Nonlinear Derivative-free Constrained Optimization with a Penalty-Interior Point Method and Direct Search

Andrea Brilli * Ana L. Custódio † Giampaolo Liuzzi‡ Everton J. Silva§

Abstract

In this work, we propose the joint use of a mixed penalty-interior point method and direct search, for addressing nonlinearly constrained derivative-free optimization problems. A merit function is considered, wherein the set of nonlinear inequality constraints is divided into two groups: one treated with a logarithmic barrier approach, and another, along with the equality constraints, addressed using a penalization term. This strategy, is adapted and incorporated into a direct search method , enabling the effective handling of general nonlinear constraints. Convergence to KKT-stationary points is established under continuous differentiability assumptions, without requiring any kind of convexity. Using CUTEst test problems, numerical experiments demonstrate the robustness, efficiency, and overall effectiveness of the proposed method, when compared with state-of-the-art solvers.

Keywords: Derivative-free optimization; Constrained optimization; Interior point methods; Direct search.

AMS Classification: 65K05, 90C30, 90C56.

* "Sapienza" University of Rome, Department of Computer Control and Management Engineering "A. Ruberti", Rome, Italy (brilli@diag.uniroma1.it).

†NOVA School of Science and Technology, Center for Mathematics and Applications (NOVA MATH), Campus de Caparica, 2829-516 Caparica, Portugal (alcustodio@fct.unl.pt). This work was funded by national funds through FCT - Fundação para a Ciência e a Tecnologia I.P., under the scope of projects UIDP/00297/2020 and UIDB/00297/2020 (Center for Mathematics and Applications).

‡ "Sapienza" University of Rome, Department of Computer Control and Management Engineering "A. Ruberti", Rome, Italy (liuzzi@diag.uniroma1.it).

§NOVA School of Science and Technology, Center for Mathematics and Applications (NOVA MATH), Campus de Caparica, 2829-516 Caparica, (ejo.silva@campus.fct.unl.pt). This work was funded by national funds through FCT - Fundação para a Ciência e a Tecnologia I.P., under the scope of projects UI/BD/151246/2021, UIDP/00297/2020, and UIDB/00297/2020 (Center for Mathematics and Applications).

1 Introduction

In this work, we consider a derivative-free optimization problem with general constraints (linear and nonlinear), defined by:

$$
\begin{array}{ll}
\min & f(\mathbf{x}) \\
\text{s.t.} & g_\ell(\mathbf{x}) \leq 0 \quad \ell = 1, \ldots, m, \\
& h_j(\mathbf{x}) = 0 \quad j = 1, \ldots, p, \\
& \mathbf{x} \in \mathcal{X},
\end{array}
\tag{P}
$$

where $f : \mathcal{X} \to \mathbb{R}$, $g_\ell : \mathcal{X} \to \mathbb{R}$, $\ell = 1, \ldots, m$, $h_j : \mathcal{X} \to \mathbb{R}$, $j = 1, \ldots, p$, are continuously differentiable in $\mathcal{X}$, and $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$, with $\mathbf{A} \in \mathbb{R}^{q \times n}$ and $\mathbf{b} \in \mathbb{R}^q$. Although, we assume that derivatives of functions f, g_l, and h_j cannot be either calculated or explicitly approximated. This is a common setting in a context of simulation-based optimization, where function evaluations are computed through complex and expensive computational simulations [3, 11].

Constrained derivative-free optimization is not new. In the context of pattern search methods, the early works are co-authored by Lewis and Torczon, first considering bound [25] or linearly constrained problems [26], and after for general nonlinear constraints [24]. When the constraints are only linear, inspired by the work of May [31], procedures were developed that allow to conform the directions to be used by the algorithms to the geometry of the feasible region imposed by the nearby constraints [23, 26, 30], including specific strategies to address the degenerated case [1]. In the presence of nonlinear constraints, augmented Lagrangian approaches have been proposed [24], reducing the problem solution to a sequence of bound constrained minimizations of an augmented Lagrangian function.

In the original presentation of Mesh Adaptive Direct Search (MADS) [7], a generalization of pattern search methods, constraints were addressed with an extreme barrier approach, only evaluating feasible points. If this saves in function evaluations, a very relevant feature for the target problem class, it does not take advantage on the local information obtained about the feasible region. Following the filter approaches proposed for derivative-based optimization [17] and already explored in pattern search methods [4], linear and nonlinear inequalities started to be treated in MADS with the progressive barrier technique [5]. Later, the approach was extended to linear equality constraints [2], by reformulating the optimization problem, possibly reducing the number of original variables.

For directional direct search methods using linesearch, other approaches have been taken to address general nonlinear constraints. These include nonsmooth exact penalty functions [27], where the original nonlinearly constrained problem is converted into the unconstrained or linearly constrained minimization of a nonsmooth exact penalty function. To overcome the limitations associated to this approach, in [29] a sequential penalty approach has been studied, based on the smoothing of the nondifferentiable exact penalty function, including a well-defined strategy for updating the penalty parameter. Recently, in the same algorithmic framework, Brilli *et al.* [10] proposed the use of a merit function that handles inequality constraints by means of a logarithmic barrier approach and equality constraints by considering a penalization term. This approach allows an easy management of relaxable and unrelaxable constraints and avoids the abrupt discontinuities introduced by the extreme barrier approach.

1.1 Our Contribution

In this work, the strategy of [10] is adapted and incorporated into generalized pattern search (GPS). Departing from SID-PSM [13, 14], an implementation of a GPS that uses polynomial models both at the search and the poll steps to improve the numerical performance of the code, LOG-DS is developed, a direct search method able to explicitly address nonlinear constraints by a mixed penalty-logarithmic barrier approach. Section 2 proposes the algorithmic structure. Convergence is established in Section 3. Details of the numerical implementation are provided in Section 4 and numerical results are reported in Section 5, comparing the numerical performance of LOG-DS with state-of-the-art solvers. We summarize our conclusions in Section 6. An appendix completes the paper, including some auxiliary technical results.

1.2 Notation

Throughout this paper, vectors will be written in lowercase boldface (e.g., $\mathbf{v} \in \mathbb{R}^n$, $n \geq 2$) while matrices will be written in uppercase boldface (e.g., $\mathbf{S} \in \mathbb{R}^{n \times p}$). The set of column vectors of a matrix $\mathbf{D}$ will be denoted by $\mathbb{D}$, and more generally sets such as $\mathbb{N}, \mathbb{Q}, \mathbb{R}$ will be denoted by blackboard letters. The set of nonnegative real numbers will be denoted by $\mathbb{R}_+$. Sequences indexed by $\mathbb{N}$ will be denoted by $\{a_k\}_{k \in \mathbb{N}}$ or $\{a_k\}$ in absence of ambiguity. Given a point $\mathbf{x} \in \mathbb{R}^n$ and a set $\mathbf{\Omega} \subset \mathbb{R}^n$, we use the notation $\mathcal{T}_{\mathbf{\Omega}}(\mathbf{x})$ to denote the tangent cone to $\mathbf{\Omega}$ computed at $\mathbf{x}$, $\mathring{\mathbf{\Omega}}$ to denote the interior of $\mathbf{\Omega}$, and $\partial \mathbf{\Omega}$ to represent its boundary.

2 A direct search algorithm for nonlinear constrained optimization

This section is devoted to the introduction of a new directional direct search algorithm, based on the sequential minimization of an adequate merit function, to solve nonlinearly constrained derivative-free optimization problems. The first part of the section details the proposed algorithmic structure. The second part analyses the role of the linear constraints and their relationship with the choice of directions to be explored in the poll step. Finally, the main assumptions required for the theoretical analysis are stated and some preliminary results are established.

2.1 LOG-DS algorithm

The strategy proposed to solve Problem (P) builds upon the original idea of [10], where a merit function is defined which guarantees strict feasibility of general inequality constraints by means of logarithmic barrier penalty terms. Equality constraints are treated by exterior penalty terms. The weakness of such an approach is that one needs an initial guess which strictly satisfies all the inequality constraints, otherwise the merit function would not be defined, being the value infinite assigned to it. In the present work, we use this requirement to partitioning the original set of nonlinear inequalities into two subsets, such that feasibility is always guaranteed for the initialization. To this aim, given an initial point $\mathbf{x}_0$, partition the index set of the nonlinear inequality constraints, namely $\{1, 2, \dots, m\}$, into the following two index sets

$$\mathcal{G}^{\texttt{log}} = \{\ell \in \{1, \dots, m\} \,|\, g_\ell(\mathbf{x}_0) < 0\},$$
$$\mathcal{G}^{\texttt{ext}} = \{\ell \in \{1, \dots, m\} \,|\, g_\ell(\mathbf{x}_0) \geq 0\}.$$

Note that $\mathcal{G}^{\texttt{log}}$ and $\mathcal{G}^{\texttt{ext}}$ depend only on the initial point $\mathbf{x}_0$ and are indeed a partition of the index set of nonlinear inequality constraints, i.e. $\mathcal{G}^{\texttt{log}} \cup \mathcal{G}^{\texttt{ext}} = \{1, \dots, m\}$ and $\mathcal{G}^{\texttt{log}} \cap \mathcal{G}^{\texttt{ext}} = \emptyset$.

The inequalities $g_\ell(\mathbf{x}) \leq 0$ with $\ell \in \mathcal{G}_{\texttt{log}}$ will be treated with a logarithmic barrier approach, whereas the ones with $\ell \in \mathcal{G}_{\texttt{ext}}$ will be aggregated into an exterior penalization term along with the equality constraints. Using the index sets $\mathcal{G}_{\texttt{log}}$ and $\mathcal{G}_{\texttt{ext}}$, we can define

$$\begin{aligned}
\Omega_{\texttt{log}} &= \{\mathbf{x} \in \mathbb{R}^n \mid g_\ell(\mathbf{x}) \leq 0,\ \ell \in \mathcal{G}_{\texttt{log}}\},\\
\Omega_{\texttt{ext}} &= \{\mathbf{x} \in \mathbb{R}^n \mid g_\ell(\mathbf{x}) \leq 0,\ \ell \in \mathcal{G}_{\texttt{ext}} \text{ and } h_j(\mathbf{x}) = 0,\ j = 1, \dots, p\}.
\end{aligned}$$

Thus, the feasible region $\mathcal{F}$ of Problem (P) is given by $\mathcal{F} = \mathcal{X} \cap \Omega_{\texttt{log}} \cap \Omega_{\texttt{ext}}$. Note that, if all inequality constraints are violated by the initial guess, then $\mathcal{G}_{\texttt{log}} = \emptyset$ and $\mathcal{G}_{\texttt{ext}} = \{1, \dots, m\}$. In this case, $\Omega_{\texttt{log}} = \mathring{\Omega}_{\texttt{log}} = \mathbb{R}^n$. However, for the sake of clarity, we assume that $\mathcal{G}_{\texttt{log}} \neq \emptyset$.

We are now ready to introduce the merit function proposed in this work, which is defined by:

$$Z(\mathbf{x}; \rho) = \begin{cases} f(\mathbf{x}) - \rho \sum_{\ell \in \mathcal{G}_{\texttt{log}}} \log(-g_\ell(\mathbf{x})) & \\ \qquad + \frac{1}{\rho^{\nu-1}} \sum_{\ell \in \mathcal{G}_{\texttt{ext}}} (\max\{g_\ell(\mathbf{x}), 0\})^\nu & \text{if } \mathbf{x} \in \mathcal{X} \text{ and } g_\ell(\mathbf{x}) < 0, \ell \in \mathcal{G}_{\texttt{log}} \\ \qquad + \frac{1}{\rho^{\nu-1}} \sum_{j=1}^{p} |h_j(\mathbf{x})|^\nu & \\ +\infty & \text{otherwise,} \end{cases} \tag{1}$$

where $\rho > 0$ is a *penalty-barrier* parameter and $\nu \in (1, 2]$ plays the role of a *smoothing exponent*. We point out that ρ is the same for the different terms of the merit function, and the range of ν comes from theoretical requirements (see Theorem B.2).

Regarding the linear constraints, they are addressed directly during the minimization of $Z(\cdot; \rho)$. Therefore, given a penalty-barrier-parameter ρ_k, the following problem is considered

$$\begin{aligned} &\min\ Z(\mathbf{x}; \rho_k) \\ &\text{s.t. } \mathbf{x} \in \mathcal{X}. \end{aligned} \tag{P_{ρ_k}}$$

Following the classic approach of penalty and interior point methods, the proposed strategy is to generate a sequence of penalty-barrier parameters $\{\rho_k\}$, and sequentially solve the related linearly constrained subproblems. The goal is to obtain a sequence of iterates $\{\mathbf{x}_k\}$ that converges to a solution of the original Problem (P). Note that, in order for that to happen, one needs the sequence $\{\rho_k\}$ to converge to zero, so that, in the limit, the merit function $Z(\cdot; \rho_k)$ aligns with the objective function $f(\cdot)$ among feasible points.

Taking into account the black-box nature of the functions defining the problem, we propose a directional direct search approach to solve the subproblems (P_{ρ_k}). In particular, as mentioned above, we propose an adaptation of the SID-PSM algorithm, i.e., an implementation of GPS. In the literature, GPS algorithms, when applied to the minimization of a function $\varphi : \mathbb{R}^n \to \mathbb{R}$, accept new points whenever a *simple decrease* of $\varphi(\cdot)$ is obtained, that is, given a current solution $\mathbf{x}_k$ and a trial point $\mathbf{y}$, whenever $\varphi(\mathbf{y}) < \varphi(\mathbf{x}_k)$ holds. This forces the method to rely on an implicit mesh to ensure the existence of limit points of the sequence of iterates $\{\mathbf{x}_k\}$. In the proposed approach, instead, the acceptance of new points relies on the notion of *sufficient decrease* [15]. The next definition adjusts the concept of *forcing function* (see [22]) and it is used to define the sufficient decrease condition required to accept new points.

Definition 2.1. *Let $\xi : [0, +\infty) \to [0, +\infty)$ be a continuous and nondecreasing function. We say that ξ is a forcing function if $\xi(t)/t \to 0$ when $t \downarrow 0$, and $\xi(t) \to 0$ implies $t \to 0$.*

A classic example of a forcing function is $\xi(t) = \gamma t^2$, with $\gamma > 0$, which is also the most commonly used in practice.

The proposed algorithmic structure is detailed below.

Algorithm LOG-DS.

Data. $\mathcal{G}_{\texttt{log}}$ and $\mathcal{G}_{\texttt{ext}}$, $\mathbf{x}_0 \in \mathcal{X} \cap \mathring{\Omega}_{\texttt{log}}$, $\mathcal{D}$ a set of sets of normalized directions, a initial step-size $\alpha_0 > 0$, a initial penalty-barrier parameter $\rho_0 > 0$, a smoothing exponent $\nu \in (1, 2]$, step-size and penalty-barrier contraction parameters $\theta_\alpha, \theta_\rho \in (0, 1)$, a step-size expansion parameter $\phi \geq 1$, and $\beta > 1$.

For $k = 0, 1, 2, \ldots$ **do**

Step 1. (Search Step, optional)

If $\mathbf{z}_k \in \mathcal{X}$ can be computed such that $Z(\mathbf{z}_k; \rho_k) \leq Z(\mathbf{x}_k; \rho_k) - \xi(\alpha_k)$
set $\mathbf{x}_{k+1} = \mathbf{z}_k$, $\alpha_{k+1} = \phi\alpha_k$, $\rho_{k+1} = \rho_k$, declare the iteration **Successful**,
set $k = k + 1$, and repeat **Step 1.**
Else go to **Step 2.**

Step 2. (Poll Step)

Select a set of normalized directions $\mathcal{D}_k \in \mathcal{D}$
If $\exists\, \mathbf{d}_k^i \in \mathcal{D}_k$: $\mathbf{x}_k + \alpha_k \mathbf{d}_k^i \in \mathcal{X}$ and $Z(\mathbf{x}_k + \alpha_k \mathbf{d}_k^i; \rho_k) \leq Z(\mathbf{x}_k; \rho_k) - \xi(\alpha_k)$
set $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k^i$, $\alpha_{k+1} = \phi\alpha_k$, $\rho_{k+1} = \rho_k$, declare the iteration **Successful**,
set $k = k + 1$, and go to **Step 1.**
Else set $\mathbf{x}_{k+1} = \mathbf{x}_k$, $\alpha_{k+1} = \theta_\alpha \alpha_k$, declare the iteration **Unsuccessful**,
and go to **Step 3.**

Step 3. (Penalty-Barrier Update Step)

Set $(g_{\min})_k = \min_{\ell \in \mathcal{G}^{\log}} \{|g_\ell(\mathbf{x}_k)|\}$

If $\alpha_{k+1} \leq \min\{\rho_k^\beta, (g_{\min})_k^2\}$ **then** set $\rho_{k+1} = \theta_\rho \rho_k$. **Else** set $\rho_{k+1} = \rho_k$.

