geq 0.$$

Using the fact that  $c^{\text{int}}(\mathbf{x}_k) < 0$ , that (i) holds for  $\mathbf{d} = \mathbf{d}_k$  when  $k$  is large, and since  $\log(t) \leq t-1$  for all  $t > 0$ , it follows that

$$\log \left( \frac{-c^{\text{int}}(\mathbf{x}_k)}{-c^{\text{int}}(\mathbf{p}_k)} \right) \leq \frac{1}{-c^{\text{int}}(\mathbf{p}_k)} (c^{\text{int}}(\mathbf{p}_k) - c^{\text{int}}(\mathbf{x}_k)).$$

Combining the two inequalities concludes the proof.  $\square$

For a function  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ , the Clarke generalized directional derivative [24] of  $f$  at  $\mathbf{x} \in \mathbb{R}^n$  along a direction  $\bar{\mathbf{d}} \in \mathbb{R}^n$  is given by

$$f^\circ(\mathbf{x}; \bar{\mathbf{d}}) = \limsup_{\substack{\mathbf{y} \rightarrow \mathbf{x} \\ t \searrow 0}} \frac{f(\mathbf{y} + t\bar{\mathbf{d}}) - f(\mathbf{y})}{t}.$$

The generalized derivative proposed by Clarke is a powerful tool for the analysis of non-smooth optimization problems. When the function  $f$  is Lipschitz near a point  $\mathbf{x}$ , then the superior limit exists and is finite, see [24, Chapter 2]. The following technical result [7, Proposition 3.9] is used in the theoretical analysis.

**Proposition 3.7.** *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be Lipschitz continuous near  $\mathbf{x}$ . Then*

$$f^\circ(\mathbf{x}, \mathbf{d}) = \limsup_{\substack{\mathbf{y} \rightarrow \mathbf{x} \\ t \searrow 0 \\ \mathbf{w} \rightarrow \mathbf{d}}} \frac{f(\mathbf{y} + t\mathbf{w}) - f(\mathbf{y})}{t}. \quad (9)$$

The next definition introduces formally the *path-refining* directions. These are similar to the *refining* directions introduced in [7], with the difference that the behaviour of the values of  $c^{\text{int}}(\cdot)$  is considered.

**Definition 3.8.** *Let  $\{\mathbf{x}_k\}_{k \in \mathcal{K}_\rho}$  be a path-following subsequence, and  $\mathcal{K}_\rho^{\mathbf{x}} \subseteq \mathcal{K}_\rho$  be the index set of an end-path subsequence with end-path point  $\bar{\mathbf{x}}$  such that the normalized sequence of poll directions  $\{\mathbf{d}_k\} \subset \mathcal{D}_k$  converges to some direction  $\bar{\mathbf{d}} \in \mathbb{R}^n$ , i.e.,  $\lim_{k \in \mathcal{K}_\rho^{\mathbf{x}}} \frac{\mathbf{d}_k}{\|\mathbf{d}_k\|} = \bar{\mathbf{d}}$ . Then,  $\bar{\mathbf{d}}$  is said to be a path-refining direction for  $\bar{\mathbf{x}}$ , if*

$$c^{\text{int}}(\mathbf{x}_k + \mathbf{d}_k) \leq c^{\text{int}}(\mathbf{x}_k) \quad \text{for all } k \in \mathcal{K}_\rho^{\mathbf{x}}, \quad (10)$$

**Remark 3.9.** Suppose that  $c^\circ(\mathbf{x}; \mathbf{d}) < 0$ . Then, by the definition of the Clarke directional derivative, for any sequences  $\{\mathbf{x}_k\} \rightarrow \mathbf{x}$  and  $\{\mathbf{d}_k\}$  satisfying

$$\lim_{k \rightarrow \infty} \frac{\mathbf{d}_k}{\|\mathbf{d}_k\|} = \mathbf{d} \quad \text{and} \quad \|\mathbf{d}_k\| \rightarrow 0,$$

there exists  $k_0$  such that  $c(\mathbf{x}_k + \mathbf{d}_k) \leq c(\mathbf{x}_k)$  for all  $k \geq k_0$ . Thus, setting  $\mathcal{K}_\rho^\mathbf{x} := \{k : k \geq k_0\}$  implies that the vector  $\mathbf{d}$  is a path-refining direction for  $\mathbf{x}$ .

It is possible to derive a property concerning the violation function  $c^{\text{ext}}(\cdot)$ , defined in (1), along path-refining directions. Such a property is presented and proved in the following proposition.

**Proposition 3.10.** Let Assumption 1 be satisfied,  $\{\mathbf{x}_k\}_{k \in \mathcal{K}_\rho}$  be a path-following subsequence, with end-path point  $\bar{\mathbf{x}}$ , and let  $\mathbf{d}$  be a path-refining direction. If  $f, c^{\text{int}}$ , and  $c^{\text{ext}}$  are Lipschitz continuous near  $\bar{\mathbf{x}}$ , then

$$(c^{\text{ext}})^\circ(\bar{\mathbf{x}}; \mathbf{d}) \geq 0. \quad (11)$$

*Proof.* By the definition of path-refining direction, a set  $\mathcal{K}_\rho^\mathbf{x}$  exists such that  $\lim_{k \in \mathcal{K}_\rho^\mathbf{x}} \frac{\mathbf{d}_k}{\|\mathbf{d}_k\|} = \bar{\mathbf{d}}$  and (10) hold. Theorem 3.6 ensures  $c^{\text{int}}(\mathbf{x}_k + \mathbf{d}_k) < 0$  for all  $\mathbf{d}_k \in \mathcal{D}_k$  and  $k \in \mathcal{K}_\rho^\mathbf{x}$  sufficiently large. Recalling  $\rho_k > 0$  for all  $k$ , and using (10), it follows for  $k \in \mathcal{K}_\rho^\mathbf{x}$  sufficiently large

$$\frac{\rho_k}{c^{\text{int}}(\mathbf{x}_k + \mathbf{d}_k)} (c^{\text{int}}(\mathbf{x}_k + \mathbf{d}_k) - c^{\text{int}}(\mathbf{x}_k)) \geq 0.$$

Thus, using (8) from Theorem 3.6, for sufficiently large  $k \in \mathcal{K}_\rho^\mathbf{x}$ , it holds

$$f(\mathbf{x}_k + \mathbf{d}_k) - f(\mathbf{x}_k) + \frac{1}{\rho_k} (c^{\text{ext}}(\mathbf{x}_k + \mathbf{d}_k) - c^{\text{ext}}(\mathbf{x}_k)) \geq 0.$$

Dividing both sides of the latter equation by  $\|\mathbf{d}_k\|$ , multiplying by  $\rho_k$  and taking the superior limit for  $k \in \mathcal{K}_\rho^\mathbf{x}$

$$\limsup_{k \in \mathcal{K}_\rho^\mathbf{x}} \left[ \rho_k \frac{f(\mathbf{x}_k + \mathbf{d}_k) - f(\mathbf{x}_k)}{\|\mathbf{d}_k\|} + \frac{c^{\text{ext}}(\mathbf{x}_k + \mathbf{d}_k) - c^{\text{ext}}(\mathbf{x}_k)}{\|\mathbf{d}_k\|} \right] \geq 0. \quad (12)$$

Since  $f(\cdot)$  is Lipschitz around  $\bar{\mathbf{x}}$ , then the sequence  $\left\{ \frac{f(\mathbf{x}_k + \mathbf{d}_k) - f(\mathbf{x}_k)}{\|\mathbf{d}_k\|} \right\}$  is bounded. Recalling that, by Theorem 3.3,  $\lim_{k \rightarrow \infty} \rho_k = 0$ , we can write

$$\limsup_{k \in \mathcal{K}_\rho^\mathbf{x}} \frac{c^{\text{ext}}(\mathbf{x}_k + \mathbf{d}_k) - c^{\text{ext}}(\mathbf{x}_k)}{\|\mathbf{d}_k\|} \geq 0.$$

Using the fact that  $c^{\text{ext}}(\cdot)$  is Lipschitz near  $\bar{\mathbf{x}}$ , by Theorem 3.7, we can write

$$(c^{\text{ext}})^\circ(\bar{\mathbf{x}}; \mathbf{d}) \geq \limsup_{k \in \mathcal{K}_\rho^\mathbf{x}} \frac{c^{\text{ext}}(\mathbf{x}_k + \mathbf{d}_k) - c^{\text{ext}}(\mathbf{x}_k)}{\|\mathbf{d}_k\|} \geq 0,$$

proving (11), and concluding the proof.  $\square$

The result of Theorem 3.10 is used to define conditions to ensure feasibility for the end-path points with respect to Problem  $(P_0)$ . Note that any sequence  $\{\mathbf{x}_k\}$  of incumbent points generated by the algorithm is such that  $c^{\text{int}}(\mathbf{x}_k) < 0$  for all  $k$ , so that any limit point  $\bar{\mathbf{x}}$  of the sequence belongs to  $\Omega^{\text{int}}$ . Therefore, in order to show that  $\bar{\mathbf{x}}$  belongs to  $\Omega$ , one has to show that  $c^{\text{ext}}(\bar{\mathbf{x}}) = 0$ . Since  $c^{\text{ext}}(\mathbf{x}) \geq 0$  for all  $\mathbf{x} \in \mathbb{R}^n$ , any point  $\bar{\mathbf{x}}$  such that  $c^{\text{ext}}(\bar{\mathbf{x}}) = 0$  is a global minimum of  $c^{\text{ext}}(\cdot)$ , i.e., for any  $\epsilon > 0$ , any  $\mathbf{y} \in \mathcal{B}(\bar{\mathbf{x}}, \epsilon)$  is such that  $c^{\text{ext}}(\mathbf{y}) \geq c^{\text{ext}}(\bar{\mathbf{x}})$ , where  $\mathcal{B}(\bar{\mathbf{x}}, \epsilon)$  is a ball centred at  $\bar{\mathbf{x}}$ . It trivially follows that  $(c^{\text{ext}})^\circ(\bar{\mathbf{x}}; \mathbf{d}) \geq 0$  for all  $\mathbf{d} \in \mathbb{R}^n$ . If, on the other hand,  $c^{\text{ext}}(\bar{\mathbf{x}}) > 0$ , there might exist directions  $\mathbf{d} \in \mathbb{R}^n$  such that  $(c^{\text{ext}})^\circ(\bar{\mathbf{x}}; \mathbf{d}) < 0$ . In view of Theorem 3.10, such directions cannot be path-refining directions.