Endfor

In the above scheme, the initial guess $\mathbf{x}_0$ is required to satisfy $\mathbf{x}_0 \in \mathcal{X} \cap \mathring{\Omega}_{\texttt{log}}$, i.e. $x \in \mathcal{X}$ and

$g_\ell(\mathbf{x}_0) < 0$ for all $\ell \in \mathcal{G}_{\texttt{log}}$. This ensures that the initial value of the merit function $Z(\mathbf{x}_0; \rho_0)$ is finite. It is important to note that $\mathcal{G}_{\texttt{log}}$ and $\mathcal{G}_{\texttt{ext}}$ are determined during the initialization phase and are not modified as the algorithm progresses. The method iterates over three steps. The first two follow the general structure proposed by Audet and Dennis [6] for GPS, whereas the third is related to the novel penalty-interior point approach.

Step 1, the *search step*, allowing the use of heuristics, is very flexible, not even requiring the projection of the generated points in some type of implicit mesh, since a sufficient decrease condition is used for the acceptance of new iterates. As it is detailed in Section 4, the original SID-PSM algorithm uses quadratic polynomial models, which are minimized to generate new trial points. The latter approach is adapted into LOG-DS as described in Section 4.1, but since it is not relevant for establishing convergence properties, for now it is omitted. If, during the search step, it is possible to find a point $\mathbf{z}_k \in \mathcal{X}$ such that, with respect to the current *incumbent* $\mathbf{x}_k$, it provides a sufficient decrease in the merit function, i.e., if $Z(\mathbf{z}_k; \rho_k) \leq Z(\mathbf{x}_k; \rho_k) - \xi(\alpha_k)$, then the algorithm sets $\mathbf{x}_{k+1} = \mathbf{z}_k$ and proceeds with iteration $k+1$. Otherwise, the poll step is invoked.

Step 2, the *poll step*, initializes with the selection of a set of normalized *poll directions* $\mathcal{D}_k$. Further considerations on the choice of $\mathcal{D}_k$ are discussed in Section 2.2. If a direction $\mathbf{d}_k^i \in \mathcal{D}_k$ is found such that the trial point $\mathbf{x}_k + \alpha_k \mathbf{d}_k^i$ belongs to $\mathcal{X}$ and achieves a sufficient decrease in the merit function, then the algorithm sets $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k^i$ and proceeds with iteration $k+1$.

Note that whenever $\mathbf{x}_{k+1} \neq \mathbf{x}_k$ one has $Z(\mathbf{x}_{k+1}; \rho_k) \leq Z(\mathbf{x}_k; \rho_k) - \xi(\alpha_k)$, meaning that, during either the search step or the poll step, the method identifies a better point as candidate to the solution of (P_{ρ_k}). We call such iterations *successful*, and we refer to the index set of successful iterations as $\mathcal{K}_s$. On the other hand, if $\mathbf{x}_{k+1} = \mathbf{x}_k$, it means that none of the trial points identified by the poll directions achieves a sufficient decrease in the merit function, i.e. $Z(\mathbf{x}_k + \alpha_k \mathbf{d}_k^i; \rho_k) > Z(\mathbf{x}_k; \rho_k) - \xi(\alpha_k)$ for all $\mathbf{d}_k^i \in \mathcal{D}_k$. We call such iterations *unsuccessful*, and we refer to the index set of unsuccessful iterations as $\mathcal{K}_u$.

Only and at all unsuccessful iterations, **Step 3** is invoked. This step concerns the update of the penalty-barrier parameter ρ_k, assessing if (P_{ρ_k}) can be considered as solved. Since providing an exact solution of the problem is, in general, impractical, one has to define a condition to determine if the precision of the solution $\mathbf{x}_k$ is acceptable. We propose the following criterion

$$\alpha_{k+1} \leq \min\{\rho_k^\beta, (g_{min})_k^2\},$$

with $\beta > 1$. It is well-known that α_k represents a measure of stationarity for problem (P_{ρ_k}) (see [22]). Hence, having $\alpha_{k+1} \leq \rho_k^\beta$ implies that we are not interested in good approximations of the solution of (P_{ρ_k}) for high values of ρ_k. Additionally, $(g_{\min})_k = \min_{\ell \in \mathcal{G}_{\texttt{log}}}\{|g_\ell(\mathbf{x}_k)|\}$ represents how close we are to the boundary of $\mathbf{\Omega}_{\texttt{log}}$. Since the merit function $Z(\mathbf{x}; \rho_k)$ takes the value infinite for all $\mathbf{x} \notin \mathring{\mathbf{\Omega}}_{\texttt{log}}$, by imposing $\alpha_{k+1} \leq (g_{\min})_k^2$ we are trying to force the algorithm to properly explore the interior of $\mathbf{\Omega}_{\texttt{log}}$ before moving on to the next subproblem. For further details see [10].

2.2 Poll directions and linear inequalities

In unconstrained minimization problems, the poll directions should be able to generate $\mathbb{R}^n$ through nonnegative linear combinations, in order to capture the behavior of the objective function (see [12, 20, 21], for further details on the topic). In the presence of linear inequalities, though, the directions must be adapted to the the geometry of the set $\mathcal{X}$, identified by the tangent cone. Let us formalize this in the following definition.

Definition 2.2 (Active constraints and tangent related sets)**.** *For every* $\mathbf{x} \in \mathcal{X}$*, i.e., such that* $\mathbf{A}\mathbf{x} \leq \mathbf{b}$*, let* $a_i^\top, i \in \{1, \dots, q\}$ *be the* i*th row of* A*. We define:*

$$\mathcal{I}_{\mathcal{X}}(\mathbf{x}) = \{i \in \{1, \dots, q\} \mid \mathbf{a}_i^\top \mathbf{x} = b_i\} \quad \textit{(set of indices of active constraints)}$$
$$\mathcal{T}_{\mathcal{X}}(\mathbf{x}) = \{\mathbf{d} \in \mathbb{R}^n \mid \mathbf{a}_i^\top \mathbf{d} \leq 0,\ i \in \mathcal{I}_{\mathcal{X}}(\mathbf{x})\} \quad \textit{(tangent cone to } \mathcal{X} \textit{ at } \mathbf{x}\textit{)}$$

Since $\mathcal{X}$ is convex by definition, for all points $\mathbf{x} \in \mathcal{X}$ one has that $\mathbf{x}+t\mathbf{d} \in \mathcal{X}$ for all $\mathbf{d} \in \mathcal{T}_{\mathcal{X}}(\mathbf{x})$ and $t > 0$ sufficiently small. Clearly, if $\mathbf{x}$ lies in the interior of $\mathcal{X}$, then $\mathcal{T}_{\mathcal{X}}(\mathbf{x}) = \mathbb{R}^n$.

For any incumbent $\mathbf{x}_k$, the tangent cone at $\mathbf{x}_k$ might be used to understand the geometry of $\mathcal{X}$ at $\mathbf{x}_k$ and to build the poll directions accordingly. While this approach might seem reasonable, in general, it is not sufficient. Indeed, the algorithm might generate a sequence of iterates $\{\mathbf{x}_k\}$ lying in the interior of $\mathcal{X}$ and converging to a point $\bar{\mathbf{x}} \in \partial\mathcal{X}$. In such case, the tangent cones $\mathcal{T}_{\mathcal{X}}(\mathbf{x}_k)$ are all equal to $\mathbb{R}^n$, and the algorithm might not be able to sample the space using directions in $\mathcal{T}_{\mathcal{X}}(\bar{\mathbf{x}}) \subset \mathbb{R}^n$. Thus, given an iterate $\mathbf{x}_k \in \mathcal{X}$, it is important to be able to capture the geometry of the set $\mathcal{X}$ *near* $\mathbf{x}_k$. In order to do that, the tangent cone is approximated by the ε-tangent cone [19, 22, 23].

Definition 2.3 (ε-Active constraints and tangent related sets)**.**

$$\mathcal{I}_{\mathcal{X}}(\mathbf{x}_k, \varepsilon) = \{i \in \{1, \dots, q\} \mid \mathbf{a}_i^\top \mathbf{x}_k \geq b_i - \varepsilon\} \quad \textit{(set of indices of } \varepsilon\textit{-active constraints)}$$
$$\mathcal{T}_{\mathcal{X}}(\mathbf{x}_k, \varepsilon) = \{\mathbf{d} \in \mathbb{R}^n \mid \mathbf{a}_i^\top \mathbf{d} \leq 0,\ i \in \mathcal{I}_{\mathcal{X}}(\mathbf{x}_k, \varepsilon)\} \quad \textit{(}\varepsilon\textit{-tangent cone at } \mathbf{x}_k\textit{)}$$

A relation between the tangent cone and the ε-tangent cone has been established in [30]. Let us recall it in the following proposition.

Proposition 2.4. *Let* $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$ *be a sequence of points in* $\mathcal{X}$*, converging to* $\mathbf{x}^* \in \mathcal{X}$*. Then, there exists an* $\varepsilon^* > 0$ *(depending only on* $\mathbf{x}^*$*) such that for any* $\varepsilon \in (0, \varepsilon^*]$ *there exists* $k_\varepsilon \in \mathbb{N}$ *such that*

$$\mathcal{I}_{\mathcal{X}}(\mathbf{x}^*) = \mathcal{I}_{\mathcal{X}}(\mathbf{x}_k, \varepsilon)$$
$$\mathcal{T}_{\mathcal{X}}(\mathbf{x}^*) = \mathcal{T}_{\mathcal{X}}(\mathbf{x}_k, \varepsilon)$$

for all $k \geq k_\varepsilon$*.*

Proposition 2.4 ensures that, if at each iteration k , the ε-tangent cone of $\mathbf{x}_k$ is used to build the set of poll directions $\mathcal{D}_k$, then, for small enough $\varepsilon > 0$, the algorithm is able to capture the geometry of $\mathcal{X}$ at the limit point $\bar{\mathbf{x}}$. The requirements on the sets of poll directions $\mathcal{D}_k$ used by the algorithm are formalized in Assumption 1 (see [28, Assumption 2]).

Assumption 1. *Let $\{\mathbf{x}_k\}$ be a sequence of points in $\mathcal{X}$. The sequence $\{\mathcal{D}_k\}$ of poll directions satisfies:*

$$\mathcal{D}_k = \{\mathbf{d}_k^i \mid \|\mathbf{d}_k^i\| = 1, i = 1, \dots, |\mathcal{D}_k|\}$$

and for some $\bar{\varepsilon} > 0$,

$$\text{cone}(\mathcal{D}_k \cap \mathcal{T}_{\mathcal{X}}(\mathbf{x}_k, \varepsilon)) = \mathcal{T}_{\mathcal{X}}(\mathbf{x}_k, \varepsilon), \quad \forall\ \varepsilon \in (0, \bar{\varepsilon}].$$

Furthermore, $\mathcal{D} = \bigcup_{k=0}^{+\infty} \mathcal{D}_k$ is a finite set, and $|\mathcal{D}_k|$ is uniformly bounded.

Strategies to conform the search directions to the geometry of the nearby feasible region, i.e. to satisfy Assumption 1, can be found in [1, 23, 26, 30].

2.3 Preliminary results

The method we propose is based on the sequential minimization of the merit function $Z(\cdot; \rho_k)$ for a decreasing sequence of penalty-barrier parameters $\{\rho_k\}$. In order for the method to work, an optimal solution has to exist for each Problem (P_{ρ_k}). To guarantee the existence of an optimal solution for a minimization problem, one of the most standard assumptions is the compactness of the lower-level sets of the objective function. In our method, this would correspond to the compactness of the lower level sets $\mathcal{L}_\rho(\alpha) = \{\mathbf{x} \in \mathcal{X} \mid Z(\mathbf{x}; \rho) \le \alpha\}$ for all $\rho > 0$. Note that the values of $Z(\mathbf{x}; \rho)$ are considered only for points lying in $\mathcal{X}$. To this end, we make use of the following assumption, and we prove the desired result in Lemma 2.5.

Assumption 2. *The set $\mathcal{X} \cap \Omega_{\texttt{log}}$ is compact and a point $\mathbf{x}_0 \in \mathcal{X}$ exists such that $g_\ell(\mathbf{x}_0) < 0$ for all $\ell \in \mathcal{G}_{\texttt{log}}$.*

Lemma 2.5. *Let Assumption 2 hold. Then, for all $\rho > 0$, $\nu \in (1, 2]$, and $\alpha \in \mathbb{R}$ the lower-level set*

$$\mathcal{L}_\rho(\alpha) = \{\mathbf{x} \in \mathcal{X} \mid Z(\mathbf{x}; \rho) \le \alpha\}$$

is compact.

Proof. By definition, $\mathcal{L}_\rho(\alpha)$ is a subset of $\mathcal{X}$, and $Z(\cdot; \rho)$ is coercive with respect to $\Omega_{\texttt{log}}$, i.e. $Z(\mathbf{x}; \rho) \to +\infty$ as $\mathbf{x} \to \partial\Omega_{\texttt{log}}$, thus, $\mathcal{L}_\rho(\alpha) \subseteq \mathcal{X} \cap \Omega_{\texttt{log}}$. Since, by Assumption 2, $\mathcal{X} \cap \Omega_{\texttt{log}}$ is compact, then $\mathcal{L}_\rho(\alpha)$ is bounded.

It remains to prove that $\mathcal{L}_\rho(\alpha)$ is closed. To this end, let us show that, for any sequence $\{\mathbf{x}_k\} \subset \mathcal{L}_\rho(\alpha)$ such that $\lim_{k\to+\infty} \mathbf{x}_k = \bar{\mathbf{x}}$, it results $\bar{\mathbf{x}} \in \mathcal{L}_\rho(\alpha)$.

Since $\mathbf{x}_k \in \mathcal{L}_\rho(\alpha)$ for all k, we have

$$Z(\mathbf{x}_k; \rho) \le \alpha. \tag{2}$$

Function $Z(.; \rho)$ is continuous at all $\mathbf{x} \in \mathcal{L}_\rho(\alpha)$. Thus, we have

$$\lim_{k\to+\infty} Z(\mathbf{x}_k; \rho) = Z(\bar{\mathbf{x}}; \rho) \le \alpha,$$

which means that $\bar{\mathbf{x}} \in \mathcal{L}_\rho(\alpha)$ and concludes the proof. □

Note that Assumption 2 arises from the observation that the logarithmic term of $Z(\cdot;\rho)$ might go to $-\infty$ if also $g_\ell(\cdot) \to -\infty$ for some $\ell \in \mathcal{G}_{\mathtt{log}}$. Since the functions are assumed to be continuous on $\mathcal{X}$, the compactness assumption prevents such situation from happening. Note that, since $\mathcal{X}$ is defined by the linear inequalities of the problem, then $\mathcal{X}$ also encompasses the possible bounds on the variables. In practical applications, one is usually able to identify some finite bounds wherein the optimal solution should lie. Therefore, the assumption that $\mathcal{X}$ is compact can be considered realistic.

We use Lemma 2.5 to prove properties that form the basis of the theoretical convergence analysis of LOG-DS. In particular, it is exploited in the proof of Theorem 2.7. Furthermore, it is important to highlight that the results presented in this work do not pertain to the entire sequence generated by the proposed algorithm, but focus on the subset of iterations wherein the penalty-barrier parameter ρ_k is updated, i.e., the sequence of inexact solutions of the linearly constrained Problems (P_{ρ_k}). Following the terminology used in the interior point methods literature (see [8]), we refer to the sequence of inexact solutions generated by LOG-DS as a *path following* sequence, and we denote the corresponding index set by $\mathcal{K}_\rho$.

Definition 2.6. *Let $\mathcal{K}_\rho$ be defined as:*

$$\mathcal{K}_\rho = \{k \in \mathbb{N} \mid \ \rho_{k+1} < \rho_k\}.$$

Given the sequence of iterates $\{\mathbf{x}_k\}$ produced by Algorithm LOG-DS, the subsequence $\{\mathbf{x}_k\}_{k\in\mathcal{K}_\rho}$ is said to be a path-following subsequence.

Let us recall that **Step 3** is invoked only at unsuccessful iterations, i.e., iterations $k \in \mathcal{K}_u = \{k \in \mathbb{N} \mid \ x_{k+1} = x_k\}$, so one has $\mathcal{K}_\rho \subseteq \mathcal{K}_u$. Additionally, according to the instructions of the algorithm, if $k \in \mathcal{K}_\rho$ then the following criterion is satisfied:

$$\alpha_{k+1} \leq \{\rho_k^\beta, (g_{min})_k^2\}, \tag{3}$$

and $\rho_{k+1} = \theta_\rho \rho_k$, with $\theta_\rho \in (0,1)$. From the updating procedure, it follows that $\{\rho_k\}$ goes to zero if and only if $\mathcal{K}_\rho$ is infinite.

In Theorem 2.7, we show that $\mathcal{K}_\rho$ is infinite and, consequently, the sequence of penalty-barrier parameters converges to zero, which is required to ensure that, in the limit, the algorithm is able to solve the original problem. Furthermore, it is possible to establish that the corresponding sequence of step-sizes also converges to zero. Since, as we mentioned above, the step-size is related to some measure of stationarity, this sets the ground to prove convergence to stationary points.

Theorem 2.7. *Let Assumption 2 hold. Let $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$, $\{\rho_k\}_{k\in\mathbb{N}}$, and $\{\alpha_k\}_{k\in\mathbb{N}}$ be the sequences of iterates, penalty parameters and step-sizes generated by LOG-DS, respectively. Then, $\mathcal{K}_\rho$ is infinite and*

$$\lim_{k\to+\infty} \rho_k = 0. \tag{4}$$

Furthermore, it holds

$$\lim_{k\to+\infty, k\in\mathcal{K}_\rho} \alpha_k = 0. \tag{5}$$

Proof. Let us start by proving that $\mathcal{K}_\rho$ is infinite. By contradiction, assume that $\mathcal{K}_\rho$ is finite, and without loss of generality, let us assume $\mathcal{K}_\rho = \emptyset$. Thus, $\rho_k = \rho_0 > 0$ for all k.

Let us consider the sets of successful and unsuccessful iterations, which we recall to be defined as $\mathcal{K}_s = \{k \in \mathbb{N} \,|\mathbf{x}_{k+1} \neq \mathbf{x}_k\}$ and $\mathcal{K}_u = \{k \in \mathbb{N} \,|\mathbf{x}_{k+1} = \mathbf{x}_k\}$. By the instructions of the algorithm, it follows

$$Z(\mathbf{x}_{k+1};\rho_0) \leq Z(\mathbf{x}_k;\rho_0) - \xi(\alpha_k), \quad \forall k \in \mathcal{K}_s, \tag{6}$$

$$Z(\mathbf{x}_{k+1};\rho_0) = Z(\mathbf{x}_k;\rho_0), \quad \forall k \in \mathcal{K}_u. \tag{7}$$

Hence, the sequence of function values $\{Z(\mathbf{x}_k;\rho_0)\}$ is monotonically nonincreasing. By Lemma 2.5, $Z(\cdot;\rho_0)$ has compact lower-level sets, thus it is bounded below. Hence,

$$\lim_{k\to+\infty} Z(\mathbf{x}_k;\rho_0) = \bar{Z}. \tag{8}$$

If $\mathcal{K}_s$ is infinite, from (8) and (6) we get

$$\lim_{\substack{k\to+\infty\\ k\in\mathcal{K}_s}} \xi(\alpha_k) = 0.$$

Recalling Definition 2.1, we get

$$\lim_{\substack{k\to+\infty\\ k\in\mathcal{K}_s}} \alpha_k = 0. \tag{9}$$

If $\mathcal{K}_u$ is infinite and $\mathcal{K}_s$ is not empty, for every $k \in \mathcal{K}_u$, let us define m_k to be the largest index such that $m_k \in \mathcal{K}_s$ and $m_k < k$. Then, we can write

$$\alpha_k = \alpha_{m_k}\phi\theta_\alpha^{k-m_k-1}.$$

When $k \to +\infty$, $k \in \mathcal{K}_u$, we have that either $m_k \to +\infty$ as well (when $\mathcal{K}_s$ is infinite) or $k - m_k - 1 \to +\infty$ (when $\mathcal{K}_s$ is finite). Thus, by (9) and the fact that $\theta_\alpha \in (0,1)$, we have

$$\lim_{\substack{k\to+\infty\\ k\in\mathcal{K}_u}} \alpha_k = \lim_{\substack{k\to+\infty\\ k\in\mathcal{K}_u}} \alpha_{m_k}\phi\theta_\alpha^{k-m_k-1} = 0. \tag{10}$$

If $\mathcal{K}_u$ is infinite and $\mathcal{K}_s$ is empty, then $\alpha_k = \alpha_0\theta_\alpha^k$ and

$$\lim_{\substack{k\to+\infty\\ k\in\mathcal{K}_u}} \alpha_k = 0, \tag{11}$$

follows.

Considering (9), (10), and (11), we can conclude that

$$\lim_{k\to+\infty} \alpha_k = 0. \tag{12}$$

By the instructions of the algorithm, for $k \in \mathcal{K}_u$ sufficiently large, since $k \notin \mathcal{K}_\rho$ and recalling (3), it holds

$$\alpha_{k+1} > \min\{\rho_0^\beta, (g_{\min})_k^2\}.$$

The latter, recalling (12), implies

$$\lim_{\substack{k\to+\infty,\\ k\in\mathcal{K}_u}} (g_{\min})_k = 0,$$

which, by the definition of $Z(\cdot;\rho_0)$, further implies

$$\lim_{\substack{k\to+\infty,\\ k\in\mathcal{K}_u}} Z(\mathbf{x}_k;\rho_0) = +\infty.$$

This is a contradiction with the monotonicity of $\{Z(\mathbf{x}_k;\rho_0)\}$, thus proving that $\mathcal{K}_\rho$ is infinite.

Let us now prove (4). Let $\mathcal{K}_\rho = \{k_1, k_2, \dots, k_j, \dots\}$. Thus, by the instructions of the algorithm, we have

$$\rho_{k_j} = \theta_\rho \rho_{k_{j-1}} = \theta_\rho^j \rho_0.$$

Since $\mathcal{K}_\rho$ is infinite, we can take the limit for $j \to +\infty$, and, recalling $\theta_\rho \in (0,1)$, we get

$$\lim_{j\to+\infty} \rho_{k_j} = \lim_{j\to+\infty} \theta_\rho^j \rho_0 = 0,$$

which, noting $\rho_{k+1} \le \rho_k$ for all k, proves (4).

Now, we prove (5). By definition of $\mathcal{K}_\rho$, we have for all $k \in \mathcal{K}_\rho$

$$\alpha_{k+1} \le \min\{\rho_k^\beta, (g_{\min})_k^2\} \le \rho_k^\beta.$$

Furthermore, we know that for all iterations $\alpha_{k+1} \ge \theta_\alpha \alpha_k$, so that we can write, for all $k \in \mathcal{K}_\rho$,

$$\theta_\alpha \alpha_k \le \alpha_{k+1} \le \rho_k^\beta.$$

Then, the proof follows by recalling that $\lim_{k\to+\infty} \rho_k = 0$. □

3 Convergence analysis

In this section, we conduct a theoretical analysis of the convergence properties of Algorithm LOG-DS. First, we define the extended Mangasarian–Fromovitz constraint qualification, which allows us to derive necessary optimality conditions for Problem (P). Then we present a key technical result, and finally the main finding of the analysis is stated in Theorem 3.5.

We use the following extended Mangasarian–Fromovitz constraint qualification (MFCQ).

Definition 3.1. *The point $\mathbf{x} \in \mathcal{X}$ is said to satisfy the MFCQ for Problem* (P) *if the two following conditions are satisfied:*

(a) *There does not exist a nonzero vector $\alpha = (\alpha_1, \dots, \alpha_p)$ such that:*

$$\left(\sum_{i=1}^{p} \alpha_i \nabla h_i(\mathbf{x})\right)^\top \mathbf{d} \ge 0, \qquad \forall \mathbf{d} \in \mathcal{T}_\mathcal{X}(\mathbf{x}). \tag{13}$$

(b) *There exists a feasible direction* $\mathbf{d} \in \mathcal{T}_{\mathcal{X}}(\mathbf{x})$, *such that:*

$$\nabla g_\ell(\mathbf{x})^\top \mathbf{d} < 0, \quad \forall \ell \in \mathcal{I}_+(\mathbf{x}), \quad \nabla h_j(\mathbf{x})^\top \mathbf{d} = 0, \quad \forall j = 1, \dots, p \tag{14}$$

where $\mathcal{I}_+(\mathbf{x}) = \{\ell \in \{1, \dots, m\} \mid g_\ell(\mathbf{x}) \geq 0\}$.

Consider the Lagrangian function $L(\mathbf{x}, \lambda, \mu)$, associated with the nonlinear constraints of Problem (P), defined by:

$$L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \lambda^\top g(\mathbf{x}) + \mu^\top h(\mathbf{x}).$$

The following proposition is a well-known result (see [8, Sec. 3.3]), which states necessary optimality conditions for Problem (P).

Proposition 3.2. *Let* $\mathbf{x}^* \in \mathcal{F}$ *be a local minimum of Problem* (P) *that satisfies the MFCQ. Then, there exist vectors* $\lambda^* \in \mathbb{R}^m$, $\mu^* \in \mathbb{R}^p$ *such that*

$$\nabla_x L(\mathbf{x}^*, \lambda^*, \mu^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0, \qquad \forall \mathbf{x} \in \mathcal{X} \tag{15}$$

$$(\lambda^*)^\top g(\mathbf{x}^*) = 0, \quad \lambda^* \geq 0. \tag{16}$$

Therefore, we consider the following definition of stationarity.

Definition 3.3 (Stationary point)**.** *A point* $\mathbf{x}^* \in \mathcal{F}$ *is said to be a stationary point for Problem* (P) *if vectors* $\lambda^* \in \mathbb{R}^m$ *and* $\mu^* \in \mathbb{R}^p$ *exist such that* (15) *and* (16) *are satisfied.*

Before proving the main convergence result, we establish a technical proposition which will be invoked in Theorem 3.5. The need for this proposition stems from noting that the trial points corresponding to failures might not strictly satisfy the inequality constraints with indices $\ell \in \mathcal{G}_{\texttt{log}}$, implying that the merit function value is, by definition, $+\infty$.

Proposition 3.4. *Given any* $\bar{\rho} > 0$, $\mathbf{x} \in \mathcal{X}$ *such that* $g_\ell(\mathbf{x}) < 0$ *for all* $\ell \in \mathcal{G}_{\texttt{log}}$, $\mathbf{d} \in \mathbb{R}^n$, *and* $\bar{\alpha} \in \mathbb{R}_+$ *such that* $\mathbf{x} + \bar{\alpha}\mathbf{d} \in \mathcal{X}$ *and*

$$Z(\mathbf{x} + \bar{\alpha}\mathbf{d}; \bar{\rho}) > Z(\mathbf{x}; \bar{\rho}) - \xi(\bar{\alpha}),$$

there exists $\hat{\alpha} \leq \bar{\alpha}$ *such that:*

$$\begin{aligned}
&\mathbf{x} + \alpha\mathbf{d} \in \mathcal{X} \quad \forall \alpha \in (0, \hat{\alpha}], \\
&\max_{\ell \in \mathcal{G}_{\texttt{log}}} g_\ell(\mathbf{x} + \alpha\mathbf{d}) < 0 \quad \forall\ \alpha \in (0, \hat{\alpha}], \\
&Z(\mathbf{x} + \hat{\alpha}\mathbf{d}; \bar{\rho}) > Z(\mathbf{x}; \bar{\rho}) - \xi(\hat{\alpha}).
\end{aligned}$$

Proof. By definition $\mathcal{X}$ is convex, then $\mathbf{x} + \alpha\mathbf{d} \in \mathcal{X}$ for all $\alpha \in [0, \bar{\alpha}]$, thus if $\max_{\ell \in \mathcal{G}_{\texttt{log}}} g_\ell(\mathbf{x} + \alpha\mathbf{d}) < 0$ for all $\alpha \in (0, \bar{\alpha}]$, by setting $\hat{\alpha} = \bar{\alpha}$, the proposition holds.
Let us assume $\max_{\ell \in \mathcal{G}_{\texttt{log}}} g_\ell(\mathbf{x} + \alpha\mathbf{d}) \geq 0$ for some $\alpha \in (0, \bar{\alpha}]$. Since $g_\ell(\mathbf{x}) < 0$ for all $\ell \in \mathcal{G}_{\texttt{log}}$, it follows that $\max_{\ell \in \mathcal{G}_{\texttt{log}}} g_\ell(\mathbf{x}) < 0$. Hence, by continuity of $g_\ell(\cdot)$, there exists a scalar $\tilde{\alpha} < \bar{\alpha}$ such that

$$\max_{\ell \in \mathcal{G}_{\texttt{log}}} g_\ell(\mathbf{x} + \alpha\mathbf{d}) < 0 \ \text{ for all } \alpha \in [0, \tilde{\alpha}),$$

$$\max_{\ell \in \mathcal{G}_{\texttt{log}}} g_\ell(\mathbf{x} + \tilde{\alpha}\mathbf{d}) = 0.$$

Recall that, by definition of $Z(\cdot;\bar{\rho})$, it holds that

$$\lim_{\alpha \to \tilde{\alpha}} Z(\mathbf{x} + \alpha\mathbf{d}; \bar{\rho}) = +\infty.$$

Thus, there exists $\hat{\alpha} \in (0, \tilde{\alpha})$, sufficiently close to $\tilde{\alpha}$, such that $g_\ell(\mathbf{x} + \hat{\alpha}\mathbf{d}) < 0$, for all $\ell \in \mathcal{G}_{\texttt{log}}$, and $Z(\mathbf{x}; \bar{\rho}) < Z(\mathbf{x} + \hat{\alpha}\mathbf{d}; \bar{\rho}) + \xi(\hat{\alpha})$, which concludes the proof. □

We are now in the condition of stating the main convergence result.

Theorem 3.5. *Let Assumption 2 hold. Let $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$ be the sequence of iterates generated by LOG-DS and let $\mathcal{K}_\rho$ be a path following sequence. Assume that the sets of directions $\{\mathcal{D}_k\}_{k\in\mathbb{N}}$, used by the algorithm, satisfy Assumption 1 and define $\mathcal{J}_k = \{i \in \{1, 2, \ldots, |\mathcal{D}_k|\} \mid \mathbf{d}_k^i \in \mathcal{D}_k \cap \mathcal{T}_{\mathcal{X}}(\mathbf{x}_k; \varepsilon)\}$, with $\varepsilon \in (0, \min\{\bar{\varepsilon}, \varepsilon^*\}]$ where ε^* and $\bar{\varepsilon}$ are the constants appearing in Proposition 2.4 and Assumption 1, respectively. Then, any limit point of $\{\mathbf{x}_k\}_{k\in\mathcal{K}_\rho}$ that satisfies the MFCQ is a stationary point of Problem* (P).