## 3.2 Feasibility

In this subsection, results involving all the introduced objects are shown, which allow to define conditions under which the end-path points generated by Algorithm 1 are feasible for Problem  $(P_0)$ .

The following provides the definition of the hypertangent cone [24].

**Definition 3.11** (Hypertangent Cone). *A vector  $\mathbf{d} \in \mathbb{R}^n$  is said to be hypertangent to  $\Omega$  at  $\mathbf{x}$  if there exist  $\epsilon > 0$  such that*

$$\mathbf{y} + t\mathbf{w} \in \Omega \text{ for all } \mathbf{y} \in \Omega \cap \mathcal{B}(\mathbf{x}, \epsilon), \quad \mathbf{w} \in \mathcal{B}(\mathbf{d}, \epsilon), \quad t \in (0, \epsilon).$$

The set of all hypertangent vectors to  $\Omega$  at  $\mathbf{x}$  is called the hypertangent cone, denoted as  $\mathcal{T}_\Omega^H(\mathbf{x})$ . The hypertangent cone appears in a necessary optimality condition for a solution  $\mathbf{x} \in \Omega$  to be a local minimizer of Problem  $(P_0)$ , for more details, see [10, Theorem 6.10]. In fact, if  $f$  is Lipschitz continuous near a local minimizer  $\mathbf{x} \in \Omega$  for Problem  $(P_0)$ , then  $f^\circ(\mathbf{x}; \mathbf{d}) \geq 0$  for every direction  $\mathbf{d} \in \mathcal{T}_\Omega^H(\bar{\mathbf{x}})$ .

The next definition introduces a first constraint qualification, which we will use in the analysis of feasibility.

**Definition 3.12** (Feasibility Constraint Qualification). *A point  $\mathbf{x} \in \mathbb{R}^n$  is said to satisfy the feasibility constraint qualification (FCQ) for problem  $(P_0)$ , if the following conditions hold:*

- (a)  $c^{\text{int}}(\cdot)$  and  $c^{\text{ext}}(\cdot)$  are Lipschitz continuous near  $\mathbf{x}$ ;
- (b) either  $\mathbf{x} \in \Omega$ , or  $\mathbf{x} \in \Omega^{\text{int}} \setminus \Omega^{\text{ext}}$  and there exists a direction  $\mathbf{d} \in \mathbb{R}^n$  such that  $(c^{\text{ext}})^\circ(\mathbf{x}; \mathbf{d}) < 0$  and
 
$$(c^{\text{int}})^\circ(\mathbf{x}; \mathbf{d}) < 0 \quad \text{if } c^{\text{int}}(\mathbf{x}) = 0. \tag{13}$$

The FCQ, as introduced in Theorem 3.12, is designed to ensure that, at any candidate limit point  $\mathbf{x}$ , the local geometry of the constraint system is sufficiently regular to guarantee that the algorithmic path can approach feasibility. This is particularly important in the nonsmooth, nonconvex setting, where classical constraints qualifications such as MFCQ [42], may not apply or may fail to capture the subtleties of the problem structure. At its core, the FCQ requires the point to be feasible with respect to the exterior constraints. Otherwise, for points that are infeasible with respect to the exterior violation function, i.e., when  $c^{\text{ext}}(\mathbf{x}) > 0$ , additional

conditions are required. When using a penalty approach to deal with nonlinear constraints, the problem of feasibility can be translated as the minimization of the norm of constraint violations, e.g.  $v(\mathbf{x}) = \sum_{\ell=1}^m [g_{\ell}^+(\mathbf{x})]^2 + \sum_{j=1}^p [h(\mathbf{x})]^2$ . When mixing the penalty approach with an interior violation function, the latter translation does not properly reflect the constraint-handling strategy as the merit function is not defined for all  $\mathbf{x} \in \mathbb{R}^n$ . A coherent representation of the feasibility problem when using the proposed PIP strategy would be

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^n} \quad & c^{\text{ext}}(\mathbf{x}) \\ \text{s.t.} \quad & c^{\text{int}}(\mathbf{x}) \leq 0. \end{aligned} \tag{P}_{feas}$$

The feasible region of  $(P_{feas})$  is  $\Omega^{\text{int}}$ . The optimization algorithm may stall at points that are stationary for  $(P_{feas})$ , thus, points such that  $(c^{\text{ext}})^{\circ}(\mathbf{x}; \mathbf{d}) \geq 0$  for all  $\mathbf{d} \in \mathcal{T}^H(\mathbf{x}, \Omega^{\text{int}})$ . By definition of hypertangent cone, and by definition of Clarke generalized derivative, it follows trivially that if a direction  $\mathbf{d}$  is such that  $(c^{\text{int}})^{\circ}(\mathbf{x}, \mathbf{d}) < 0$ , as in (13), then  $\mathbf{d}$  belongs to  $\mathcal{d} \in \mathcal{T}^H(\mathbf{x}, \Omega^{\text{int}})$ . Therefore, FCQ ensures that points are not stationary for  $(P_{feas})$ .

For our asymptotic analysis, it is well known that, for direct search methods, a density assumption on the sequence of sets of poll directions  $\mathcal{D}_k$  [7] is needed. Let us now formalize the notion of density over the unit sphere.

**Definition 3.13.** *Let  $\mathcal{K}$  be an infinite subset of indices. A sequence of sets of directions  $\{\mathcal{D}_k\}_{k \in \mathcal{K}}$  is said to be dense in the unit sphere  $\mathcal{B}(\mathbf{0}_n, 1)$  if, for any  $\bar{\mathbf{d}} \in \mathcal{B}(\mathbf{0}_n, 1)$ , there exists a sequence  $\{\mathbf{d}_k\}_{k \in \mathcal{K}^{\mathbf{d}}}$ , with  $\mathbf{d}_k \in \mathcal{D}_k$  and  $\mathcal{K}^{\mathbf{d}} \subseteq \mathcal{K}$  such that  $\lim_{\substack{k \rightarrow +\infty \\ k \in \mathcal{K}^{\mathbf{d}}}} \frac{\mathbf{d}_k}{\|\mathbf{d}_k\|} = \bar{\mathbf{d}}$ .*

In what follows, we show that any end-path point satisfying the FCQ (as in Theorem 3.12) is feasible for  $(P_0)$ .

**Theorem 3.14.** *Let Assumption 1 hold and  $\{\mathbf{x}_k\}$  be the sequence generated by Algorithm 1. Let  $\{\mathbf{x}_k\}_{k \in \mathcal{K}_{\rho}}$  be a path-following subsequence of  $\{\mathbf{x}_k\}$  and consider  $\mathcal{K}_{\rho}^{\mathbf{x}} \subseteq \mathcal{K}_{\rho}$  to be the index set of an end-path subsequence with end-path point  $\bar{\mathbf{x}}$ . Assume that  $f$  is Lipschitz continuous near  $\bar{\mathbf{x}}$ , that the sequence of sets of poll directions  $\{\mathcal{D}_k\}$  is dense in the unit sphere and that  $\bar{\mathbf{x}}$  satisfies the FCQ. Then,  $\bar{\mathbf{x}}$  is feasible for problem  $(P_0)$ .*

*Proof.* By definition of the iterates  $\mathbf{x}_k$  generated by Algorithm 1, it follows that  $c^{\text{int}}(\mathbf{x}_k) < 0$  for all  $k$ , and continuity implies that  $c^{\text{int}}(\bar{\mathbf{x}}) \leq 0$  (i.e.,  $\bar{\mathbf{x}} \in \Omega^{\text{int}}$ ). Thus, under the assumption that  $\bar{\mathbf{x}}$  satisfies the FCQ, it suffices to prove that  $\bar{\mathbf{x}} \in \Omega^{\text{ext}}$  in order to deduce that  $\bar{\mathbf{x}}$  is feasible for problem  $(P_0)$ . The analysis is divided into two cases.

**Case (i)  $c^{\text{int}}(\bar{\mathbf{x}}) < 0$ :** Since the sets  $\{\mathcal{D}_k\}$  are dense in the unit sphere, then for any  $\mathbf{d} \in \mathcal{B}(\mathbf{0}_n, 1)$ , there exists a subsequence of indices  $\mathcal{K}_{\rho}^{\mathbf{d}} \subseteq \mathcal{K}_{\rho}^{\mathbf{x}}$  such that  $\lim_{\substack{k \rightarrow +\infty \\ k \in \mathcal{K}_{\rho}^{\mathbf{d}}}} \frac{\mathbf{d}_k}{\|\mathbf{d}_k\|} = \mathbf{d}$ . In the

following, we use  $\mathbf{p}_k = \mathbf{x}_k + \mathbf{d}_k$ . Hence, by Theorem 3.6, we have, for sufficiently large  $k \in \mathcal{K}_{\rho}^{\mathbf{d}}$ ,

$$f(\mathbf{p}_k) - f(\mathbf{x}_k) + \frac{1}{\rho_k} (c^{\text{ext}}(\mathbf{p}_k) - c^{\text{ext}}(\mathbf{x}_k)) \geq \frac{\rho_k}{c^{\text{int}}(\mathbf{p}_k)} (c^{\text{int}}(\mathbf{p}_k) - c^{\text{int}}(\mathbf{x}_k)).$$

Hence,

$$\rho_k \frac{f(\mathbf{p}_k) - f(\mathbf{x}_k)}{\|\mathbf{d}_k\|} + \frac{c^{\text{ext}}(\mathbf{p}_k) - c^{\text{ext}}(\mathbf{x}_k)}{\|\mathbf{d}_k\|} \geq \frac{\rho_k^2}{c^{\text{int}}(\mathbf{p}_k)} \left( \frac{c^{\text{int}}(\mathbf{p}_k) - c^{\text{int}}(\mathbf{x}_k)}{\|\mathbf{d}_k\|} \right).$$