Proof. First note that, by Theorem 2.7, we have

$$\lim_{\substack{k\to+\infty \\ k\in\mathcal{K}_\rho}} \rho_k = 0,$$

$$\lim_{\substack{k\to+\infty \\ k\in\mathcal{K}_\rho}} \alpha_k = 0.$$

Now, let $\mathbf{x}^*$ be any limit point of $\{\mathbf{x}_k\}_{k\in\mathcal{K}_\rho}$. Then, there exists a set $\mathcal{K}_\rho^{\mathbf{x}} \subseteq \mathcal{K}_\rho$ such that

$$\lim_{\substack{k\to+\infty \\ k\in\mathcal{K}_\rho^{\mathbf{x}}}} \rho_k = 0,$$

$$\lim_{\substack{k\to+\infty \\ k\in\mathcal{K}_\rho^{\mathbf{x}}}} \alpha_k = 0,$$

$$\lim_{\substack{k\to+\infty \\ k\in\mathcal{K}_\rho^{\mathbf{x}}}} \mathbf{x}_k = \mathbf{x}^*,$$

with $\alpha_{k+1} < \alpha_k$ and $\rho_{k+1} < \rho_k$, for all $k \in \mathcal{K}_\rho^{\mathbf{x}}$. Recall that $\mathcal{D}_k = \{\mathbf{d}_k^1, \mathbf{d}_k^2, \ldots, \mathbf{d}_k^{|\mathcal{D}_k|}\}$. Then, for all $k \in \mathcal{K}_\rho^{\mathbf{x}}$ sufficiently large, we know that $\mathbf{x}_k + \alpha_k\mathbf{d}_k^i \in \mathcal{X}$ for all $i \in \mathcal{J}_k$. For every $i \in \mathcal{J}_k$, if $g_\ell(\mathbf{x}_k + \alpha_k\mathbf{d}_k^i) < 0$, for all $\ell \in \mathcal{G}_{\texttt{log}}$, by the instructions of the algorithm we have

$$Z(\mathbf{x}_k + \alpha_k\mathbf{d}_k^i; \rho_k) > Z(\mathbf{x}_k; \rho_k) - \xi(\alpha_k). \tag{17}$$

Otherwise, i.e. when an index $\ell \in \mathcal{G}_{\texttt{log}}$ exists such that $g_\ell(\mathbf{x}_k + \alpha_k\mathbf{d}_k^i) \geq 0$, i.e. $Z(\mathbf{x}^k + \alpha_k\mathbf{d}_k^i; \rho_k) = +\infty$, Proposition 3.4 allows us to ensure the existence of a scalar $\hat{\alpha}_k^i \leq \alpha_k$ such that

$$Z(\mathbf{x}_k + \hat{\alpha}_k^i\mathbf{d}_k^i; \rho_k) > Z(\mathbf{x}_k; \rho_k) - \xi(\hat{\alpha}_k^i). \tag{18}$$

Applying the mean value theorem to (18) (or (17) setting $\hat{\alpha}_k^i = \alpha_k$), we can write

$$-\xi(\hat{\alpha}_k^i) \le Z(\mathbf{x}_k + \hat{\alpha}_k^i \mathbf{d}_k^i; \rho_k) - Z(\mathbf{x}_k; \rho_k) = \hat{\alpha}_k^i \nabla Z(\mathbf{y}_k^i; \rho_k)^\top \mathbf{d}_k^i, \tag{19}$$

for all $k \in \mathcal{K}_\rho^{\mathtt{x}}$ sufficiently large and all $i \in \mathcal{J}_k$, where $\mathbf{y}_k^i = \mathbf{x}_k + t_k^i \hat{\alpha}_k^i \mathbf{d}_k^i$, with $t_k^i \in (0,1)$. Thus, we have

$$\nabla Z(\mathbf{y}_k^i; \rho_k)^\top \mathbf{d}_k^i \ge -\frac{\xi(\hat{\alpha}_k^i)}{\hat{\alpha}_k^i}, \quad \forall i \in \mathcal{J}_k. \tag{20}$$

By considering the expression of $Z(\mathbf{x}; \rho_k)$, for all i in $\mathcal{J}_k$ and for $k \in \mathcal{K}_\rho^{\mathtt{x}}$ sufficiently large

$$\nabla Z(\mathbf{y}_k^i; \rho_k)^\top \mathbf{d}_k^i = \left[\nabla f(\mathbf{y}_k^i) + \sum_{\ell \in \mathcal{G}^{\mathtt{log}}} \frac{\rho_k}{-g_\ell(\mathbf{y}_k^i)} \nabla g_\ell(\mathbf{y}_k^i) + \nu \left(\sum_{\ell \in \mathcal{G}^{\mathtt{ext}}} \left(\frac{\max\{g_\ell(\mathbf{y}_k^i), 0\}}{\rho_k} \right)^{\nu-1} \nabla g_\ell(\mathbf{y}_k^i) + \sum_{j=1}^{p} \left(\frac{|h_j(\mathbf{y}_k^i)|}{\rho_k} \right)^{\nu-1} \nabla h_j(\mathbf{y}_k^i) \right) \right]^\top \mathbf{d}_k^i \ge -\frac{\xi(\hat{\alpha}_k^i)}{\hat{\alpha}_k^i}. \tag{21}$$

By Assumption 1, we can extract a further subset of indices $\mathcal{K}_\rho^{\mathtt{x},\mathtt{D}} \subseteq \mathcal{K}_\rho^{\mathtt{x}}$ such that the sets of poll directions $\mathcal{D}_k$ are equal to a fixed set $\mathcal{D}^*$ for all $k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{D}}$. Therefore, we can write

$$\begin{aligned}
&\lim_{\substack{k \to +\infty \\ k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{D}}}} \rho_k = 0 \\
&\lim_{\substack{k \to +\infty \\ k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{D}}}} \alpha_k = 0, \\
&\lim_{\substack{k \to +\infty \\ k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{D}}}} \mathbf{x}_k = \mathbf{x}^*, \\
&\mathcal{J}_k = \mathcal{J}^*, \quad \forall\ k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{D}}, \\
&\mathbf{d}_k^i = \bar{\mathbf{d}}^i, \quad \forall\ i \in \mathcal{J}^*,\ k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{D}}, \\
&\mathcal{D}^* = \{\bar{\mathbf{d}}^i\}_{i \in \mathcal{J}^*}.
\end{aligned}$$

When $k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{D}}$ is sufficiently large, for all $i \in \mathcal{J}^*$, with $\mathbf{y}_k^i = \mathbf{x}_k + t_k^i \hat{\alpha}_k^i \bar{\mathbf{d}}^i$, and $t_k^i \in (0,1)$, since $\hat{\alpha}_k^i \le \alpha_k$, by Theorem 2.7, we have that, $\lim_{\substack{k \to +\infty \\ k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{D}}}} \mathbf{y}_k^i = \mathbf{x}^*$.

Let us define the following approximations to the Lagrange multipliers of each constraint:

- for $\ell = 1, \dots, m$ set $\lambda_\ell(\mathbf{x}; \rho) = \begin{cases} \dfrac{\rho}{-g_\ell(\mathbf{x})}, & \text{if } \ell \in \mathcal{G}_{\mathtt{log}} \\ \nu \left(\dfrac{\max\{g_\ell(\mathbf{x}), 0\}}{\rho} \right)^{\nu-1}, & \text{if } \ell \in \mathcal{G}_{\mathtt{ext}} \end{cases}$

- for $j = 1, \dots, p$ set $\mu_j(\mathbf{x}; \rho) = \nu \left(\dfrac{|h_j(\mathbf{x})|}{\rho} \right)^{\nu-1}$.

The sequences $\{\lambda_\ell(\mathbf{x}_k;\rho_k)\}_{k\in\mathcal{K}_\rho^{\mathtt{x},\mathtt{D}}}$ and $\{\mu_j(\mathbf{x}_k;\rho_k)\}_{k\in\mathcal{K}_\rho^{\mathtt{x},\mathtt{D}}}$, are bounded (see Appendix B). Thus, it is possible to consider $\mathcal{K}_\rho^{\mathtt{x},\mathtt{D},\lambda} \subseteq \mathcal{K}_\rho^{\mathtt{x},\mathtt{D}}$, such that

$$\lim_{\substack{k\to+\infty \\ k\in\mathcal{K}_\rho^{\mathtt{x},\mathtt{D},\lambda}}} \lambda_\ell(\mathbf{x}_k;\rho_k) = \lambda_\ell^*, \qquad \ell = 1,\dots,m \tag{22}$$

$$\lim_{\substack{k\to+\infty \\ k\in\mathcal{K}_\rho^{\mathtt{x},\mathtt{D},\lambda}}} \mu_j(\mathbf{x}_k;\rho_k) = \mu_j^*, \qquad j = 1,\dots,p. \tag{23}$$

Note that, by definition of the multiplier functions, we have $\lambda_\ell^* = 0$ for $\ell \notin \mathcal{I}_+(\mathbf{x}^*)$.

Multiplying (21) by $\rho_k^{\nu-1}$ and taking the limit for $k\to+\infty$, $k\in\mathcal{K}_\rho^{\mathtt{x},\mathtt{D},\lambda}$, recalling $\nu\in(1,2]$, we have that $\rho_k^{\nu-1}\to 0$, so we obtain the following

$$\left(\sum_{\ell\in\mathcal{G}^{\mathtt{ext}}} \nu\max\{g_\ell(\mathbf{x}^*),0\}^{\nu-1}\nabla g_\ell(\mathbf{x}^*) + \sum_{j=1}^p \nu|h_j(\mathbf{x}^*)|^{\nu-1}\nabla h_j(\mathbf{x}^*)\right)^\top \bar{\mathbf{d}}^i \geq 0, \quad \forall \bar{\mathbf{d}}^i \in D^*.$$

From Proposition 2.4 and Assumption 1, since $\varepsilon\in(0,\min\{\bar\varepsilon,\varepsilon^*\}]$, for all $k\in\mathcal{K}_\rho^{\mathtt{x},\mathtt{D},\lambda}$ sufficiently large it holds

$$\mathcal{T}_X(\mathbf{x}^*) = \mathcal{T}_X(\mathbf{x}_k;\varepsilon) = \mathrm{cone}(\mathcal{D}_k\cap\mathcal{T}_X(\mathbf{x}_k;\varepsilon)) = \mathrm{cone}(\mathcal{D}^*).$$

Then, for every $\mathbf{d}\in\mathcal{T}_X(\mathbf{x}^*)$, there exist nonnegative numbers β_i such that

$$\mathbf{d} = \sum_{i\in\mathcal{J}^*}\beta_i\bar{\mathbf{d}}^i, \text{with } \bar{\mathbf{d}}^i\in\mathcal{D}^*. \tag{24}$$

Let us recall that, by assumption, $\mathbf{x}^*$ satisfies the MFCQ conditions, and let $\mathbf{d}$ be the direction satisfying (14) in point (b) of the MFCQ conditions. Then we have,

$$\begin{aligned} 0 \leq \quad & \sum_{i\in\mathcal{J}^*}\beta_i\left(\sum_{\ell\in\mathcal{G}^{\mathtt{ext}}} \nu\max\{g_\ell(\mathbf{x}^*),0\}^{\nu-1}\nabla g_\ell(\mathbf{x}^*) + \sum_{j=1}^p \nu|h_j(\mathbf{x}^*)|^{\nu-1}\nabla h_j(\mathbf{x}^*)\right)^\top \bar{\mathbf{d}}^i = \\ & \left(\sum_{\ell\in\mathcal{G}^{\mathtt{ext}}} \nu\max\{g_\ell(\mathbf{x}^*),0\}^{\nu-1}\nabla g_\ell(\mathbf{x}^*) + \sum_{j=1}^p \nu|h_j(\mathbf{x}^*)|^{\nu-1}\nabla h_j(\mathbf{x}^*)\right)^\top \mathbf{d} = \\ & \left(\sum_{\ell\in\mathcal{I}_+(\mathbf{x}^*)\cap\mathcal{G}^{\mathtt{ext}}} \nu\max\{g_\ell(\mathbf{x}^*),0\}^{\nu-1}\nabla g_\ell(\mathbf{x}^*) + \sum_{j=1}^p \nu|h_j(\mathbf{x}^*)|^{\nu-1}\nabla h_j(\mathbf{x}^*)\right)^\top \mathbf{d}, \end{aligned} \tag{25}$$

where the last equality follows by considering that $\max\{g_\ell(\mathbf{x}^*),0\} = 0$, for all $\ell\in\mathcal{G}_{\mathtt{ext}}\setminus\mathcal{I}_+(\mathbf{x}^*)$.

Again by (14), $\nabla g_\ell(\mathbf{x}^*)^T\mathbf{d} < 0$, for all $\ell\in\mathcal{I}_+(\mathbf{x}^*)$, and $\nabla h_j(\mathbf{x}^*)^\top\mathbf{d} = 0$, for all j. Then, we get $\max\{g_\ell(\mathbf{x}^*),0\} = 0$ for all $\ell\in\mathcal{I}_+(\mathbf{x}^*)\cap\mathcal{G}_{\mathtt{ext}}$, and we can conclude that $g_\ell(\mathbf{x}^*)\leq 0$ for all

$\ell \in \mathcal{G}_{\mathtt{ext}}$. Furthermore, (25) and $g_\ell(\mathbf{x}^*) \le 0$ for all $\ell \in \mathcal{G}_{\mathtt{ext}}$, imply

$$\left(\sum_{\ell\in\mathcal{G}^{\mathtt{ext}}} \nu \max\{g_\ell(\mathbf{x}^*),0\}^{\nu-1}\nabla g_\ell(\mathbf{x}^*) + \sum_{j=1}^{p} \nu|h_j(\mathbf{x}^*)|^{\nu-1}\nabla h_j(\mathbf{x}^*)\right)^\top \mathbf{d} = \left(\sum_{j=1}^{p} \nu|h_j(\mathbf{x}^*)|^{\nu-1}\nabla h_j(\mathbf{x}^*)\right)^\top \mathbf{d} \ge 0, \quad \text{for all } \mathbf{d} \in \mathcal{T}_{\mathcal{X}}(\mathbf{x}^*).$$

Using (13), we get $h_j(\mathbf{x}^*) = 0$ for all $j = 1,\dots,p$. Therefore, recalling that $g_\ell(\mathbf{x}^*) \le 0$ for all $\ell \in \mathcal{G}_{\mathtt{log}}$, we obtain that point $\mathbf{x}^*$ is feasible.

By simple manipulations, inequality (21) can be rewritten as

$$\begin{aligned}\Bigg(&\nabla f(\mathbf{y}_k^i) + \sum_{\ell=1}^{m} \nabla g_\ell(\mathbf{y}_k^i)\lambda_\ell(\mathbf{x}_k;\rho_k)\\ &+ \sum_{\ell=1}^{m} \nabla g_\ell(\mathbf{y}_k^i)(\lambda_\ell(\mathbf{y}_k^i;\rho_k) - \lambda_\ell(\mathbf{x}_k;\rho_k)) + \sum_{j=1}^{p} \nabla h_j(\mathbf{y}_k^i)\mu_j(\mathbf{x}_k;\rho_k)\\ &+ \sum_{j=1}^{p} \nabla h_j(\mathbf{y}_k^i)(\mu_j(\mathbf{y}_k^i;\rho_k) - \mu_j(\mathbf{x}_k;\rho_k))\Bigg)^\top \bar{\mathbf{d}}^i \ge -\frac{\xi(\hat{\alpha}_k^i)}{\hat{\alpha}_k^i}, \quad \forall i \in \mathcal{J}^*\end{aligned} \tag{26}$$

Taking limits for $k \to +\infty$, $k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{D},\lambda}$ and considering (48) and (49) in appendix B, we get:

$$\left(\nabla f(\mathbf{x}^*) + \sum_{\ell=1}^{m} \nabla g_\ell(\mathbf{x}^*)\lambda_\ell^* + \sum_{j=1}^{p} \nabla h_j(\mathbf{x}^*)\mu_j^*\right)^\top \bar{\mathbf{d}}^i \ge 0, \quad \forall i \in \mathcal{J}^*. \tag{27}$$

Again, by (24) and (27), we have, for all $\mathbf{d} \in \mathcal{T}_{\mathcal{X}}(\mathbf{x}^*)$,

$$\left(\nabla f(\mathbf{x}^*) + \sum_{\ell=1}^{m} \nabla g_\ell(\mathbf{x}^*)\lambda_\ell^* + \sum_{j=1}^{p} \nabla h_j(\mathbf{x}^*)\mu_j^*\right)^\top \mathbf{d} = \sum_{i\in\mathcal{J}^*}\beta_i\left(\nabla f(\mathbf{x}^*) + \sum_{\ell=1}^{m} \nabla g_\ell(\mathbf{x}^*)\lambda_\ell^* + \sum_{j=1}^{p} \nabla h_j(\mathbf{x}^*)\mu_j^*\right)^\top \bar{\mathbf{d}}^i \ge 0.$$

Since $\mathbf{x}^*$ is feasible and considering that, by definition of λ_ℓ^* for $\ell \in \mathcal{I}_+(\mathbf{x}^*)$, we have $\lambda_\ell^* g_\ell(\mathbf{x}^*) = 0$, for all $\ell = 1,\dots,m$, $\mathbf{x}^*$ is a KKT point for problem (P), thus concluding the proof. □

4 Implementation details of LOG-DS

In this section, we describe a practical implementation for LOG-DS, based on the original implementation of SID-PSM [13, 14].

4.1 LOG-DS vs SID-PSM

LOG-DS enhances SID-PSM with the ability of handling general constraints using a penalty-interior point method. Thus, the original structure and algorithmic options of SID-PSM implementation were considered. In this subsection, we provide a general overview of the main features of SID-PSM, highlighting the differences with LOG-DS. For more details on SID-PSM, the original references [13, 14] could be used.

The main difference between LOG-DS and SID-PSM is the use of a merit function to address constraints, instead of an extreme barrier approach. In SID-PSM, only feasible points are evaluated, being the function value set equal to $+\infty$ for infeasible ones. LOG-DS allows the violation of nonlinear constraints. The merit function also replaces the original objective function throughout the different algorithmic steps. So, during the Search Step, quadratic polynomial models are built for both the objective and constraints functions and they are analytically combined to produce a model of the merit function. The model is minimized inside a ball with a radius length directly related to the step-size parameter.

The sets of points used in model computation do not require feasibility regarding the nonlinear constraints, always resulting from previous evaluations of the merit function. No function evaluations are spent solely for the purpose of model building. Depending on the number of points available, minimum Frobenius norm models, quadratic interpolation, or regression approaches can be used [13] to compute the model coefficients.

The original SID-PSM only requires a simple decrease for the acceptance of new points, whereas in LOG-DS points are accepted if they satisfy the sufficient decrease condition

$$Z(\mathbf{x}_{k+1};\rho_k) \le Z(\mathbf{x}_k;\rho_k) - \gamma\alpha_k^2,$$

with $\gamma = 10^{-9}$.

The algorithm proceeds with an *opportunistic* Poll Step, accepting the first tentative point satisfying the sufficient decrease condition. Before initiating the polling procedure, previously evaluated points are again used to build a *simplex gradient* [9], which is used as an ascent indicator. Poll directions are reordered according to the largest angle made with this ascent indicator (see [14]). In LOG-DS, the simplex gradient is built on the merit function values, while the original SID-PSM exploits the values of the objective function.

4.2 Penalty parameter details

In the initialization of LOG-DS, we define the two sets of indices $\mathcal{G}_{\texttt{log}}$ and $\mathcal{G}_{\texttt{ext}}$, considering the values of the inequality constraints at the initial point $\mathbf{x}_0 \in \mathcal{X}$:

$$\mathcal{G}_{\texttt{log}} = \{\ell \in \{1,\ldots,m\} \mid g_\ell(\mathbf{x}_0) < 0\} \text{ and } \mathcal{G}_{\texttt{ext}} = \{\ell \in \{1,\ldots,m\} \mid g_\ell(\mathbf{x}_0) \ge 0\}.$$

Moreover, we define two penalty parameters, one corresponding to the logarithmic barrier term ($\rho^{\texttt{log}}$) and another one associated with the penalty exterior component ($\rho^{\texttt{ext}}$), and we initialize them as $\rho_0^{\texttt{log}} = 10^{-1}$ and $\rho_0^{\texttt{ext}} = \frac{1}{\max\{|f(x_0)|,10\}}$. Note that we treat $\rho^{\texttt{ext}}$ as $\rho^{\nu-1}$ in equation (1). Therefore, we can write the merit function as

$$Z(\mathbf{x};\rho_k) = f(\mathbf{x}) - \rho_k^{\mathtt{log}} \sum_{\ell\in\mathcal{G}_{\mathtt{log}}} \log(-g_\ell(\mathbf{x})) + \frac{1}{\rho_k^{\mathtt{ext}}}\left(\sum_{\ell\in\mathcal{G}_{\mathtt{ext}}} (\max\{0, g_\ell(\mathbf{x})\})^\nu + \sum_{j=1}^{p} |h_j(\mathbf{x})|^\nu\right). \quad (28)$$

The penalty parameters are updated only at unsuccessful iterations, considering two different criteria:

$$\alpha_{k+1} \leq \min\{(\rho_k^{\mathtt{log}})^\beta, (g_{\min})_k^2\}, \text{ for updating } \rho_k^{\mathtt{log}}; \quad (29)$$

$$\alpha_{k+1} \leq \min\{(\rho_k^{\mathtt{log}})^\beta, (\rho_k^{\mathtt{ext}})^\beta, (g_{\min})_k^2\}, \text{ when updating } \rho_k^{\mathtt{ext}}. \quad (30)$$

Recall that $(g_{\min})_k$ is the minimum absolute value for the constraints in $\mathcal{G}_{\mathtt{log}}$ at iterate x_k. In the implementation, we have considered $\beta = 1 + 10^{-9}$ and $\nu = 2$. Thus, $\rho^{\nu-1} = \rho$ in equation (1).

If inequality (29) holds, LOG-DS uses the following rule

$$\rho_{k+1}^{\mathtt{log}} = \zeta\rho_k^{\mathtt{log}}, \quad (31)$$

whereas if inequality (30) holds, LOG-DS performs the following update

$$\rho_{k+1}^{\mathtt{ext}} = \zeta\rho_k^{\mathtt{ext}}, \quad (32)$$

in both cases with $\zeta = 10^{-2}$.

The use of two different penalty parameters, for the two different terms of penalization, is a practical need to be able to properly scale the constraints and the different ways they are handled. In particular, in our numerical experience, the logarithmic term seemed not to suffer with the different scales of the objective function, while the exterior penalty seemed to be very sensitive to it. In practice, if the exterior penalty parameter is set too high, the algorithm might be slow at reaching feasible solutions. If it is set too low, the algorithm might not be good at reaching solutions with the best objective function value, even though it might be able to attain feasibility. While scaling the initial exterior penalty parameter with respect to the initial value of the objective function improved our numerical results, it is possible that it might not work for specific problems. Indeed, we are implicitly assuming that the gradient of the objective function is closely related to the objective function value, which might be true for some real problems, but it is certainly not true in general. Scaling the objective function and the constraints for general nonlinear optimization problems is currently an active field of research.

Furthermore, to keep our method aligned with the theoretical framework proposed, in Appendix C we provide some results related to our choices.

5 Numerical experiments

This section is dedicated to the numerical experiments and performance evaluation of the proposed mixed penalty-logarithmic barrier derivative-free optimization algorithm, LOG-DS, on a collection of test problems available in the literature.

We considered the test set used in [10], part of the CUTEst collection [18]. Specifically, we included problems where the number of variables does not exceed 50 and at least one inequality constraint is strictly satisfied by the initialization provided, ensuring $\mathbf{x}_0 \in \mathring{\Omega}_{\mathcal{G}^{\log}}$.

Table 1 details the test set, providing the name, the number of variables, n, the number of inequality constraints, m, the number of inequality constraints treated by the logarithmic barrier $\bar{m} = |\mathcal{G}^{\log}|$, and the number of nonlinear equalities, p, for each problem. Additional information can be found in [10, 18].

5.1 Performance and data profiles

Numerical experiments will be analyzed using performance [16] and data [32] profiles. To provide a brief overview of these tools, consider a set of solvers $\mathcal{S}$ and a set of problems $\mathcal{P}$. Let $t_{p,s}$ represent the number of function evaluations required by solver s to satisfy the convergence test adopted for problem p.

For an accuracy $\tau = 10^{-k}$, where $k \in \{1, 3, 5\}$, we adopted the convergence test:

$$f_M - f(\mathbf{x}) \geq (1 - \tau)(f_M - f_L), \tag{33}$$

where f_M represents the objective function value of the worst feasible point determined by all solvers for problem p, and f_L is the best objective function value obtained by all solvers, corresponding to a feasible point of problem p.

The convergence test given by (33) requires a significant reduction in the objective function value by comparison with the worst feasible point f_M. We assign an infinite value to the objective function at points that violate the feasibility conditions, defined by $c(\mathbf{x}) > 10^{-4}$, where

$$c(\mathbf{x}) = \sum_{i=1}^{m} \max\{g_i(\mathbf{x}), 0\} + \sum_{j=1}^{p} |h_j(\mathbf{x})|.$$

The performance of solver $s \in \mathcal{S}$ is measured by the fraction of problems in which the performance ratio is at most α, given by:

$$\rho_s(\alpha) = \frac{1}{|P|} \left| \left\{ p \in P \mid \frac{t_{p,s}}{\min\left\{t_{p,s'} : s' \in S\right\}} \leq \alpha \right\} \right|.$$

A performance profile provides an overview of how well a solver performs across a set of optimization problems. Particularly relevant is the value $\rho_s(1)$, that reflects the efficiency of the solver, i.e., the percentage of problems for which the algorithm performs the best. Robustness, as the percentage of problems that the algorithm is able to solve, can be perceived for high values of α.

Data profiles focus on the behavior of the algorithm during the optimization process. A data profile measures the percentage of problems that can be solved (given the accuracy τ) with κ estimates of simplex gradients and is defined by:

$$d_s(\kappa) = \frac{1}{|P|} \left| \left\{ p \in P \mid t_{p,s} \leq \kappa \left(n_p + 1\right) \right\} \right|,$$

where n_p represents the dimension of problem p.

Problem	n	m	$\bar{m}$	p
ANTWERP	27	10	2	8
DEMBO7	16	21	16	0
ERRINBAR	18	9	1	8
HS117	15	5	5	0
HS118	15	29	28	0
LAUNCH	25	29	20	9
LOADBAL	31	31	20	11
MAKELA4	21	40	20	0
MESH	33	48	17	24
OPTPRLOC	30	30	28	0
RES	20	14	2	12
SYNTHES2	11	15	1	1
SYNTHES3	17	23	1	2
TENBARS1	18	9	1	8
TENBARS4	18	9	1	8
TRUSPYR1	11	4	1	3
TRUSPYR2	11	11	8	3
HS12	2	1	1	0
HS13	2	1	1	0
HS16	2	2	2	0
HS19	2	2	1	0
HS20	2	3	3	0
HS21	2	1	1	0
HS23	2	5	4	0
HS30	3	1	1	0
HS43	4	3	3	0
HS65	3	1	1	0
HS74	4	5	2	3
HS75	4	5	2	3
HS83	5	6	5	0
HS95	6	4	3	0
HS96	6	4	3	0
HS97	6	4	2	0
HS98	6	4	2	0
HS100	7	4	4	0
HS101	7	6	2	0
HS104	8	6	3	0
HS105	8	1	1	0
HS113	10	8	8	0
HS114	10	11	8	3
HS116	13	15	10	0
S365	7	5	2	0
ALLINQP	24	18	9	3
BLOCKQP1	35	16	1	15
BLOCKQP2	35	16	1	15
BLOCKQP3	35	16	1	15
BLOCKQP4	35	16	1	15
BLOCKQP5	35	16	1	15

Problem	n	m	$\bar{m}$	p
CAMSHAPE	30	94	90	0
CAR2	21	21	5	16
CHARDIS1	28	14	13	0
EG3	31	90	60	1
GAUSSELM	29	36	11	14
GPP	30	58	58	0
HADAMARD	37	93	36	21
HANGING	15	12	8	0
JANNSON3	30	3	2	1
JANNSON4	30	2	2	0
KISSING	37	78	32	12
KISSING1	33	144	113	0
KISSING2	33	144	113	0
LIPPERT1	41	80	64	16
LIPPERT2	41	80	64	16
LUKVLI1	30	28	28	0
LUKVLI10	30	28	14	0
LUKVLI11	30	18	3	0
LUKVLI12	30	21	6	0
LUKVLI13	30	18	3	0
LUKVLI14	30	18	18	0
LUKVLI15	30	21	7	0
LUKVLI16	30	21	13	0
LUKVLI17	30	21	21	0
LUKVLI18	30	21	21	0
LUKVLI2	30	14	7	0
LUKVLI3	30	2	2	0
LUKVLI4	30	14	4	0
LUKVLI6	31	15	15	0
LUKVLI8	30	28	14	0
LUKVLI9	30	6	6	0
MANNE	29	20	10	0
MOSARQP1	36	10	10	0
MOSARQP2	36	10	10	0
NGONE	29	134	106	0
NUFFIELD	38	138	28	0
OPTMASS	36	30	6	24
POLYGON	28	119	94	0
POWELL20	30	30	15	0
READING4	30	60	30	0
SINROSNB	30	58	29	0
SVANBERG	30	30	30	0
VANDERM1	30	59	29	30
VANDERM2	30	59	29	30
VANDERM3	30	59	29	30
VANDERM4	30	59	29	30
YAO	30	30	1	0
ZIGZAG	28	30	5	20

Table 1: Test set selected from the CUTEst collection. Parameters n, m, $\bar{m}$, and p denote, respectively, the number of variables, of inequality constraints, of inequality constraints treated by the logarithmic barrier, and of equality constraints for the given problem.

5.2 Results and analysis

This subsection aims to demonstrate the good numerical performance of the LOG-DS algorithm, when compared against SID-PSM and other state-of-the-art solvers.

5.2.1 Strategies for addressing linear constraints

In this subsection, our focus lies on evaluating the performance of LOG-DS using two distinct approaches for managing linear inequality constraints, other than bounds. The first approach treats each linear inequality constraint (other than bounds) as a general inequality, i.e. addressing it within the proposed penalty-barrier approach. The second approach consists on the use of an extreme barrier method, adjusting the directions so that Assumption 1 is satisfied for the linear constraints (including bounds).

It is important to understand the relevance of comparing the two strategies. The logarithmic barrier approach is well-known for handling linear constraints very efficiently, especially in the presence of a large number. Moreover, conforming the directions to the nearby linear constraints might affect the geometry of the generated points, impacting the quality of the surrogate models built to improve the performance of the algorithm. Using the penalty approach allows us to keep using the coordinate directions, which are known to have good geometry for building linear models. Any other orthonormal basis could be used, though the coordinate directions additionally conform to bound constraints on the variables, so that we can treat them separately from the penalty approach.

The works by Lucidi et al. [30] and Lewis and Torczon [26] propose methods for computing directions conforming to linear inequality constraints but do not consider degeneracy. Abramson et al. [1] provide a detailed algorithm for generating a set directions with the desired property, regardless of whether the constraints are degenerate or not.

In order to use Definition 2.3, of ε-active constraints, we assume a preliminary scaling of the constraints. For this purpose, we multiply the ith constraint by $\|\mathbf{a}_i\|^{-1}$, since the vectors $\mathbf{a}_i$, $i \in \mathcal{I}_{\mathcal{X}}(\mathbf{x}_k)$, are not null (we have that $\|\mathbf{a}_i\| \neq 0$ for all $i \in \mathcal{I}_{\mathcal{X}}(\mathbf{x}_k)$). Therefore, we consider

$$\bar{\mathbf{a}}_i = \frac{\mathbf{a}_i}{\|\mathbf{a}_i\|}, \quad \bar{b}_i = \frac{b_i}{\|\mathbf{a}_i\|}, i \in \mathcal{I}_{\mathcal{X}}(\mathbf{x}_k). \tag{34}$$

The ε-active index set is computed using the matrix $\bar{\mathbf{A}}$, a scaled version of the matrix $\mathbf{A}$, and the vector $\bar{\mathbf{b}}$. Consequently, we have that $\|\bar{\mathbf{a}}_i\| = 1$ for all $i \in \mathcal{I}_{\mathcal{X}}(\mathbf{x}_k)$ and $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \leq \mathbf{b}\} = \{\mathbf{x} \in \mathbb{R}^n \mid \bar{\mathbf{A}}\mathbf{x} \leq \bar{\mathbf{b}}\}$.