Taking the superior limit for  $k \in \mathcal{K}_\rho^x$ , we obtain:

$$\limsup_{\substack{k \rightarrow \infty \\ k \in \mathcal{K}_\rho^d}} \rho_k \frac{f(\mathbf{p}_k) - f(\mathbf{x}_k)}{\|\mathbf{d}_k\|} + \frac{c^{\text{ext}}(\mathbf{p}_k) - c^{\text{ext}}(\mathbf{x}_k)}{\|\mathbf{d}_k\|} \geq \limsup_{\substack{k \rightarrow \infty \\ k \in \mathcal{K}_\rho^d}} \frac{\rho_k^2}{c^{\text{int}}(\mathbf{p}_k)} \left( \frac{c^{\text{int}}(\mathbf{p}_k) - c^{\text{int}}(\mathbf{x}_k)}{\|\mathbf{d}_k\|} \right). \quad (14)$$

Since  $\bar{\mathbf{x}}$  is an end-path point, we have  $\{\mathbf{x}_k\}_{k \in \mathcal{K}_\rho^d} \rightarrow \bar{\mathbf{x}}$ . Also, since  $\|\mathbf{d}_k\| \leq \Delta_k$  and  $\{\Delta_k\}_{k \in \mathcal{K}_\rho^d} \rightarrow 0$  (by Theorem 3.4), it follows that  $\{\mathbf{p}_k\}_{k \in \mathcal{K}_\rho^x} \rightarrow \bar{\mathbf{x}}$ . Using  $\{\rho_k\}_{k \in \mathcal{K}_\rho} \rightarrow 0$ ,  $c^{\text{int}}(\bar{\mathbf{x}}) < 0$ , and the Lipschitz continuity of  $c^{\text{int}}(\cdot)$  near  $\bar{\mathbf{x}}$  (as it satisfies the FCQ), we can write

$$\lim_{k \in \mathcal{K}_\rho^d} \frac{\rho_k^2}{c^{\text{int}}(\mathbf{p}_k)} = 0 \quad \text{and} \quad \left\{ \frac{c^{\text{int}}(\mathbf{p}_k) - c^{\text{int}}(\mathbf{x}_k)}{\|\mathbf{d}_k\|} \right\} \text{ is bounded.}$$

Hence, the right-hand side of (14) equals zero.

From  $\{\rho_k\}_{k \in \mathcal{K}_\rho} \rightarrow 0$ , and Lipschitz continuity of  $f(\cdot)$  near  $\bar{\mathbf{x}}$ , it follows that

$$\limsup_{\substack{k \rightarrow \infty \\ k \in \mathcal{K}_\rho^d}} \left[ \rho_k \frac{f(\mathbf{p}_k) - f(\mathbf{x}_k)}{\|\mathbf{d}_k\|} + \frac{c^{\text{ext}}(\mathbf{p}_k) - c^{\text{ext}}(\mathbf{x}_k)}{\|\mathbf{d}_k\|} \right] = \limsup_{\substack{k \rightarrow \infty \\ k \in \mathcal{K}_\rho^d}} \frac{c^{\text{ext}}(\mathbf{p}_k) - c^{\text{ext}}(\mathbf{x}_k)}{\|\mathbf{d}_k\|} \geq 0.$$

This implies that  $(c^{\text{ext}})^\circ(\bar{\mathbf{x}}; \mathbf{d}) \geq 0$  for all  $\mathbf{d} \in \mathcal{B}(\mathbf{0}_n, 1)$ , which leads to

$$(c^{\text{ext}})^\circ(\bar{\mathbf{x}}; \mathbf{d}) \geq 0, \quad \forall \mathbf{d} \in \mathbb{R}^n. \quad (15)$$

Since  $\bar{\mathbf{x}}$  satisfies the FCQ, if  $c^{\text{ext}}(\bar{\mathbf{x}}) > 0$ , then  $\bar{\mathbf{x}} \notin \Omega^{\text{ext}}$ , and there would exist a direction  $\mathbf{d} \in \mathcal{B}(\mathbf{0}_n, 1)$  such that  $(c^{\text{ext}})^\circ(\bar{\mathbf{x}}; \mathbf{d}) < 0$ , which contradicts (15). Therefore,  $c^{\text{ext}}(\bar{\mathbf{x}}) = 0$ , and  $\bar{\mathbf{x}}$  is feasible for problem  $(P_0)$ .

**Case (ii)  $c^{\text{int}}(\bar{\mathbf{x}}) = 0$ :** By contradiction, assume that  $\bar{\mathbf{x}} \notin \Omega^{\text{ext}}$ . Since  $\bar{\mathbf{x}}$  satisfies the FCQ, then there exists a direction  $\mathbf{d}$  such that  $(c^{\text{int}})^\circ(\bar{\mathbf{x}}; \mathbf{d}) < 0$  and  $(c^{\text{ext}})^\circ(\bar{\mathbf{x}}; \mathbf{d}) < 0$ .

Because  $(c^{\text{int}})^\circ(\bar{\mathbf{x}}; \mathbf{d}) < 0$ , Theorem 3.9 implies that  $\mathbf{d}$  is a path-refining direction. Hence, by Theorem 3.10, we obtain  $(c^{\text{ext}})^\circ(\bar{\mathbf{x}}; \mathbf{d}) \geq 0$  which contradicts  $(c^{\text{ext}})^\circ(\bar{\mathbf{x}}; \mathbf{d}) < 0$ . Therefore,  $\bar{\mathbf{x}} \in \Omega^{\text{ext}}$ , and thus  $\bar{\mathbf{x}}$  is feasible for problem  $(P_0)$ .  $\square$

### 3.3 Stationarity

This subsection uses the Clarke subdifferential, a generalization of the gradient for non-differentiable functions, to provide stationarity properties on the sequence of iterates generated by Algorithm 1. First, a notion of stationarity for nonsmooth constrained problems is introduced. Then a constraint qualification is presented under which the end-path points are proved to be stationary.

The stationarity results hold when the problem does not present equality constraints. This is not surprising as the additional theoretical complexity caused by the presence of nonsmooth equality constraints is well-known for mathematical programming problems, especially when barrier or penalty techniques are used to solve them [28, 46]. Stationarity results of the Penalty-Interior Point Method in presence of equality constraints are proposed in [12, 22, 20].

The Clarke subdifferential of a function is defined as follows.

**Definition 3.15** (Clarke’s Subdifferential [24]). *For any  $\mathbf{x} \in \mathbb{R}^n$ , Clarke’s subdifferential of  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  at  $\mathbf{x}$  is defined as:*

$$\partial f(\mathbf{x}) := \{\mathbf{v} \in \mathbb{R}^n \mid f^\circ(\mathbf{x}; \mathbf{d}) \geq \mathbf{v}^\top \mathbf{d}, \text{ for all } \mathbf{d} \in \mathbb{R}^n\}.$$

When  $f(\cdot)$  is Lipschitz near a point  $\mathbf{x}$ , then  $\partial f(\mathbf{x})$  is nonempty and compact.

We provide a constraint qualification variant, similar to [5, Assumption A3], for our first-order stationarity analysis.

**Definition 3.16** (Stationarity Constraint Qualification). *A point  $\mathbf{x} \in \Omega$  is said to satisfy the stationarity constraint qualification (SCQ) for problem  $(P_0)$  with  $p = 0$ , if the following conditions hold:*

- (a)  $c^{\text{int}}(\cdot)$  and  $c^{\text{ext}}(\cdot)$  are Lipschitz continuous near  $\mathbf{x}$ ;
- (b)  $\forall \mathbf{d} \in \mathcal{T}_\Omega^H(\mathbf{x})$ ,  $(c^{\text{int}})^\circ(\mathbf{x}; \mathbf{d}) < 0$ ;
- (c)  $\forall \mathbf{d} \in \mathcal{T}_\Omega^H(\mathbf{x})$ , there exists  $\epsilon > 0$  such that  $(c^{\text{ext}})^\circ(\mathbf{y}; \mathbf{d}) < 0$ , for any  $\mathbf{y} \in \mathcal{B}(\mathbf{x}, \epsilon) \setminus \Omega^{\text{ext}}$ .

Theorem 3.16 requires that at  $\mathbf{x}$ , the Clarke directional derivative of  $c^{\text{int}}$  in any hyper-tangent direction is strictly negative; and for the exterior constraints, the Clarke directional derivative of  $c^{\text{ext}}$  is strictly negative for all points outside  $\Omega^{\text{ext}}$  in a neighbourhood around  $\mathbf{x}$ .

As the algorithm progresses and the penalty/barrier parameters are driven to their limits, any accumulation point (end-path point) must be feasible with respect to the original constraints. The SCQ rules out pathological cases where the algorithm converges to a point on the boundary of the feasible set, without generating strictly feasible points, due to the lack of descent directions for the constraint violations (e.g., [5, Appendix A]).

The SCQ ensures that, locally, the feasible region is not “tangentially degenerate”—that is, there are always directions that can strictly decrease any active constraint violation. This is analogous to the classical MFCQ, but adapted to the nonsmooth, aggregated, and split-barrier/penalty structure of our problem. The SCQ is incompatible with equality constraints, as it implicitly implies that  $\mathcal{T}_\Omega^H(\mathbf{x})$  is nonempty: the hypertangent cone is necessarily empty when the dimension of the feasible region is strictly less than  $n$ .

The next result is not related directly to Algorithm 1, but it is useful for our stationarity analysis.

**Proposition 3.17** ([24, Theorem 2.3.7]). *Let  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  and let  $c : \mathbb{R}^n \rightarrow \mathbb{R}$  be Lipschitz continuous in an open set containing the line segment  $[\mathbf{x}, \mathbf{y}]$ . Then, there exists  $\mathbf{u} \in (\mathbf{x}, \mathbf{y})$  and  $\boldsymbol{\xi} \in \partial c(\mathbf{u})$  such that*

$$c(\mathbf{x}) - c(\mathbf{y}) = \boldsymbol{\xi}^\top (\mathbf{x} - \mathbf{y}). \tag{16}$$

The following proposition is similar to a result with the progressive barrier [4] technique used to handle inequality constraints.