To compute the set of directions $\mathcal{D}_k$ that conforms to the geometry of the nearby constraints, we use the algorithm proposed in [1, Alg. 4.4.2]. The latter is divided into two parts: the first constructs the index set corresponding to the ε-active non-redundant constraints, and the second computes the set of directions $\mathcal{D}_k$, which include the generators of the cone $\mathcal{T}_{\mathcal{X}}(\mathbf{x}_k, \varepsilon)$. For identifying a constraint as being ε-active, we considered $\varepsilon = 10^{-5}$. When linear constraints (other than bounds) are included in the merit function, we used $\mathcal{D}_k = [\mathbf{1} \ -\mathbf{1} \quad \mathbb{I}_n \ -\mathbb{I}_n]$ as the default set of directions, for every k.

Figure 1 depicts the performance of LOG-DS using the two strategies described before, considering a maximum number of 2000 function evaluations and a minimum stepsize tolerance equal to 10^{-8}.

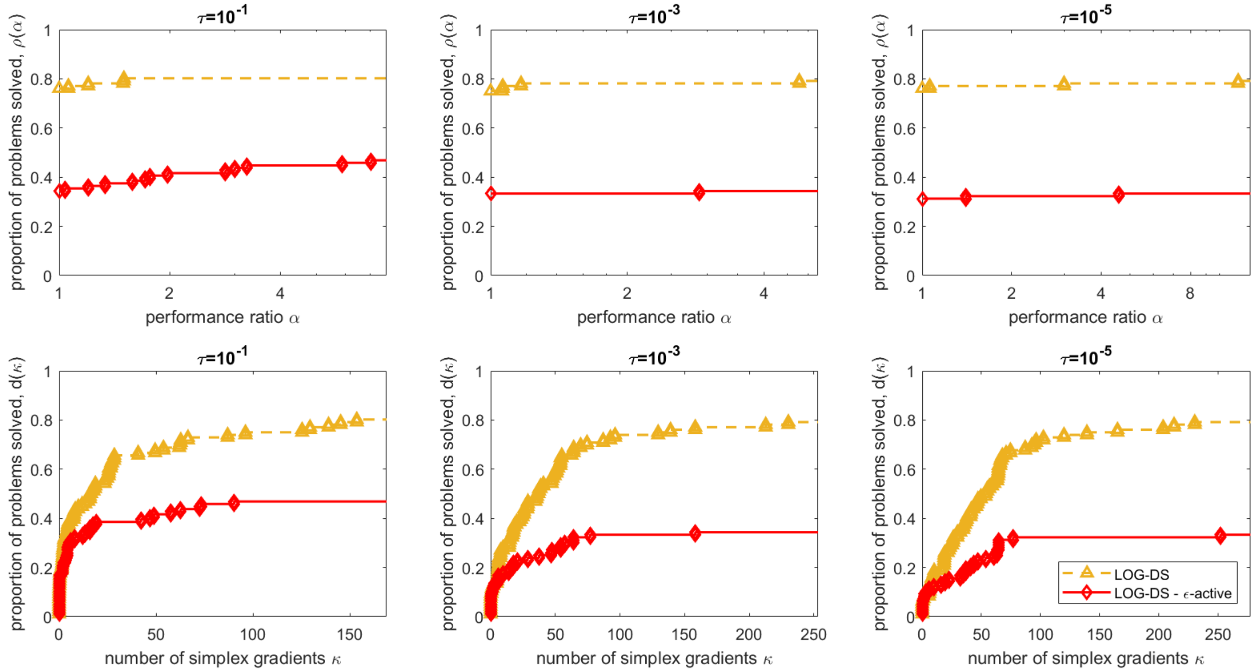

Figure 1: Performance (on top) and data (on bottom) profiles comparing LOG-DS using two different approaches to address linear inequality constraints.

As we can see, the performance of the LOG-DS algorithm when addressing the linear inequality constraints within the penalty approach outperforms the competing strategy, addressing the linear constraints directly. Therefore, in the rest of the work, the experiments will be carried out using the winning strategy. Note that the significant difference in the performance might be due to the specific choice of the tested problems and/or the specific strategy used to conform the directions to the linear constraints, rather than to a flaw in the approach by itself.

5.2.2 Comparison with the original SID-PSM algorithm

We are proposing an alternative strategy to address constraints within the SID-PSM algorithm. Thus, we start by illustrating that the use of a mixed penalty-logarithmic barrier is competitive against the extreme barrier approach. The latter is typically used for problems without equality constraints and for which a feasible point is given as initialization, so we selected a subset of problems satisfying these conditions. The subset consists of a total of 28 problems, highlighted in Table 1.

Figure 2 presents the comparison between LOG-DS, which exploits the mixed penalty-logarithmic barrier, and the original SID-PSM, which employs an extreme barrier approach. The default values of SID-PSM were considered for both algorithms, allowing a maximum number of 2000 function evaluations and a minimum stepsize equal to 10^{-8}.

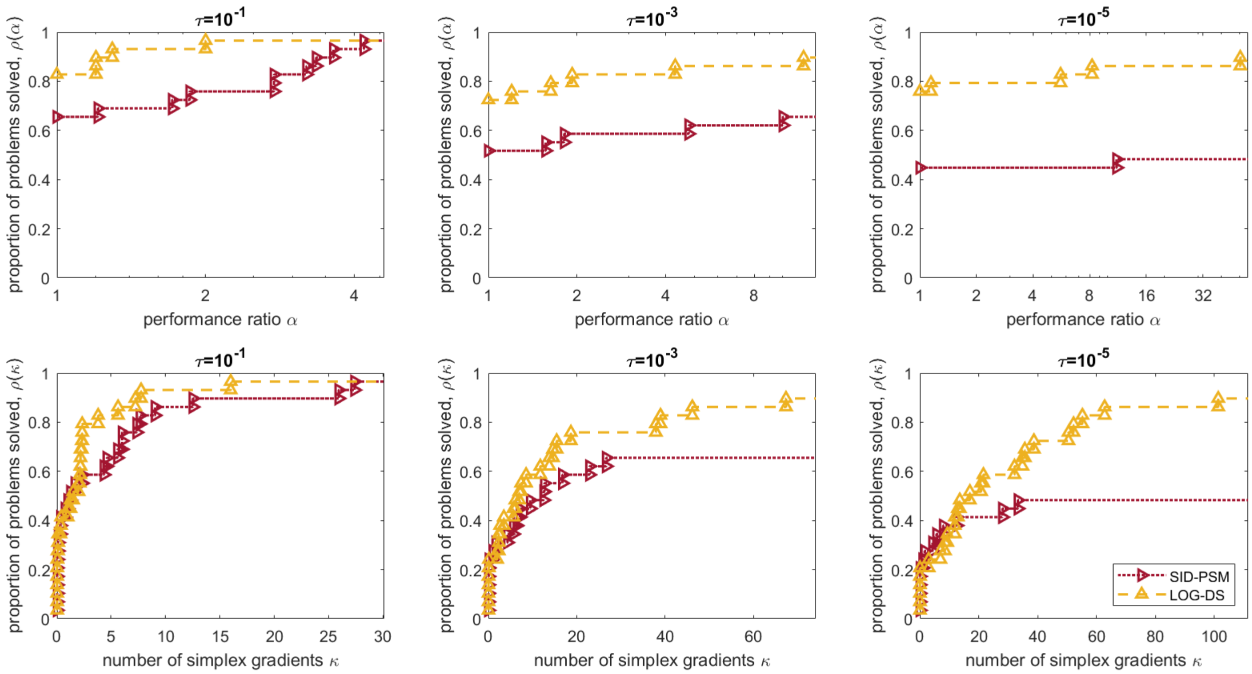

Figure 2: Performance (on top) and data (on bottom) profiles comparing LOG-DS and SID-PSM.

As Figure 2 shows, LOG-DS presents a better performance than SID-PSM, especially when a higher precision is considered. Furthermore, the possibility of initializing LOG-DS with infeasible points, regarding some of the constraints, allows to handle a wider class of practical problems.

5.2.3 Comparison with state-of-the-art solvers

This subsection focuses on comparing LOG-DS against state-of-the-art derivative-free optimization solvers that are able to address general nonlinear constraints. Comparisons were made with MADS [7], implemented in the well-known NOMAD package (version 4), which can be freely obtained at `https://www.gerad.ca/en/software/nomad`. Additionally, the X-LOG-DFL algorithm [10], available through the DFL library as the LOGDFL package at `https://github.com/DerivativeFreeLibrary/LOGDFL`, was also tested. Comparison with LOG-DFL [10] is particularly relevant since it uses the same merit function of LOG-DS. Default settings were considered for all codes and results, reported in Figure 3, were obtained for a maximum budget of 2000 function evaluations.

It can be observed that LOG-DS presents the best performance, for any of the three precision levels considered, both in terms of efficiency and robustness, across the different computational budgets.

Figure 4 compares the different solvers considering the subset of problems with only inequality constraints (61 out of 96 problems), again allowing a maximum of 2000 function

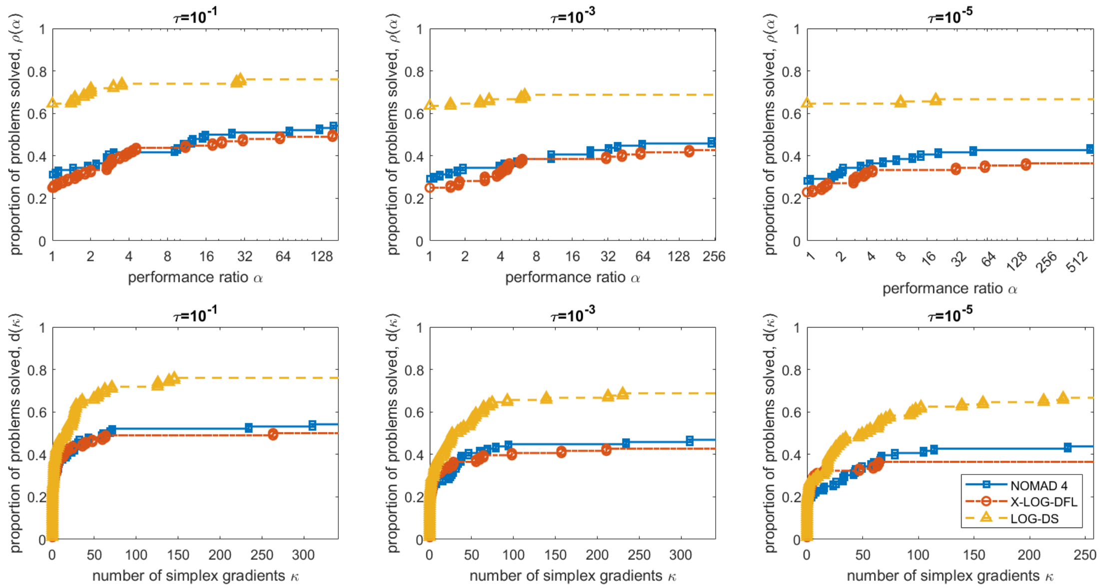

Figure 3: Performance (on top) and data (on bottom) comparing LOG-DS, NOMAD, and X-LOG-DFL, on the complete problem collection.

evaluations. Once more, LOG-DS is clearly the solver with the best performance.

In summary, considering the outcomes of the different numerical experiments, we can conclude that LOG-DS is the most efficient and robust solver across different scenarios, making it the top-performing choice.

6 Conclusions

The primary objective of this work was to extend the approach introduced in [10] to generalized pattern search, allowing to efficiently address nonlinear constraints.

To accomplish it, we adapted the SID-PSM algorithm, a generalized pattern search method, where polynomial models are used both at the search and at the poll steps to improve the numerical performance. We proposed a new algorithm, LOG-DS, that keeps the basic algorithmic features of SID-PSM but uses a mixed penalty-logarithmic barrier merit function to address general nonlinear and linear constraints.

Under standard assumptions, not requiring convexity of the functions defining the problem, convergence was established towards stationary points. Furthermore, an extensive numerical experimentation allowed to compare the performance of LOG-DS with several state-of-the-art solvers on a large set of test problems from the CUTEst collection. The numerical results indicate the robustness, efficiency, and overall effectiveness of the proposed algorithm.

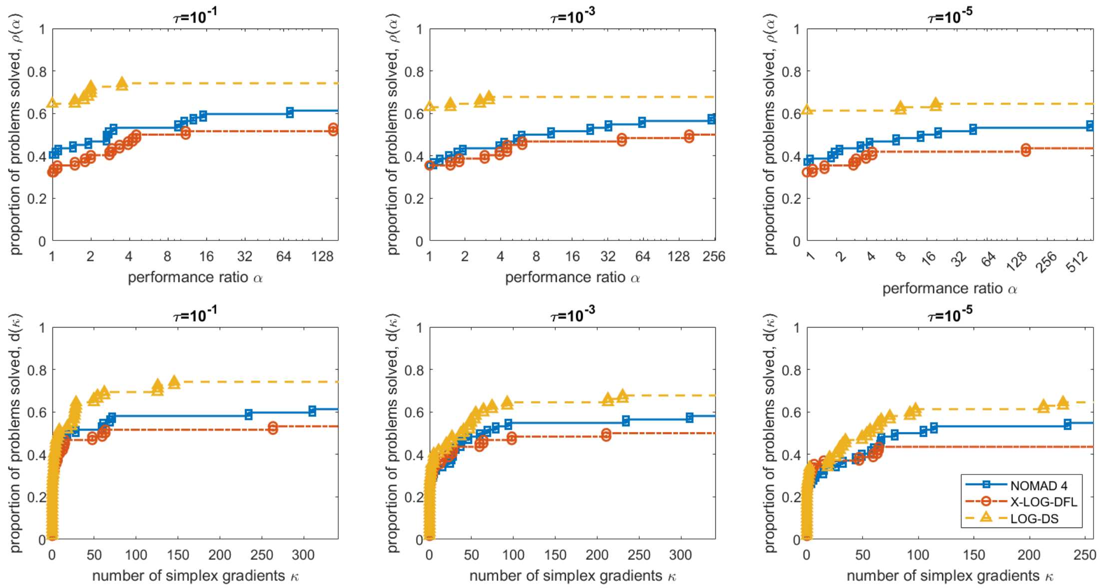

Figure 4: Performance (on top) and data (on bottom) profiles comparing LOG-DS, NOMAD, and LOG-DFL, on the subset of problems with only inequality constraints.

A Appendix

In this Appendix, we present two technical results involving inequalities for real numbers and exponents. The first lemma provides a bound for the sum of powers of positive real numbers.

Lemma A.1. *[33, Equation (2), p. 37] Let $a \in \mathbb{R}_+$, $b \in \mathbb{R}_+$, $a + b > 0$ and $p \in [0, 1]$. It results*

$$(a + b)^p \leq a^p + b^p.$$

The next result shows bounds concerning the difference of powers of absolute values and maximum functions, which are fundamental for proving the boundedness of the Lagrange multipliers.

Lemma A.2. *Let $a \in \mathbb{R}$, $b \in \mathbb{R}$ and $p \in [0, 1]$. The following inequalities hold*

$$||a|^p - |b|^p| \leq |a - b|^p, \tag{35}$$

and

$$|\max\{a, 0\}^p - \max\{b, 0\}^p| \leq |a - b|^p. \tag{36}$$

Proof. To prove (35) consider

$$\begin{aligned} |a|^p &= |a - b + b|^p \\ \text{(Triangular inequality)} &\leq (|a - b| + |b|)^p \\ \text{(Lemma A.1)} &\leq |a - b|^p + |b|^p \end{aligned} \tag{37}$$

On the other hand,

$$\begin{aligned} |b|^p &= |b-a+a|^p \\ \text{(Triangular inequality)} \quad &\leq (|b-a|+|a|)^p \\ \text{(Lemma A.1)} \quad &\leq |a-b|^p + |a|^p \end{aligned} \tag{38}$$

Then, from inequalities (37) and (38) we derive inequality (35).
Now, to prove (36), we will analyze four different cases:

i) If $a \leq 0$ and $b \leq 0$, the result holds trivially.

ii) If $a \leq 0$ and $b > 0$, then $|-b^p| = b^p \leq |b-a|^p = |a-b|^p$.

iii) If $a > 0$ and $b \leq 0$, then $|a|^p = a^p \leq |a-b|^p$.

iv) If $a > 0$ and $b > 0$, using (35) we conclude that:

$$|\max\{a,0\}^p - \max\{b,0\}^p| = |a^p - b^p| = ||a|^p - |b|^p| \leq |a-b|^p.$$

□

B Appendix

First, we present a result regarding a property of sequences of scalars, which is used to prove the boundedness of the Lagrange multipliers.

Lemma B.1. *Let $\{a_k^i\}_{k\in\mathbb{N}}$, $i = 1,\dots,\ell$, be sequences of scalars. Two cases can occur:*

(i) It results $\lim_{k\to+\infty} a_k^i = 0$, $i = 1,\dots,\ell$. In particular, all the sequences are bounded;

(ii) An index $j \in \{1,\dots,\ell\}$, an infinite index set $\mathcal{K}_j \subseteq \{0,1,\dots\}$ and a positive scalar $\bar{a}_j$ exist such that

$$|a_k^j| > \bar{a}_j > 0, \quad \forall k \in \mathcal{K}_j,$$

i.e., at least one sequence is not convergent to zero. Then, there exists an index $s \in \{1,\dots,\ell\}$ and an infinite subset $\mathcal{K} \subseteq \mathbb{N}$ such that:

$$\lim_{\substack{k\to+\infty,\\ k\in\mathcal{K}}} \frac{a_k^i}{|a_k^s|} = z_i, \quad |z_i| < +\infty, \quad i = 1,\dots,\ell, \tag{39}$$

i.e., all the sequences $\left\{\frac{a_k^i}{|a_k^s|}\right\}_{k\in\mathcal{K}}$ are bounded.

Proof. Point (i) directly follows from the properties of convergent sequences.

Then, let us now assume that at least one sequence is not convergent to zero. If this is the case, we can reorder the sequences in such a way that:

- $\lim_{k\to+\infty} a_k^1 = \lim_{k\to+\infty} a_k^2 = \cdots = \lim_{k\to+\infty} a_k^{r-1} = 0$;
- $\{a_k^i\}$, $i = r, \ldots, \ell$, are not convergent to zero.

We now prove that an index $s \in \{r, \ldots, \ell\}$ and an infinite index set $\hat{\mathcal{K}}$ exist such that

$$\lim_{\substack{k\to+\infty,\\ k\in\hat{\mathcal{K}}}} \frac{a_k^i}{|a_k^s|} = z_i, \quad |z_i| < +\infty, \quad i = r, \ldots, \ell.$$

This is (obviously) true when $\ell = r$. Indeed, when $\ell = r$, there is $\mathcal{K}_r$ such that

$$\lim_{\substack{k\to+\infty,\\ k\in\mathcal{K}_r}} \frac{a_k^r}{|a_k^r|} = \pm 1.$$

Furthermore, we have

$$\lim_{\substack{k\to+\infty,\\ k\in\mathcal{K}_r}} \frac{a_k^i}{|a_k^r|} = 0, \quad i = 1, \ldots, r-1.$$

The thesis follows by choosing $s = r$ and $\mathcal{K} = \hat{\mathcal{K}} = \mathcal{K}_r$.

Now, we prove the thesis by induction on ℓ. Then, we have sequences $\{a_k^i\}$, $i = 1, \ldots, \ell-1$ such that:

- $\lim_{k\to+\infty} a_k^1 = \lim_{k\to+\infty} a_k^2 = \cdots = \lim_{k\to+\infty} a_k^{r-1} = 0$;
- $\{a_k^i\}$, $i = r, \ldots, \ell-1$, are not convergent to zero,

and an index $\hat{\imath} \in \{r, \ldots, \ell-1\}$ and an infinite index set $\hat{\mathcal{K}}$ exist such that

$$\lim_{\substack{k\to+\infty,\\ k\in\hat{\mathcal{K}}}} \frac{a_k^i}{|a_k^{\hat{\imath}}|} = z_i, \quad |z_i| < +\infty, \quad i = r, \ldots, \ell-1.$$

1. If $\lim_{k\to+\infty} a_k^\ell = 0$, then we have

   $$\lim_{\substack{k\to+\infty,\\ k\in\hat{\mathcal{K}}}} \frac{a_k^j}{|a_k^{\hat{\imath}}|} = z_j, \quad j = 1, \ldots, \ell-1$$
   $$\lim_{\substack{k\to+\infty,\\ k\in\hat{\mathcal{K}}}} \frac{a_k^\ell}{|a_k^{\hat{\imath}}|} = 0;$$

   and the thesis follows by choosing $s = \hat{\imath}$ and $\mathcal{K} = \hat{\mathcal{K}}$.

2. Suppose now that $\{a_k^\ell\}$ is not convergent to zero. In this situation, two subcases can occur:

(a) the sequence $\left\{\frac{a_k^\ell}{a_k^{\hat{i}}}\right\}_{k\in\hat{\mathcal{K}}}$ is bounded;

(b) an infinite index set $\mathcal{K}_1 \subseteq \hat{\mathcal{K}}$ exists such that $\left\{\frac{a_k^\ell}{a_k^{\hat{i}}}\right\}_{k\in\mathcal{K}_1}$ is unbounded.

In the first case, an infinite index set $\mathcal{K}_2 \subseteq \hat{\mathcal{K}}$ exists such that $\left\{\frac{a_k^\ell}{a_k^{\hat{i}}}\right\}_{k\in\mathcal{K}_2}$ is convergent. Then, the thesis follows by taking $s = \hat{i}$ and $\mathcal{K} = \mathcal{K}_2$.

In the second case, we have

$$\lim_{\substack{k\to+\infty,\\ k\in\mathcal{K}_1}} \frac{a_k^{\hat{i}}}{a_k^\ell} = 0.$$

Furthermore, for every $i = 1, \ldots, \ell - 1$, we have

$$\lim_{\substack{k\to+\infty,\\ k\in\mathcal{K}_1}} \frac{a_k^i}{a_k^\ell} = \lim_{\substack{k\to+\infty,\\ k\in\mathcal{K}_1}} \frac{a_k^i}{a_k^{\hat{i}}}\frac{a_k^{\hat{i}}}{a_k^\ell} = 0,$$

and again the thesis is proved with $s = \ell$ and $\mathcal{K} = \mathcal{K}_1$.

Thus, the proof is concluded. $\square$

In the next result, we establish the boundedness of the subsequence of Lagrange multipliers

Theorem B.2. *In the conditions of Theorem 3.5, the subsequences* $\{\lambda_\ell(\mathbf{x}_k;\rho_k)\}_{k\in\mathcal{K}_\rho^{\mathbf{x}}}$, $\ell = 1, \ldots, m$ *and* $\{\mu_j(\mathbf{x}_k;\rho_k)\}_{k\in\mathcal{K}_\rho^{\mathbf{x}}}$, $j = 1, \ldots, p$, *defined as:*

$$\lambda_\ell(\mathbf{x}_k;\rho_k) = \begin{cases} \dfrac{\rho_k}{-g_\ell(\mathbf{x}_k)}, & \text{if } \ell \in \mathcal{G}^{\texttt{log}} \\ \nu\left(\dfrac{\max\{g_\ell(\mathbf{x}_k), 0\}}{\rho_k}\right)^{\nu-1}, & \text{if } \ell \in \mathcal{G}^{\texttt{ext}} \end{cases}$$

$$\mu_j(\mathbf{x}_k;\rho_k) = \nu\left(\frac{|h_j(\mathbf{x}_k)|}{\rho_k}\right)^{\nu-1}, \quad j = 1, \ldots, p,$$

are bounded.

Proof. Recalling the expressions of $\lambda_\ell(\mathbf{x};\rho)$, $\ell = 1, \ldots, m$, and of $\mu_j(\mathbf{x};\rho)$, $j = 1, \ldots, p$, we can rewrite inequality (21) as

$$\begin{aligned}&\Bigg(\nabla f(\mathbf{y}_k^i) + \sum_{\ell=1}^m \lambda_\ell(\mathbf{y}_k^i;\rho_k)\nabla g_\ell(\mathbf{y}_k^i) + \\ &\qquad \sum_{j=1}^p \mu_j(\mathbf{y}_k^i;\rho)\nabla h_j(\mathbf{y}_k^i)\Bigg)^\top \mathbf{d}_k^i \geq -\frac{\xi(\hat{\alpha}_k^i)}{\hat{\alpha}_k^i}, \quad \forall\, i \in \mathcal{J}_k \text{ and } k \in \mathcal{K}_\rho^{\mathbf{x}},\end{aligned} \tag{40}$$

where $\mathbf{y}_k^i = \mathbf{x}_k + t_k^i\hat{\alpha}_k^i\mathbf{d}_k^i$, with $t_k^i \in (0,1)$ and $\hat{\alpha}_k^i \leq \alpha_k$.

We start by establishing that

$$\lim_{\substack{k\to+\infty \\ k\in\mathcal{K}_\rho^{\mathbf{x}}}} |\lambda_\ell(\mathbf{x}_k;\rho_k) - \lambda_\ell(\mathbf{y}_k^i;\rho_k)| = 0, \quad \ell \in \mathcal{G}^{\texttt{log}}, \quad \forall\, i \in \mathcal{J}_k \tag{41}$$

In fact,

$$\begin{aligned} \left|\frac{\rho_k}{-g_\ell(\mathbf{x}_k)} - \frac{\rho_k}{-g_\ell(\mathbf{y}_k^i)}\right| &= \rho_k \left|\frac{g_\ell(\mathbf{x}_k) - g_\ell(\mathbf{y}_k^i)}{(-g_\ell(\mathbf{y}_k^i))(-g_\ell(\mathbf{x}_k))}\right| \\ &= \rho_k \frac{\left|\nabla g_\ell(\mathbf{u}_k^\ell)^\top \left(\mathbf{x}_k - \mathbf{y}_k^i\right)\right|}{|g_\ell(\mathbf{y}_k^i)||g_\ell(\mathbf{x}_k)|} \\ &\le \rho_k \frac{\left\|\nabla g_\ell(\mathbf{u}_k^\ell)\right\| \left\|\mathbf{y}_k^i - \mathbf{x}_k\right\|}{|g_\ell(\mathbf{y}_k^i)||g_\ell\,(\mathbf{x}_k)\,|}, \end{aligned} \tag{42}$$

where $\mathbf{u}_k^\ell = \mathbf{x}_k + \tilde{t}_k^\ell(\mathbf{y}_k^i - \mathbf{x}_k)$ with $\tilde{t}_k^\ell \in (0,1)$.

Then, there is $c_1 > 0$ such that

$$\begin{aligned} \rho_k \left\|\nabla g_\ell(\mathbf{u}_k^\ell)\right\| \frac{\left\|\mathbf{y}_k^i - \mathbf{x}_k\right\|}{|g_\ell(\mathbf{y}_k^i)||g_\ell\,(\mathbf{x}_k)\,|} &\le \rho_k c_1 \frac{\|\mathbf{y}_k^i - \mathbf{x}_k\|}{|g_\ell(\mathbf{y}_k^i)||g_\ell(\mathbf{x}_k)|} \\ &= \rho_k c_1 \frac{\|\mathbf{x}_k + t_k^i \hat{\alpha}_k^i \mathbf{d}_k^i - \mathbf{x}_k\|}{|g_\ell(\mathbf{y}_k^i)||g_\ell(\mathbf{x}_k)|} \\ &= \rho_k c_1 \frac{\|t_k^i \hat{\alpha}_k^i \mathbf{d}_k^i\|}{|g_\ell(\mathbf{y}_k^i)||g_\ell(\mathbf{x}_k)|} \\ &= \rho_k c_1 \frac{t_k^i \hat{\alpha}_k^i \|\mathbf{d}_k^i\|}{|g_\ell(\mathbf{y}_k^i)||g_\ell(\mathbf{x}_k)|}. \end{aligned} \tag{43}$$

Now, we will prove that there is another constant $c_2 > 0$ such that

$$\frac{1}{|g_\ell(\mathbf{y}_k^i)|} \le c_2 \frac{1}{|g_\ell(\mathbf{x}_k)|}. \tag{44}$$

Suppose, in order to arrive to a contradiction, that c_2 does not exist. This would imply that there exists $\mathcal{K}_\rho^{\mathbf{x},\mathbf{g}} \subseteq \mathcal{K}_\rho^{\mathbf{x}} \subseteq \mathcal{K}_\rho$ such that

$$\lim_{\substack{k\to+\infty \\ k\in\mathcal{K}_\rho^{\mathbf{x},\mathbf{g}}}} \frac{\frac{1}{|g_\ell(\mathbf{y}_k^i)|}}{\frac{1}{|g_\ell(\mathbf{x}_k)|}} = \lim_{\substack{k\to+\infty \\ k\in\mathcal{K}_\rho^{\mathbf{x},\mathbf{g}}}} \frac{|g_\ell(\mathbf{x}_k)|}{|g_\ell(\mathbf{y}_k^i)|} = +\infty. \tag{45}$$

Let us consider the case where

$$\lim_{\substack{k\to+\infty \\ k\in\mathcal{K}_\rho^{\mathbf{x},\mathbf{g}}}} |g_\ell(\mathbf{x}_k)| = 0.$$

Since $g_\ell(\mathbf{x}_k) < 0$ and $g_\ell(\mathbf{y}_k^i) < 0$ for all $k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{g}}$, by (45) there exists $\overline{k} \in \mathbb{N}$ such that, for all $k \geq \overline{k}$, $k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{g}}$, we have

$$-g_\ell(\mathbf{x}_k) > -g_\ell(\mathbf{y}_k^i) = -g_\ell(\mathbf{x}_k + t_k^i \hat{\alpha}_k^i \mathbf{d}_k^i).$$

Using the Lipschitz continuity of g_ℓ, $\ell = 1, \ldots, m$ and the fact that $\|\mathbf{d}_k^i\| = 1$ for all $i \in \mathcal{J}_k$, we get

$$-g_\ell(\mathbf{x}_k + t_k^i \hat{\alpha}_k^i \mathbf{d}_k^i) \geq -g_\ell(\mathbf{x}_k) - L_{g_\ell} \|t_k^i \hat{\alpha}_k^i \mathbf{d}_k^i\| = -g_\ell(\mathbf{x}_k) - L_{g_\ell} t_k^i \hat{\alpha}_k^i.$$

The definition of $\mathcal{K}_\rho$ guarantees that

$$\alpha_{k+1} \leq \min\{\rho_k^\beta, (g_{\mathtt{min}})_k^2\}, \qquad \alpha_{k+1} = \theta_\alpha \alpha_k,$$

so that

$$\alpha_k \leq \frac{\min\{\rho_k^\beta, (g_{\mathtt{min}})_k^2\}}{\theta_\alpha} \tag{46}$$

Hence, since $\hat{\alpha}_k^i \leq \alpha_k$, we have

$$-g_\ell(\mathbf{x}_k) - L_{g_\ell} t_k^i \hat{\alpha}_k^i \geq -g_\ell(\mathbf{x}_k) - L_{g_\ell} t_k^i \frac{1}{\theta_\alpha} (g_\ell(\mathbf{x}_k))^2, \quad \forall k \geq \overline{k},\ k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{g}}$$

Thus,

$$\begin{aligned} \lim_{\substack{k\to+\infty \\ k\in\mathcal{K}_\rho^{\mathtt{x},\mathtt{g}}}} \frac{-g_\ell(\mathbf{x}_k)}{-g_\ell(\mathbf{y}_k^i)} &= \lim_{\substack{k\to+\infty \\ k\in\mathcal{K}_\rho^{\mathtt{x},\mathtt{g}}}} \frac{-g_\ell(\mathbf{x}_k)}{-g_\ell(\mathbf{x}_k + t_k^i \hat{\alpha}_k^i \mathbf{d}_k^i)} \\ &\leq \lim_{\substack{k\to+\infty \\ k\in\mathcal{K}_\rho^{\mathtt{x},\mathtt{g}}}} \frac{-g_\ell(\mathbf{x}_k)}{-g_\ell(\mathbf{x}_k) - L_{g_\ell} t_k^i \frac{1}{\theta_\alpha} (g_\ell(\mathbf{x}_k))^2} = 1, \end{aligned}$$

which leads to a contradiction, proving (44).