**Proposition 3.18.** *Let  $\mathbf{x} \in \Omega$  be a point that satisfies the SCQ and let  $\mathbf{d} \in \mathcal{T}_\Omega^H(\mathbf{x})$ . Then, there exists  $\epsilon > 0$  such that, for any  $\mathbf{y} \in \mathcal{B}(\mathbf{x}, \epsilon)$  and  $\mathbf{w} \in \mathcal{B}(\mathbf{d}, \epsilon)$ , the following holds for all  $t \in (0, \epsilon)$*

$$c^{\text{int}}(\mathbf{y} + t\mathbf{w}) \leq c^{\text{int}}(\mathbf{y}), \quad (17a)$$

$$c^{\text{ext}}(\mathbf{y} + t\mathbf{w}) \leq c^{\text{ext}}(\mathbf{y}). \quad (17b)$$

*Proof.* Throughout,  $c^{\text{int}}$  and  $c^{\text{ext}}$  are locally Lipschitz near  $\mathbf{x}$  by condition (a) of the SCQ, hence Clarke's generalized directional derivatives  $(c^{\text{int}})^\circ(\cdot; \cdot)$  and  $(c^{\text{ext}})^\circ(\cdot; \cdot)$  are well-defined. The proof of (17a) follows using the Lipschitz continuity of  $c^{\text{int}}$  near  $\mathbf{x}$  and condition (b) of SCQ, i.e.,  $(c^{\text{int}})^\circ(\mathbf{x}, \mathbf{d}) < 0$  for all  $\mathbf{d} \in \mathcal{T}_\Omega^H(\mathbf{x})$ . Regarding (17b), let us suppose, by contradiction, that for any  $\epsilon > 0$ , there exist  $\mathbf{y}_\epsilon \in \mathcal{B}(\mathbf{x}, \epsilon)$ ,  $\mathbf{w}_\epsilon \in \mathcal{B}(\mathbf{d}, \epsilon)$  and  $t_\epsilon \in (0, \epsilon)$  such that

$$c^{\text{ext}}(\mathbf{y}_\epsilon + t_\epsilon \mathbf{w}_\epsilon) > c^{\text{ext}}(\mathbf{y}_\epsilon). \quad (18)$$

If  $c^{\text{ext}}(\mathbf{y}_\epsilon) = 0$ , since  $\mathbf{d} \in \mathcal{T}_\Omega^H(\mathbf{x})$ , by the definition of  $\mathcal{T}_\Omega^H(\mathbf{x})$ , for small enough  $\epsilon$  it holds  $\mathbf{y}_\epsilon + t_\epsilon \mathbf{w}_\epsilon \in \Omega$ , which implies  $c^{\text{ext}}(\mathbf{y}_\epsilon + t_\epsilon \mathbf{w}_\epsilon) = 0$ , and hence contradicts (18).

If  $c^{\text{ext}}(\mathbf{y}_\epsilon) > 0$ , since  $\mathcal{T}_\Omega^H(\mathbf{x})$  is an open convex set, for small enough  $\epsilon$ , the vector  $\mathbf{w}_\epsilon \in \mathcal{T}_\Omega^H(\mathbf{x})$ , and the line segment  $[\mathbf{y}_\epsilon, \mathbf{y}_\epsilon + t_\epsilon \mathbf{w}_\epsilon] \subset \mathcal{B}(\mathbf{x}, \epsilon)$ . Therefore, by condition (c) of SCQ,  $(c^{\text{ext}})^\circ(\mathbf{u}; \mathbf{w}_\epsilon) < 0$  for all  $\mathbf{u} \in [\mathbf{y}_\epsilon, \mathbf{y}_\epsilon + t_\epsilon \mathbf{w}_\epsilon]$ . Using Theorem 3.17 and (18), there exists  $\mathbf{u}_\epsilon \in [\mathbf{y}_\epsilon, \mathbf{y}_\epsilon + t_\epsilon \mathbf{w}_\epsilon]$  and  $\xi_\epsilon \in \partial c^{\text{ext}}(\mathbf{u}_\epsilon)$  such that

$$t_\epsilon \mathbf{w}_\epsilon^\top \xi_\epsilon = c^{\text{ext}}(\mathbf{y}_\epsilon + t_\epsilon \mathbf{w}_\epsilon) - c^{\text{ext}}(\mathbf{y}_\epsilon) > 0.$$

Hence, by Theorem 3.15 of  $\partial c^{\text{ext}}(\mathbf{u}_\epsilon)$ , it follows that  $(c^{\text{ext}})^\circ(\mathbf{u}; \mathbf{w}_\epsilon) \geq \mathbf{w}_\epsilon^\top \xi_\epsilon > 0$ , again, a contradiction. Hence, (17b) holds.  $\square$

Next is our final result.

**Theorem 3.19.** *Consider problem  $(P_0)$  with  $p = 0$ . Let Assumption 1 hold and  $\{\mathbf{x}_k\}$  be the sequence of iterates generated by Algorithm 1. Let  $\{\mathbf{x}_k\}_{\mathcal{K}_\rho}$  be a path-following subsequence of  $\{\mathbf{x}_k\}$  and consider  $\mathcal{K}_\rho^x \subseteq \mathcal{K}_\rho$  to be the index set of a relative end-path subsequence, with end-path point  $\bar{\mathbf{x}}$ . Assume that  $f$  is Lipschitz continuous around  $\bar{\mathbf{x}}$ , that the sequence of sets of poll directions  $\{\mathcal{D}_k\}$  is dense in the unit sphere, and that  $\bar{\mathbf{x}}$  satisfies the SCQ. Then,  $f^\circ(\bar{\mathbf{x}}; \mathbf{d}) \geq 0$  for all  $\mathbf{d} \in \mathcal{T}_\Omega^H(\bar{\mathbf{x}})$ .*

*Proof.* Since the sets  $\{\mathcal{D}_k\}$  are dense in the unit sphere, for any  $\mathbf{d} \in \mathcal{B}(\mathbf{0}_n, 1)$ , there exists a subsequence of  $\mathcal{K}_\rho^d \subseteq \mathcal{K}_\rho^x$  such that  $\lim_{\substack{k \rightarrow +\infty \\ k \in \mathcal{K}_\rho^d}} \frac{\mathbf{d}_k}{\|\mathbf{d}_k\|} = \mathbf{d}$ . In the following, we use  $\mathbf{p}_k = \mathbf{x}_k + \mathbf{d}_k$ .

Theorem 3.6 ensures that for sufficiently large  $k \in \mathcal{K}_\rho^d$ ,

$$\frac{f(\mathbf{p}_k) - f(\mathbf{x}_k)}{\|\mathbf{d}_k\|} \geq \frac{\rho_k}{c^{\text{int}}(\mathbf{p}_k)} \left( \frac{c^{\text{int}}(\mathbf{p}_k) - c^{\text{int}}(\mathbf{x}_k)}{\|\mathbf{d}_k\|} \right) - \frac{1}{\rho_k} \left( \frac{c^{\text{ext}}(\mathbf{p}_k) - c^{\text{ext}}(\mathbf{x}_k)}{\|\mathbf{d}_k\|} \right). \quad (19)$$

Since  $\{\mathbf{x}_k\}_{k \in \mathcal{K}_\rho^d} \rightarrow \bar{\mathbf{x}}$ , and, for all  $\mathbf{d} \in \mathcal{T}_\Omega^H(\bar{\mathbf{x}})$ , there exists a subsequence of  $\overline{\mathcal{K}}_\rho^d \subseteq \mathcal{K}_\rho^d$  such that  $\left\{ \frac{\mathbf{d}_k}{\|\mathbf{d}_k\|} \right\}_{k \in \overline{\mathcal{K}}_\rho^d} \rightarrow \mathbf{d}$ . Thus, as  $\bar{\mathbf{x}}$  satisfies SCQ, Theorem 3.18 ensures that (17b) and (17a) hold by setting  $\mathbf{w} = \mathbf{d}_k / \|\mathbf{d}_k\|$ ,  $t = \|\mathbf{d}_k\|$  and  $\mathbf{y} = \mathbf{x}_k$  for  $k \in \overline{\mathcal{K}}_\rho^d$  sufficiently large. Hence, for all  $\mathbf{d} \in \mathcal{T}_\Omega^H(\bar{\mathbf{x}})$  and sufficiently large  $k \in \overline{\mathcal{K}}_\rho^d$ ,

$$-(c^{\text{ext}}(\mathbf{p}_k) - c^{\text{ext}}(\mathbf{x}_k)) \geq 0 \quad \text{and} \quad \frac{c^{\text{int}}(\mathbf{p}_k) - c^{\text{int}}(\mathbf{x}_k)}{c^{\text{int}}(\mathbf{p}_k)} \geq 0,$$

which implies that (19) leads to  $\frac{f(\mathbf{p}_k) - f(\mathbf{x}_k)}{\|\mathbf{d}_k\|} \geq 0$  for all  $\mathbf{d} \in \mathcal{T}_\Omega^H(\bar{\mathbf{x}})$  and  $k \in \overline{\mathcal{K}}_\rho^d$  sufficiently large. Hence, we deduce that  $f^\circ(\bar{\mathbf{x}}; \mathbf{d}) \geq 0$  for all  $\mathbf{d} \in \mathcal{T}_\Omega^H(\bar{\mathbf{x}})$ .  $\square$

## 4 Practical implementation

This section details some practical implementation aspects.

### 4.1 Coding algorithm

All MADS variants in this work are implemented in NOMAD 4 [13] with the default granular mesh called GMesh. In this variant of the mesh, the frame size  $\Delta_k$  is updated using **increase** or **decrease** functions instead of being multiplied by  $\theta_\Delta$  or  $\theta_\Delta^{-1}$ . The convergence properties of MADS with GMesh, the updating functions and the initialization of  $\Delta_0$  are detailed in [15].

The MADS Penalty-Interior Point (MADS-PIP) algorithm is coded in a development version forked from release 4.5.2<sup>1</sup>. In the code, the MADS-PIP algorithm class derives from the MADS progressive barrier (MADS-PB) algorithm class in which the objective function is changed, the infeasibility measure is ignored and the condition of **Step 3** is tested at each unsuccessful iteration. Coding this new algorithm relies on the NOMAD 4 capability to dynamically change at each iteration the computation of the objective function and infeasibility measure from all blackbox problem outputs. Once the partition sets  $\mathcal{G}^{\text{int}}$  and  $\mathcal{G}^{\text{ext}}$  have been identified, the objective function values can be computed as the merit function  $z(\mathbf{x}_k, \rho_k)$  of Theorem 2.1 and the infeasibility measure can be fixed to zero. The penalty parameter  $\rho$  and the partition sets can be updated periodically as detailed in the next section.

### 4.2 Switches and numerical choices

As suggested in Section 2, the PIP strategy can benefit from dynamical updates of the constraints index partition sets  $\mathcal{G}^{\text{int}}$  and  $\mathcal{G}^{\text{ext}}$ . We propose a slight modification of **Step 4** of Algorithm 1, as follows.

The update in **Step 4** consists of verifying whether the inequality constraints contributing to the exterior violation function are strictly feasible at the new incumbent solution. That is, at each iteration  $k$ , the algorithm checks the condition  $g_\ell(\mathbf{x}_{k+1}) < 0$  for each  $\ell \in \mathcal{G}^{\text{ext}}$ .

---

<sup>1</sup><https://github.com/bbopt/nomad>

---

▷ **Step 4: Stopping criterion check** ◀

if some stopping criterion is satisfied **then**

**Stop**

else

    Set  $\mathcal{G}^{\text{int}} \leftarrow \mathcal{G}^{\text{int}} \cup \{\ell\}$  and  $\mathcal{G}^{\text{ext}} \leftarrow \mathcal{G}^{\text{ext}} \setminus \{\ell\}$  for each  $\ell \in \mathcal{G}^{\text{ext}}$  for which  $g_\ell(\mathbf{x}_{k+1}) < 0$

$k \leftarrow k + 1$  and go to **Step 0**

---

If the latter holds for an index  $\bar{\ell}$ , then  $\bar{\ell}$  is removed from  $\mathcal{G}^{\text{ext}}$  and assigned to  $\mathcal{G}^{\text{int}}$ , hence, the inequality constraint is moved into the interior violation function. Since the number of inequality constraints is finite, the update in **Step 4** can be performed only a finite number of times. When the last of such updates occurs, the algorithm follows the theoretical analysis of the previous section. Finally, note that at unsuccessful iterations it holds  $\mathbf{x}_{k+1} = \mathbf{x}_k$ ; therefore, the constraint index partition sets can only change at successful iterations. It follows that the penalty parameter and the constraint index partition sets cannot be updated during the same iteration.

In the computational experiments, the rule in the new step of the algorithm has been implemented considering a tolerance, that is

$$g_\ell(\mathbf{x}_{k+1}) \leq -\varepsilon^{\text{ext}},$$

where  $\varepsilon^{\text{ext}}$  was set equal to  $10^{-14}$ . The tolerance allows for a choice that is more robust to cases where the points may be extremely close to the boundary of the feasible region and machine precision might cause inconsistencies.

The implementation of the PIP strategy comes with non-trivial numerical choices. First, we consider the numerical issue of having constraint functions that are not well-scaled with the objective function. To address this problem, the logarithmic and penalty terms of the merit function (see Theorem 2.1) are multiplied by scaling constants  $b^{\text{int}}$  and  $b^{\text{ext}}$ ; i.e., the merit function can be written as

$$z(\mathbf{x}; \rho) = f(\mathbf{x}) - b^{\text{int}} \rho \log(-c^{\text{int}}(\mathbf{x})) + \frac{b^{\text{ext}}}{\rho} c^{\text{ext}}(\mathbf{x}).$$

Note that such constants do not affect the theoretical properties of the algorithm. After some preliminary tests, the choice of  $b^{\text{int}}$  did not seem to affect the performance of the algorithm significantly; hence, it is set equal to 1. The remaining constant is set using the following rule,

$$b^{\text{ext}} = \begin{cases} 1 & \text{if } f(\mathbf{x}_0) = 0, \\ \max\{1, 10^{\lfloor \log_{10}(|f(\mathbf{x}_0)|) \rfloor}\} & \text{otherwise.} \end{cases}$$

The idea behind the latter formula is to increase the effect of the penalty term to the same order of magnitude as the value of the objective function at the initial point. This allows the objective function and the exterior constraint function to have comparable contributions to the merit function value. Note that when the order of magnitude of  $f(\mathbf{x}_0)$  is smaller than 1, the constant  $b^{\text{ext}}$  is set to 1. Setting  $b^{\text{ext}}$  as a power of 10 is due to faster and more robust numerical computation. Finally, the penalty-barrier parameter  $\rho$  is initialized as  $\rho_0 = 10^{-1}$ .

When an iteration  $k$  is unsuccessful, the algorithm performs the penalty parameter update step. For the practical implementation, the updating criterion is relaxed by multiplying both terms by positive constants  $b_\rho$  and  $b_c$ . The penalty-barrier parameter  $\rho_k$  is reduced whenever

$$\Delta_{k+1} \leq \min\{b_\rho \rho_k^\beta, b_c [\phi^{\text{prox}}(\mathbf{x}_k)]^2\},$$

where  $\beta = 1 + 10^{-9}$ ,  $b_\rho = 10$ , and  $b_c = 10^{10}$ . For values of  $\phi^{\text{prox}}(\mathbf{x}_k)$  on the order of  $10^{-5}$ , the quantity  $b_c [\phi^{\text{prox}}(\mathbf{x}_k)]^2$  is on the order of 1. This means that the influence of the second term of the criterion is not masked by the scaling constant, it is solely relaxed. Note that this choice is consistent with the theoretical analysis. Finally, the penalty contraction parameter is set as  $\theta_\rho = 10^{-2}$ .

We conclude this section by describing a strategy to update the current incumbent at each iteration when the unconstrained subproblem is updated. Whenever the penalty-barrier parameter is updated or the constraint index partition sets are changed, the merit function is modified, and the current incumbent point might not be the best candidate among the previously visited ones. Therefore, the algorithm computes analytically the merit function value for all the points stored during the optimization routine and selects the one with the lowest merit function value.

## 5 Computational experiments

This section proposes a series of computational experiments to compare MADS with Penalty-Interior Point (MADS-PIP) to MADS using the progressive barrier (MADS-PB). For one of the real-life blackbox optimization problems, the comparison considers CMA-ES without parameter tuning from the python module PYCMA<sup>2</sup>.

We benchmark algorithms for solving constrained problems by examining their data profiles within a selected blackbox evaluation budget. In this work, we compute the data profile function  $d_a(k)$  of [11, Definition 3.3] to measure the portion of problem instances that an algorithm  $\tau$ -solves with a precision  $\tau$  within  $k$  groups of  $n + 1$  evaluations. The best known objective function value  $f^*$  considered here is the lowest feasible value achieved by any algorithm on a problem instance. A problem instance combines a test problem with an initial point and a fixed random seed.

Since MADS-PB and MADS-PIP handle constraints in fundamentally different ways, we also examine their efficiency in obtaining a first feasible point starting from an infeasible one. In what follows, when deemed relevant, data profiles show the portion of problem instances for which a feasible point was found within  $k$  groups of  $n + 1$  evaluations.

For problems with equality constraints, a strict feasibility cannot be obtained during optimization. During the benchmarking process, a relaxed feasibility criterion is considered. After some preliminary testing, a point is considered feasible if  $|h_j(\mathbf{x})| < 10^{-8}$  for each constraint index  $j$ .

---

<sup>2</sup><https://github.com/CMA-ES/pycma>

## 5.1 Settings

In this work, all MADS variants used the Ortho 2N polling directions [1]. Several search methods are enabled by default for the search step. In this work, the default Nelder-Mead search method [16] is disabled to focus solely on MADS. For the comparisons, we have considered the effect of enabling the quadratic model search method [25] or not, noted MADS-PIP QMS and MADS-PB QMS in what follows. Quadratic models are built for each output individually before any combination into a non-quadratic surrogate of the merit function  $z$  (MADS-PIP) or the infeasibility measure (MADS-PB). The search step solves a surrogate subproblem to find a new candidate point within a box whose side lengths are proportional to the frame size (typically four times larger) and centred on the incumbent point  $\mathbf{x}_k$ . The surrogate subproblem is solved with the same algorithm as the main optimization problem  $P_0$ . But, for MADS-PIP QMS, during the subproblem iterations the penalty parameters  $\rho_k$  are fixed to values taken from the main problem iteration.

The code of MADS progressive barrier was also adapted to handle equality constraints. The infeasibility of each equality constraint  $|h_j(\mathbf{x})|$  is added into the overall infeasibility measure when above a selected feasibility threshold. In other words, we have considered that if  $|h_j(\mathbf{x})| < \epsilon$  the constraint can be relaxed and not included in the infeasibility measure. If not, the quantity  $|h_j(\mathbf{x})|^2$  is added into the infeasibility measure. Relaxing feasibility must be done with care because it alters a problem’s solution. Some testing was done with several small  $\epsilon$ s. To be consistent with the choice made previously for benchmarking,  $\epsilon = 10^{-8}$  was chosen as a good compromise for all the tests on problems with equality constraints.

## 5.2 Inequality constrained optimization

This section describes the computational experiments conducted on three inequality-constrained test sets.

### Results on problems from the literature

Table 1 lists fourteen inequality constrained problems selected from the literature. The number of variables  $n$  varies from 2 to 13. The number of general inequality constraints  $m$  ranges from 1 to 15 and there are no equality constraint. Feasible and infeasible starting points are considered. A problem instance is defined by a single initial point and a random seed. Each problem with a selected initial point is run 10 times with different seeds for a total of 780 runs for each algorithm tested.

Figure 1 shows the obtained results on the selection of problems from the literature. The results show that with or without the quadratic model, MADS-PIP performs slightly better than MADS-PB. On this test set, the impact of the models is much more important than the choice in the strategy (PIP versus PB) to handle the constraints.

### Results on a real blackbox: The solar6 problem

solar is a blackbox optimization benchmarking framework [2]. The solar6 instance is considered here for benchmarking. It is a simulator of a thermal power plant using the molten salt cycle

Problem	$n$	$m$	$\# x_0$	Problem	$n$	$m$	$\# x_0$
CHENWANG_F2[23]	8	6	11	MEZMONTES[44]	2	2	1
CHENWANG_F3[23]	10	8	10	OPTENG_RBF[33]	3	4	1
CRESCENT[5]	10	2	1	PENTAGON[41]	6	15	1
DISK[5]	10	1	1	SNAKE[5]	2	2	1
G210[8]	10	2	1	SPRING[47]	3	4	10
MAD6[41]	5	7	10	TAOWANG_F2[49]	7	4	10
MDO*	10	10	10	ZHAOWANG_F5[50]	13	9	10

\*MDO problem with the same analytical expressions for disciplines than the MDO problem in Section 5.3 but with a different approach to solve the multidisciplinary analyses. Source code available at [https://github.com/bbopt/aircraft\\_range](https://github.com/bbopt/aircraft_range).

Table 1: Description of a set of analytical problems from the literature without any equality constraints.

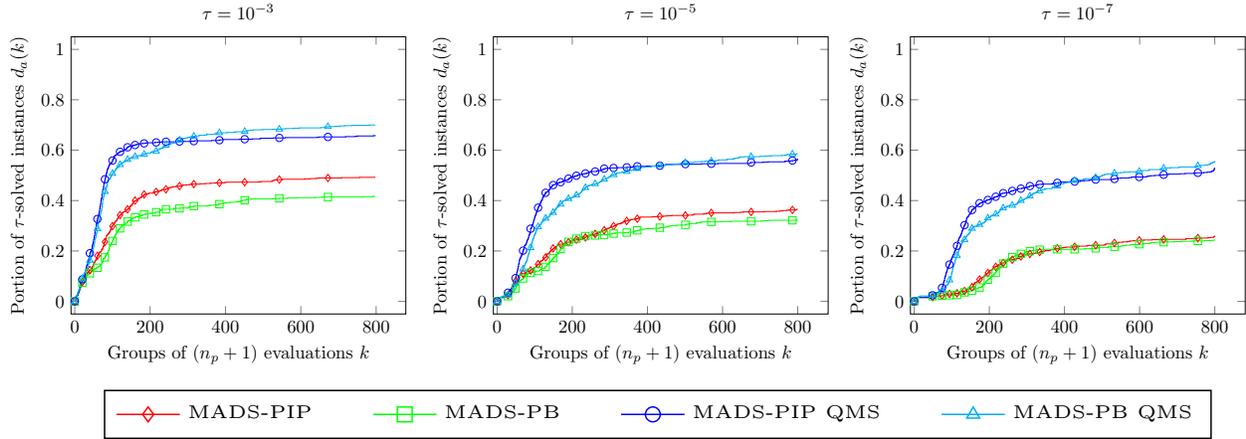


Figure 1: Data profiles on a selection of analytical problems from the literature.

and power block models. The objective is to minimize the cost of the thermal storage units so that the power plant is able to sustain a 120MW electrical daily power output. The heliostat field is not simulated in `solar6`. The problem has 5 variables, 6 general inequality constraints and no equality constraints. The thirty different initial points proposed are those of the `solar` repository<sup>3</sup> and a single seed per optimization is considered. Hence, each algorithm is run on 30 problem instances.

The average optimization time for each `solar` problem instance was approximately 14 hours on a 12th Gen Intel(R) Core(TM) i7 processor. Each of the 1,500 blackbox evaluations within an optimization run was executed on a single core, while multiple (at most 12) optimization runs were performed in parallel.

Figure 2 includes CMA-ES to the comparison. In this problem, introducing quadratic models seems to degrade the algorithm’s overall performance, and the induced lack of smoothness adversely affects CMA-ES. MADS-PIP QMS outperforms MADS-PB QMS, but without models MADS-PIP is preferable to MADS-PB only when  $\tau$  is small.

On this first real-life blackbox problem, the difference in performance between MADS-PB

<sup>3</sup><https://github.com/bbopt/solar>

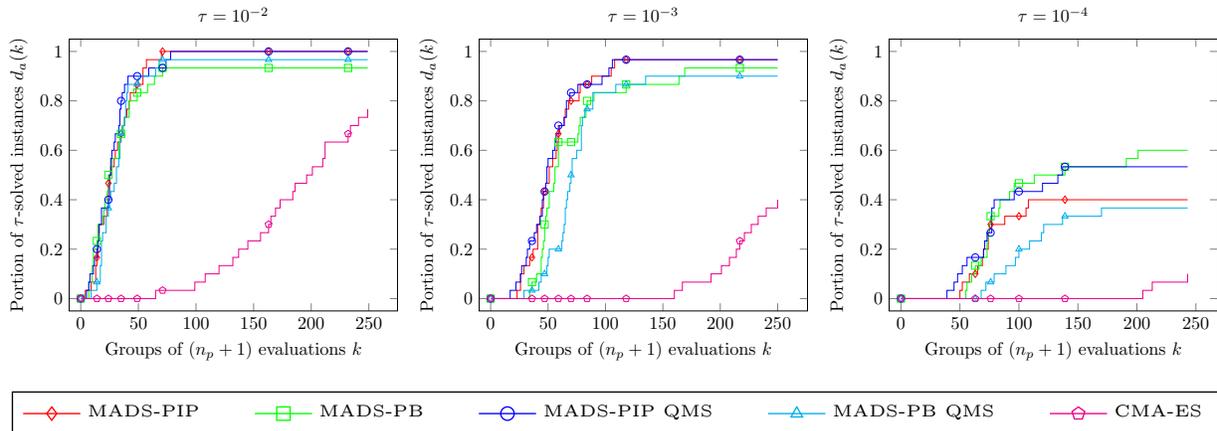


Figure 2: Data profiles obtained on 30 inequality constrained problem instances of `solar6`.

and MADS-PIP is not as pronounced as in Section 5.3.

### 5.3 Equality and inequality constrained optimization

This section analyzes two sets of optimization problems constrained by equalities and inequalities.

#### Results on CUTEst problems

Table 2 lists the 25 CUTEst problems with both equality and inequality constraints. The problem dimensions range from 4 to 41, the number of inequality constraints varies between 3 and 90, and there are up to 21 equality constraints. The problem instances are generated using five different random seeds.

Problem	$n$	$m$	$p$	Problem	$n$	$m$	$p$	Problem	$n$	$m$	$p$
ALLINQP	24	18	3	GAUSSELM	29	36	14	LOADBAL	31	31	11
ANTWERP	27	10	8	HADAMARD	37	93	21	RES	20	14	12
BLOCKQP1	35	16	15	HS74	4	5	3	SYNTHES2	11	15	1
BLOCKQP2	35	16	15	HS75	4	5	3	SYNTHES3	17	23	2
BLOCKQP3	35	16	15	HS114	10	11	3	TENBARS4	18	9	8
BLOCKQP4	35	16	15	JANNSON3	30	3	1	TRUSPYR2	11	11	3
BLOCKQP5	35	16	15	KISSING	37	78	12	ZIGZAG	28	30	20
EG3	31	90	1	LIPPERT1	41	80	16				
ERRINBAR	18	9	8	LIPPERT2	41	80	16				

Table 2: Set of 25 equality and inequality constrained problems selected from the CUTEst collection. The parameters  $n$ ,  $m$ , and  $p$  denote, respectively, the number of variables, of inequality constraints, and of equality constraints.

Figure 3 compares MADS-PIP and MADS-PB on the 25 CUTEst problems with equality and inequality constraints. The plot on the left of Figure 3 shows the portion of instances for

which the methods generate a feasible point for equality constrained problems. Here MADS-PB is clearly outperformed by MADS-PIP. This is not surprising because MADS-PB is not designed to handle equality constraints. For MADS-PIP, the quadratic model helps to quickly find a feasible point. The two plots on the right show the data profiles. Again, MADS-PIP is superior to MADS-PB, and the use of quadratic models is again shown to be very useful.

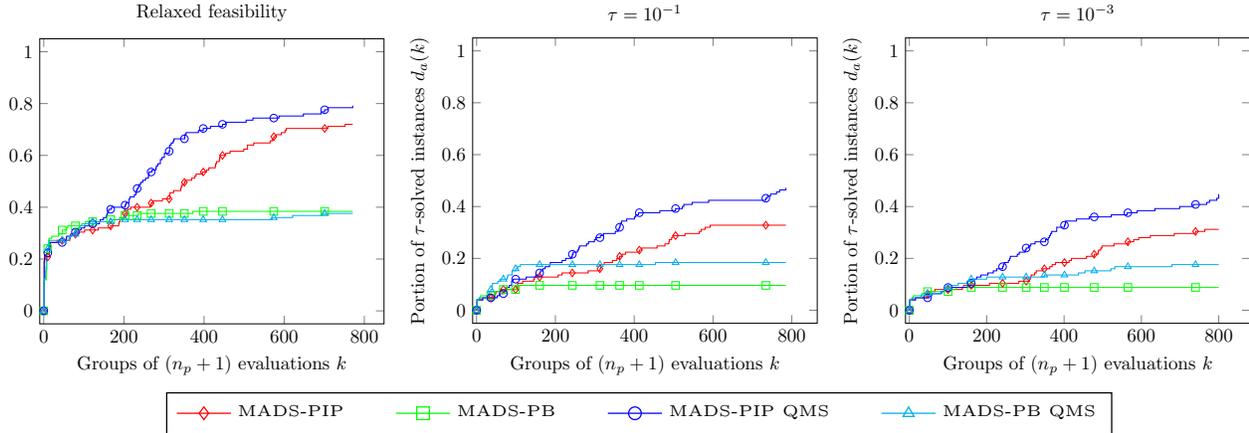


Figure 3: Data profiles on 25 CUTEst analytical problems having equality constraints. The left plot displays the portion of feasible instances obtained. Middle and right plots show the portion of  $\tau$ -solved instances for  $\tau = 10^{-1}$  (middle) and  $\tau = 10^{-3}$  (right).

### Results on a real blackbox: The aircraft range MDO problem

The Aircraft Range multidisciplinary design optimization (MDO) problem that has both equality and inequality constraints is also considered for benchmarking. A simulator solves the multidisciplinary analyses (MDA) to predict the operation of a supersonic business jet for a given set of variables. In the simulator, the aerodynamics, structure and propulsion disciplines are modelled by analytical expressions typical for an early conceptual design stage. The analytical expressions given in [48] have been converted to C++ and made available in GitHub<sup>4</sup>. These analytical expressions are fast to evaluate and are regarded as unknown in this blackbox simulator.

We consider the variant of the problem in which the discipline coupling is solved by adding interdisciplinary (coupling) variables and consistency equality constraints to the MDO problem. This type of approach is referred to as “individual discipline feasible” (IDF) [27]. The MDO problem in Section 5.2 uses the same discipline analyses but solves the MDA with an iterative fixed point method that is often referred to as multidisciplinary feasible (MDF). The problem has 13 bounded variables scaled between 0 and 100, 10 general inequality constraints and 3 general equality constraints.

<sup>4</sup>[https://github.com/bbopt/aircraft\\_range](https://github.com/bbopt/aircraft_range)

Ten different initial points and ten random seeds are considered. A problem instance is defined by a single initial point and a random seed. The leftmost plot in Figure 4 shows that again, both MADS-PB variants fail to produce feasible solutions to 70% of the problem instances. MADS-PIP has the opposite behaviour, and MADS-PIP QMS produces a feasible solution to 95% of the problem instances. The performance profiles on the right show that MADS-PB fails on 100% of the instances. MADS-PB QMS is clearly the dominant method.

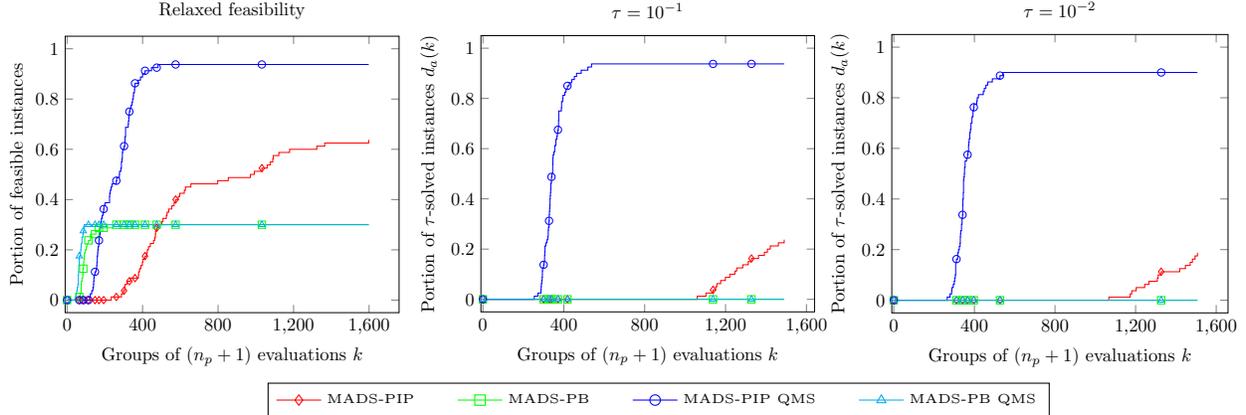


Figure 4: Data profiles on 100 problem instances of Aircraft Range MDO. The left plot displays the portion of feasible instances obtained. Middle and right plots show the portion of  $\tau$ -solved instances for  $\tau = 10^{-1}$  (middle) and  $\tau = 10^{-2}$  (right).

In summary, the results on the real-life test problem suggest that MADS-PB and MADS-PIP are competitive on problems with only inequality constraints. However, as soon as equality constraints are present, MADS-PB fails as it is not designed to tackle such problems. MADS-PIP is the first variant of MADS to handle equality constraints.

## 6 Conclusions

The main purpose of this work is to extend MADS to handle equality and inequality constraints efficiently. The constraints are partitioned into two distinct sets. The interior set comprises certain inequality constraints, while the exterior set consists of the remaining inequality constraints together with all equality constraints. The interior set is managed using a logarithmic barrier based on an aggregated measure of interior violation, whereas the exterior set is treated via an exterior penalty approach.

These components are combined into a merit function, yielding the proposed penalty-interior-point method (MADS-PIP). This construction allows MADS to solve a sequence of unconstrained subproblems while progressively driving the penalty parameter to zero, so that the merit function approximates the original constrained problem.

This work shows that MADS-PIP generates an infinite sequence of path-following updates and that the frame size parameters along these subsequences shrink to zero, ensuring that end-path limit points exist. Under mild Lipschitz continuity and density assumptions, together

with appropriate constraint qualifications for both feasibility (FCQ) and stationarity (SCQ), end-path limit points are shown to be feasible for the original problem and are shown to satisfy first-order Clarke-stationarity conditions for the inequality-constrained case. These results provide rigorous nonsmooth convergence guaranties for the penalty interior-point strategy when combined with the MADS framework.

Furthermore, MADS-PIP was implemented in NOMAD 4 [13], and extensive tests were conducted on both CUTEst and other challenging blackbox problems. These experiments indicate that MADS-PIP is competitive with the progressive barrier approach (MADS-PB) on inequality-only constrained problems and demonstrate a clear advantage when equality constraints are present.

## References

- [1] M.A. Abramson, C. Audet, J.E. Dennis, Jr., and S. Le Digabel. “OrthoMADS: A Deterministic MADS Instance with Orthogonal Directions”. In: *SIAM J. Optim.* 20.2 (2009), pp. 948–966. DOI: 10.1137/080716980. URL: <https://dx.doi.org/10.1137/080716980>.
- [2] N. Andrés-Thió, C. Audet, M. Diago, A.E. Gheribi, S. Le Digabel, X. Lebeuf, M. Lemyre Garneau, and C. Tribes. “Solar: a solar thermal power plant simulator for blackbox optimization benchmarking”. In: *Optim. Eng.* 26.3 (2025), pp. 1815–1861. DOI: 10.1007/s11081-024-09952-x. URL: <https://dx.doi.org/10.1007/s11081-024-09952-x>.
- [3] C. Audet and J.E. Dennis, Jr. “A Pattern Search Filter Method for Nonlinear Programming without Derivatives”. In: *SIAM J. Optim.* 14.4 (2004), pp. 980–1010. DOI: 10.1137/S105262340138983X. URL: <https://dx.doi.org/10.1137/S105262340138983X>.
- [4] C. Audet and J.E. Dennis, Jr. “A Progressive Barrier for Derivative-Free Nonlinear Programming”. In: *SIAM J. Optim.* 20.1 (2009), pp. 445–472. DOI: 10.1137/070692662. URL: <https://dx.doi.org/10.1137/070692662>.
- [5] C. Audet and J.E. Dennis, Jr. “A Progressive Barrier for Derivative-Free Nonlinear Programming”. In: *SIAM J. Optim.* 20.1 (2009), pp. 445–472. DOI: 10.1137/070692662. URL: <https://dx.doi.org/10.1137/070692662>.
- [6] C. Audet and J.E. Dennis, Jr. “Analysis of Generalized Pattern Searches”. In: *SIAM J. Optim.* 13.3 (2003), pp. 889–903. DOI: 10.1137/S1052623400378742. URL: <https://dx.doi.org/10.1137/S1052623400378742>.
- [7] C. Audet and J.E. Dennis, Jr. “Mesh Adaptive Direct Search Algorithms for Constrained Optimization”. In: *SIAM J. Optim.* 17.1 (2006), pp. 188–217. DOI: 10.1137/040603371. URL: <https://dx.doi.org/10.1137/040603371>.
- [8] C. Audet, J.E. Dennis, Jr., and S. Le Digabel. “Parallel Space Decomposition of the Mesh Adaptive Direct Search Algorithm”. In: *SIAM J. Optim.* 19.3 (2008), pp. 1150–1170. DOI: 10.1137/070707518. URL: <https://dx.doi.org/10.1137/070707518>.

- [9] C. Audet, T. Denorme, Y. Diouane, S. Le Digabel, and C. Tribes. *Adaptive direct search algorithms for constrained optimization*. Tech. rep. G-2025-53. Les cahiers du GERAD, 2025. URL: <https://www.gerad.ca/fr/papers/G-2025-53>.
- [10] C. Audet and W. Hare. *Derivative-Free and Blackbox Optimization*. Springer Series in Operations Research and Financial Engineering. Cham, Switzerland: Springer, 2017. DOI: 10.1007/978-3-319-68913-5. URL: <https://dx.doi.org/10.1007/978-3-319-68913-5>.
- [11] C. Audet, W. Hare, and C. Tribes. *Benchmarking constrained, multi-objective and surrogate-assisted derivative-free optimization methods*. Tech. rep. G-2025-36. Les cahiers du GERAD, 2025. URL: <https://www.gerad.ca/en/papers/G-2025-36>.
- [12] C. Audet, S. Le Digabel, and M. Peyrega. “Linear equalities in blackbox optimization”. In: *Comput. Optim. Appl.* 61.1 (2015), pp. 1–23. DOI: 10.1007/s10589-014-9708-2. URL: <https://dx.doi.org/10.1007/s10589-014-9708-2>.
- [13] C. Audet, S. Le Digabel, V. Rochon Montplaisir, and C. Tribes. “Algorithm 1027: NOMAD version 4: Nonlinear optimization with the MADS algorithm”. In: *ACM Trans. Math. Soft.* 48.3 (2022), 35:1–35:22. DOI: 10.1145/3544489. URL: <https://dx.doi.org/10.1145/3544489>.
- [14] C. Audet, S. Le Digabel, and R. Saltet. “Quantifying uncertainty with ensembles of surrogates for blackbox optimization”. In: *Comput. Optim. Appl.* 83 (2022), pp. 29–66. DOI: 10.1007/s10589-022-00381-z. URL: <https://doi.org/10.1007/s10589-022-00381-z>.
- [15] C. Audet, S. Le Digabel, and C. Tribes. “The Mesh Adaptive Direct Search Algorithm for Granular and Discrete Variables”. In: *SIAM J. Optim.* 29.2 (2019), pp. 1164–1189. DOI: 10.1137/18M1175872. URL: <https://dx.doi.org/10.1137/18M1175872>.
- [16] C. Audet and C. Tribes. “Mesh-based Nelder-Mead algorithm for inequality constrained optimization”. In: *Comput. Optim. Appl.* 71.2 (2018), pp. 331–352. DOI: 10.1007/s10589-018-0016-0. URL: <https://link.springer.com/article/10.1007/s10589-018-0016-0>.
- [17] D.P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. New York: Academic Press, 1982.
- [18] D.P. Bertsekas. *Nonlinear Programming*. 2nd ed. Athena Scientific, 1999.
- [19] A. Brilli, A. Cristofari, G. Liuzzi, and S. Lucidi. *Complexity results and active-set identification of a derivative-free method for bound-constrained problems*. 2025. arXiv: 2402.10801 [math.OA]. URL: <https://arxiv.org/abs/2402.10801>.
- [20] A. Brilli, A.L. Custódio, G. Liuzzi, and E.J. Silva. *Nonlinear Derivative-free Constrained Optimization with a Mixed Penalty-Logarithmic Barrier Approach and Direct Search*. 2024. arXiv: 2407.21634 [math.OA]. URL: <https://doi.org/10.48550/arXiv.2407.21634>.

- [21] A. Brilli, M. Kimiaei, G. Liuzzi, and S. Lucidi. “Worst Case Complexity Bounds for Linesearch-Type Derivative-Free Algorithms”. In: *J. Optimiz. Theory App.* 203.1 (2024), pp. 419–454. ISSN: 1573-2878. DOI: 10.1007/s10957-024-02519-x. URL: <https://doi.org/10.1007/s10957-024-02519-x>.
- [22] A. Brilli, G. Liuzzi, and S. Lucidi. “An interior point method for nonlinear constrained derivative-free optimization”. In: *Optim. Methods Softw.* 40.3 (2025), pp. 611–649. DOI: 10.1080/10556788.2025.2453110. eprint: <https://doi.org/10.1080/10556788.2025.2453110>. URL: <https://doi.org/10.1080/10556788.2025.2453110>.
- [23] X. Chen and N. Wang. “Optimization of short-time gasoline blending scheduling problem with a DNA based hybrid genetic algorithm”. In: *Chem. Eng. Process. Process Intensif.* 49.10 (2010), pp. 1076–1083. DOI: 10.1016/j.cep.2010.07.014. URL: <http://dx.doi.org/10.1016/j.cep.2010.07.014>.
- [24] F.H. Clarke. *Optimization and Nonsmooth Analysis*. Society for Industrial and Applied Mathematics, 1990. DOI: 10.1137/1.9781611971309. URL: <https://dx.doi.org/10.1137/1.9781611971309>.
- [25] A.R. Conn and S. Le Digabel. “Use of quadratic models with mesh-adaptive direct search for constrained black box optimization”. In: *Optim. Methods Softw.* 28.1 (2013), pp. 139–158. DOI: 10.1080/10556788.2011.623162. URL: <https://dx.doi.org/10.1080/10556788.2011.623162>.
- [26] A.R. Conn, K. Scheinberg, and L.N. Vicente. “Global convergence of general derivative-free trust-region algorithms to first and second order critical points”. In: *SIAM J. Optim.* 20.1 (2009), pp. 387–415. DOI: 10.1137/060673424. URL: <https://dx.doi.org/10.1137/060673424>.
- [27] E.J. Cramer, J.E. Dennis, Jr., P.D. Frank, R.M. Lewis, and G.R. Shubin. “Problem Formulation for Multidisciplinary Optimization”. In: *SIAM Journal of Optimization* 4.4 (1994), pp. 754–776. DOI: 10.1137/0804044. URL: <https://dx.doi.org/10.1137/0804044>.
- [28] G. Di Pillo and F. Facchinei. “Exact barrier function methods for Lipschitz programs”. In: *Appl. Math. Opt.* 32.1 (1995), pp. 1–31. ISSN: 1432-0606. DOI: 10.1007/BF01189901. URL: <https://doi.org/10.1007/BF01189901>.
- [29] Y. Diouane. “A merit function approach for evolution strategies”. In: *EURO J. Comput. Optim.* 9 (2021), p. 100001. DOI: 10.1016/j.ejco.2020.100001. URL: <https://doi.org/10.1016/j.ejco.2020.100001>.
- [30] G. Fasano, G. Liuzzi, S. Lucidi, and F. Rinaldi. “A Linesearch-Based Derivative-Free Approach for Nonsmooth Constrained Optimization”. In: *SIAM J. Optim.* 24.3 (2014), pp. 959–992. DOI: 10.1137/130940037. URL: <https://dx.doi.org/10.1137/130940037>.
- [31] R. Fletcher and S. Leyffer. “Nonlinear Programming Without a Penalty Function”. In: *Math. Program. Series A*, 91 (2002), pp. 239–269. DOI: 10.1007/s101070100244. URL: <https://dx.doi.org/10.1007/s101070100244>.

- [32] S. Gratton and L.N. Vicente. “A Merit Function Approach for Direct Search”. In: *SIAM J. Optim.* 24.4 (2014), pp. 1980–1998. DOI: 10.1137/130917661. URL: <https://dx.doi.org/10.1137/130917661>.
- [33] S. Kitayama, M. Arakawa, and K. Yamazaki. “Sequential Approximate Optimization using Radial Basis Function network for engineering optimization”. In: *Optim. Eng.* 12.4 (2011), pp. 535–557. DOI: 10.1007/s11081-010-9118-y. URL: <https://dx.doi.org/10.1007/s11081-010-9118-y>.
- [34] S. Le Digabel and S.M. Wild. “A taxonomy of constraints in black-box simulation based optimization”. In: *Optim. Eng.* 25.2 (2024), pp. 1125–1143. DOI: 10.1007/s11081-023-09839-3. URL: <https://dx.doi.org/10.1007/s11081-023-09839-3>.
- [35] R.M. Lewis and V. Torczon. “A Globally Convergent Augmented Lagrangian Pattern Search Algorithm for Optimization With General Constraints and Simple Bounds”. In: *SIAM J. Optim.* 12.4 (2002), pp. 1075–1089. DOI: 10.1137/S1052623498339727. URL: <https://dx.doi.org/10.1137/S1052623498339727>.
- [36] R.M. Lewis and V. Torczon. “Pattern Search Algorithms for Bound Constrained Minimization”. In: *SIAM J. Optim.* 9.4 (1999), pp. 1082–1099. DOI: 10.1137/S1052623496300507. URL: <https://dx.doi.org/10.1137/S1052623496300507>.
- [37] R.M. Lewis and V. Torczon. “Pattern search methods for linearly constrained minimization”. In: *SIAM J. Optim.* 10.3 (2000), pp. 917–941. DOI: <https://doi.org/10.1137/S1052623497331373>.
- [38] G. Liuzzi and S. Lucidi. “A Derivative-Free Algorithm for Inequality Constrained Nonlinear Programming via Smoothing of an  $\ell_\infty$  Penalty Function”. In: *SIAM J. Optim.* 20.1 (2009), pp. 1–29. DOI: 10.1137/070711451. URL: <https://dx.doi.org/10.1137/070711451>.
- [39] G. Liuzzi, S. Lucidi, and M. Sciandrone. “Sequential Penalty Derivative-Free Methods for Nonlinear Constrained Optimization”. In: *SIAM J. Optim.* 20.5 (2010), pp. 2614–2635. DOI: 10.1137/090750639. URL: <https://dx.doi.org/10.1137/090750639>.
- [40] S. Lucidi and M. Sciandrone. “A Derivative-Free Algorithm for Bound Constrained Optimization”. In: *Comput. Optim. Appl.* 21.2 (2002), pp. 119–142. ISSN: 1573-2894. DOI: 10.1023/A:1013735414984. URL: <https://doi.org/10.1023/A:1013735414984>.
- [41] L. Lukšan and J. Vlček. *Test Problems for Nonsmooth Unconstrained and Linearly Constrained Optimization*. Tech. rep. V-798. ICS AS CR, 2000.
- [42] O.L. Mangasarian and S. Fromovitz. “The Fritz John necessary optimality conditions in the presence of equality and inequality constraints”. In: *J. Math. Anal. Appl.* 17.1 (1967), pp. 37–47. ISSN: 0022-247X. DOI: [https://doi.org/10.1016/0022-247X\(67\)90163-1](https://doi.org/10.1016/0022-247X(67)90163-1). URL: <https://www.sciencedirect.com/science/article/pii/0022247X67901631>.
- [43] M.D. McKay, R.J. Beckman, and W.J. Conover. “A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code”. In: *Technometrics* 21.2 (1979), pp. 239–245. URL: <http://www.jstor.org/stable/1268522>.

- [44] E. Mezura-Montes and C.A. Coello. “Useful Infeasible Solutions in Engineering Optimization with Evolutionary Algorithms”. In: *Proceedings of the 4th Mexican International Conference on Advances in Artificial Intelligence*. MICAI’05. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 652–662. DOI: 10.1007/11579427\_66. URL: [https://dx.doi.org/10.1007/11579427\\_66](https://dx.doi.org/10.1007/11579427_66).
- [45] J. Nocedal and S.J. Wright. *Numerical Optimization*. Second. Springer Series in Operations Research and Financial Engineering. Berlin: Springer, 2006. URL: <http://www.springer.com/mathematics/book/978-0-387-30303-1>.
- [46] E. Polak, D.Q. Mayne, and Y. Wardi. “On the Extension of Constrained Optimization Algorithms from Differentiable to Nondifferentiable Problems”. In: *SIAM J. Control Optim.* 21.2 (1983), pp. 179–203. DOI: 10.1137/0321010. eprint: <https://doi.org/10.1137/0321010>. URL: <https://doi.org/10.1137/0321010>.
- [47] J.F. Rodríguez, J.E. Renaud, and L.T. Watson. “Trust Region Augmented Lagrangian Methods for Sequential Response Surface Approximation and Optimization”. In: *J. Mech. Design* 120.1 (1998), pp. 58–66. DOI: 10.1115/1.2826677. URL: <https://dx.doi.org/10.1115/1.2826677>.
- [48] J. Sobieszczanski-Sobieski, J.S. Agte, and R.R. Sandusky, Jr. *Bi-Level Integrated System Synthesis (BLISS)*. Tech. rep. NASA/TM-1998-208715. NASA, Langley Research Center, 1998.
- [49] J. Tao and N. Wang. “DNA Double Helix Based Hybrid GA for the Gasoline Blending Recipe Optimization Problem”. In: *Chem. Eng. Technol.* 31.3 (2008), pp. 440–451. DOI: 10.1002/ceat.200700322. URL: <https://dx.doi.org/10.1002/ceat.200700322>.
- [50] J. Zhao and N. Wang. “A bio-inspired algorithm based on membrane computing and its application to gasoline blending scheduling”. In: *Comput. Chem. Eng.* 35.2 (2011), pp. 272–283. DOI: 10.1016/j.compchemeng.2010.01.008. URL: <http://dx.doi.org/10.1016/j.compchemeng.2010.01.008>.