Now, by considering the other case

$$\lim_{\substack{k\to+\infty \\ k\in\mathcal{K}_\rho^{\mathtt{x},\mathtt{g}}}} |g_\ell(\mathbf{x}_k)| = c < +\infty,$$

we have

$$\lim_{\substack{k\to+\infty \\ k\in\mathcal{K}_\rho^{\mathtt{x},\mathtt{g}}}} \frac{-g_\ell(\mathbf{x}_k)}{-g_\ell(\mathbf{y}_k^i)} = \lim_{\substack{k\to+\infty \\ k\in\mathcal{K}_\rho^{\mathtt{x},\mathtt{g}}}} \frac{-g_\ell(\mathbf{x}_k)}{-g_\ell(\mathbf{x}_k + t_k^i \hat{\alpha}_k^i \mathbf{d}_k^i)} < +\infty.$$

Again, this leads to a contradiction, proving (44).

Hence, the existence of the constant $c_2 > 0$, (44), and recalling that $\hat{\alpha}_k^i \leq \alpha_k$, allow us to write

$$\frac{\rho_k c_1 t_k^i \hat{\alpha}_k^i}{|g_\ell(\mathbf{y}_k^i)||g_\ell(\mathbf{x}_k)|} \leq \frac{\rho_k c_1 c_2 t_k^i \alpha_k}{|g_\ell(\mathbf{x}_k)|^2}.$$

The instructions of Step 3 imply that $\mathbf{x}_{k+1} = \mathbf{x}_k$, so that $(g_{\min})_k = \min_{\ell\in\mathcal{G}^{\mathtt{log}}}\{|g_\ell(\mathbf{x}_{k+1})|\} = \min_{\ell\in\mathcal{G}^{\mathtt{log}}}\{|g_\ell(\mathbf{x}_k)|\}$. Recalling (46), we get

$$\frac{\rho_k\alpha_k}{(g_{\mathtt{min}})_k^2} \le \frac{\rho_k}{\theta_\alpha}.$$

Then, recalling Theorem 2.7, (41) is proved.

Furthermore,

$$\lim_{\substack{k\to+\infty\\ k\in\mathcal{K}_\rho^{\mathbf{x}}}} |\lambda_\ell(\mathbf{x}_k;\rho_k) - \lambda_\ell(\mathbf{y}_k^i;\rho_k)| = 0, \quad \ell\in\mathcal{G}^{\mathtt{ext}}, \quad \forall\, i\in\mathcal{J}_k \tag{47}$$

can be established. In fact,

$$\begin{aligned}
\left|\frac{\nu}{\rho_k^{\nu-1}}(\max\{g_\ell(\mathbf{x}_k),0\})^{\nu-1} \right.&-\left.\frac{\nu}{\rho_k^{\nu-1}}\left(\max\{g_\ell(\mathbf{y}_k^i),0\}\right)^{\nu-1}\right| = \\
&= \frac{\nu}{\rho_k^{\nu-1}}\left|\max\{g_\ell(\mathbf{x}_k),0\}^{\nu-1} - \max\{g_\ell(\mathbf{x}_k)+\nabla g_\ell(\mathbf{u}_k^\ell)^\top(\mathbf{x}_k-\mathbf{y}_k^i),0\}^{\nu-1}\right| \\
(\text{Lemma A.2} - (36)\ )\quad &\le \frac{\nu}{\rho_k^{\nu-1}}\left|g_\ell(\mathbf{x}_k) - g_\ell(\mathbf{x}_k) - \nabla g_\ell(\mathbf{u}_k^\ell)^\top(\mathbf{x}_k-\mathbf{y}_k^i)\right|^{\nu-1} \\
&= \frac{\nu}{\rho_k^{\nu-1}}|\nabla g_\ell(\mathbf{u}_k^\ell)^\top(\mathbf{x}_k-\mathbf{y}_k^i)|^{\nu-1} \le \frac{\nu}{\rho_k^{\nu-1}}\left\|\nabla g_\ell(\mathbf{u}_k^\ell)\right\|^{\nu-1}\left\|\mathbf{x}_k-\mathbf{y}_k^i\right\|^{\nu-1} \\
&\le c_3\frac{\nu}{\rho_k^{\nu-1}}\left\|(\mathbf{x}_k-(\mathbf{x}_k+t_k^i\hat{\alpha}_k^i\mathbf{d}_k^i))\right\|^{\nu-1} = c_3\frac{\nu}{\rho_k^{\nu-1}}(t_k^i\hat{\alpha}_k^i)^{\nu-1}\left\|\mathbf{d}_k^i\right\|^{\nu-1} \\
&\le c_3\frac{\nu}{\rho_k^{\nu-1}}(\hat{\alpha}_k^i)^{\nu-1}\left\|\mathbf{d}_k^i\right\|^{\nu-1} \le c_3\nu\left(\frac{\alpha_k}{\rho_k}\right)^{\nu-1}\left\|\mathbf{d}_k^i\right\|^{\nu-1} \\
&\le c_3\nu\left(\frac{\rho_k^{\beta-1}}{\theta_\alpha}\right)^{\nu-1}\left\|\mathbf{d}_k^i\right\|^{\nu-1} = c_3\nu\theta_\alpha^{1-\nu}\rho_k^{(\beta-1)(\nu-1)}\left\|\mathbf{d}_k^i\right\|^{\nu-1},
\end{aligned}$$

where $\mathbf{u}_k^\ell = \mathbf{x}_k + \tilde{t}_k^\ell(\mathbf{y}_k^i - \mathbf{x}_k)$ with $\tilde{t}_k^\ell \in (0,1)$, and $c_3 > 0$. Thus, using $\beta > 1$, $\nu\in(1,2]$, $\|\mathbf{d}_k^i\| = 1$ for all $i\in\mathcal{J}_k$, and recalling Theorem 2.7, (47) is proved. Therefore, we have that

$$\lim_{\substack{k\to+\infty\\ k\in\mathcal{K}_\rho^{\mathbf{x}}}} |\lambda_\ell(\mathbf{x}_k;\rho_k) - \lambda_\ell(\mathbf{y}_k^i;\rho_k)| = 0, \quad \ell = 1,\dots,m, \quad \forall\, i\in\mathcal{J}_k \tag{48}$$

Let us now prove that

$$\lim_{\substack{k\to+\infty\\ k\in\mathcal{K}_\rho^{\mathbf{x}}}} |\mu_j(\mathbf{x}_k;\rho_k) - \mu_j(\mathbf{y}_k^i;\rho_k)| = 0, \quad j = 1,\dots,p, \quad \forall i\in\mathcal{J}_k \tag{49}$$

In fact, recalling that $\nu \in (1,2]$ so that $\nu - 1 \in (0,1]$, we have

$$\begin{aligned}
\left|\nu\left|\frac{h_j(\mathbf{x}_k)}{\rho_k}\right|^{\nu-1} - \nu\left|\frac{h_j(\mathbf{y}_k^i)}{\rho_k}\right|^{\nu-1}\right| &= \frac{\nu}{\rho_k^{\nu-1}}\left|\left|h_j(\mathbf{x}_k)\right|^{\nu-1} - \left|h_j(\mathbf{x}_k) + \nabla h_j(\mathbf{u}_k^j)^\top(\mathbf{y}_k^i - \mathbf{x}_k)\right|^{\nu-1}\right| \\
\text{(Lemma A.2 – (35))} \quad &\le \frac{\nu}{\rho_k^{\nu-1}}\left|h_j(\mathbf{x}_k) - h_j(\mathbf{x}_k) - \nabla h_j(\mathbf{u}_k^j)^\top(\mathbf{y}_k^i - \mathbf{x}_k)\right|^{\nu-1} \\
&= \frac{\nu}{\rho_k^{\nu-1}}\left|\nabla h_j(\mathbf{u}_k^j)^\top(\mathbf{y}_k^i - \mathbf{x}_k)\right|^{\nu-1} \\
&\le \frac{\nu}{\rho_k^{\nu-1}}\left\|\nabla h_j(\mathbf{u}_k^j)\right\|^{\nu-1}\left\|(\mathbf{y}_k^i - \mathbf{x}_k)\right\|^{\nu-1},
\end{aligned} \tag{50}$$

where $\mathbf{u}_k^j = \mathbf{x}_k + \tilde{t}_k^j(\mathbf{y}_k^i - \mathbf{x}_k)$, with $\tilde{t}_k^j \in (0,1)$ Now, recalling that h_j, $j = 1,\dots,p$ are continuously differentiable functions and $\mathbf{y}_k^i = \mathbf{x}_k + t_k^i\hat{\alpha}_k^i\mathbf{d}_k^i$, with $t_k^i \in (0,1)$ and $\|\mathbf{d}_k^i\| = 1$, from (50) and the fact that $\hat{\alpha}_k^i \le \alpha_k$ we can write

$$\begin{aligned}
\left|\nu\left|\frac{h_j(\mathbf{x}_k)}{\rho_k}\right|^{\nu-1} - \nu\left|\frac{h_j(\mathbf{y}_k^i)}{\rho_k}\right|^{\nu-1}\right| &\le \frac{\nu}{\rho_k^{\nu-1}}c_4\left(t_k^i\hat{\alpha}_k^i\right)^{\nu-1} \\
&\le c_4\frac{\nu}{\rho_k^{\nu-1}}\left(t_k^i\alpha_k\right)^{\nu-1} \le c_4\nu\left(\frac{\alpha_k}{\rho_k}\right)^{\nu-1} \\
&\le c_4\nu\left(\frac{\rho_k^{\beta-1}}{\theta_\alpha}\right)^{\nu-1} = c_4\nu\theta_\alpha^{1-\nu}\rho_k^{(\beta-1)(\nu-1)}.
\end{aligned}$$

Given that $\beta > 1$, $\nu \in (1,2]$, and recalling Theorem 2.7, we can conclude that (49) holds.

Now, we are able to prove the boundness of the sequences $\{\lambda_\ell(\mathbf{x}_k;\rho_k)\}_{k\in\mathcal{K}_\rho^{\mathbf{x}}}$, $\ell = 1,\dots,m$ and $\{\mu_j(\mathbf{x}_k;\rho_k)\}_{k\in\mathcal{K}_\rho^{\mathbf{x}}}$, $j = 1,\dots,p$.

In fact, we can rewrite (40) as

$$\begin{aligned}
&\left(\nabla f(\mathbf{y}_k^i) + \sum_{\ell=1}^m \nabla g_\ell(\mathbf{y}_k^i)\lambda_\ell(\mathbf{x}_k;\rho_k) + \right. \\
&+ \sum_{\ell=1}^m \nabla g_\ell(\mathbf{y}_k^i)(\lambda_\ell(\mathbf{y}_k^i;\rho_k,) - \lambda_\ell(\mathbf{x}_k;\rho_k)) + \sum_{j=1}^p \nabla h_j(\mathbf{y}_k^i)\mu_j(\mathbf{x}_k;\rho_k) + \\
&\left. + \sum_{j=1}^p \nabla h_j(\mathbf{y}_k^i)(\mu_j(\mathbf{y}_k^i;\rho_k) - \mu_j(\mathbf{x}_k;\rho_k))\right)^\top \mathbf{d}_k^i \ge -\frac{\xi(\hat{\alpha}_k^i)}{\hat{\alpha}_k^i}, \quad \forall i \in \mathcal{J}_k \text{ and } k \in \mathcal{K}_\rho^{\mathbf{x}}.
\end{aligned} \tag{51}$$

Let

$$\{a_k^1, \ldots, a_k^m\} = \{\lambda_1(\mathbf{x}_k; \rho_k), \ldots, \lambda_m(\mathbf{x}_k; \rho_k)\},$$
$$\{a_k^{m+1}, \ldots, a_k^{m+p}\} = \{\mu_1(\mathbf{x}_k; \rho_k), \ldots, \mu_p(\mathbf{x}_k; \rho_k)\}.$$

Assume, by contradiction, that there exists at least one index $l \in \{1, \ldots, m+p\}$ such that

$$\lim_{\substack{k \to +\infty \\ k \in \mathcal{K}_\rho^{\mathtt{x}}}} |a_k^l| = +\infty. \tag{52}$$

Hence, the sequence $\{a_k^i\}$, $i = 1, \ldots, m+p$, cannot be all convergent to zero. Then, from Lemma B.1, there exists an infinite subset $\mathcal{K}_\rho^{\mathtt{x},\mathtt{a}} \subseteq \mathcal{K}_\rho^{\mathtt{x}}$ and an index $s \in \{1, \ldots, m+p\}$ such that,

$$\lim_{\substack{k \to +\infty \\ k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{a}}}} \frac{a_k^i}{|a_k^s|} = z_i, \quad |z_i| < +\infty, \quad i = 1, \ldots, m+p \tag{53}$$

If there is an unique index l that satisfies (52), then $s = l$. If we have more than one index satisfying the equation, then s is selected as one of the indexes such that $\{a_k^s\}_{k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{a}}}$ tends to $+\infty$ faster than the others. Note also that

$$z_s = 1, \qquad \text{and} \qquad |a_k^s| \to +\infty. \tag{54}$$

Dividing the relation (51) by $|a_k^s|$, we have

$$\begin{aligned}
&\left(\frac{\nabla f(\mathbf{y}_k^i)}{|a_k^s|} + \sum_{\ell=1}^m \frac{\nabla g_\ell(\mathbf{y}_k^i) a_k^\ell}{|a_k^s|} \right. \\
&+ \sum_{\ell=1}^m \nabla g_\ell(\mathbf{y}_k^i) \frac{\lambda_\ell(\mathbf{y}_k^i; \rho_k) - \lambda_\ell(\mathbf{x}_k; \rho_k)}{|a_k^s|} + \sum_{j=1}^p \frac{\nabla h_j(\mathbf{y}_k^i) a_k^{m+j}}{|a_k^s|} \\
&\left. + \sum_{j=1}^p \nabla h_j(\mathbf{y}_k^i) \frac{\mu_j(\mathbf{y}_k^i; \rho_k) - \mu_j(\mathbf{x}_k; \rho_k)}{|a_k^s|} \right)^\top \mathbf{d}_k^i \geq -\frac{\xi(\hat{\alpha}_k^i)}{\hat{\alpha}_k^i |a_k^s|}, \quad \forall i \in \mathcal{J}_k \text{ and } k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{a}}.
\end{aligned} \tag{55}$$

Reasoning as for (21), by Assumption 1 we can extract a further subsequence $\mathcal{K}_\rho^{\mathtt{x},\mathtt{a},\mathtt{D}} \subseteq \mathcal{K}_\rho^{\mathtt{x},\mathtt{a}}$ such that for all $k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{a},\mathtt{D}}$ it holds $\mathcal{J}_k = \mathcal{J}^*$, $\mathbf{d}_k^i = \bar{\mathbf{d}}^i$ for all $i \in \mathcal{J}^*$, and $\mathcal{D}_k = \mathcal{D}^* = \{\bar{\mathbf{d}}^i\}_{i \in \mathcal{J}^*}$. Since $\lim_{k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{a},\mathtt{D}}} \mathbf{x}_k = \mathbf{x}^*$, by Assumption 1 and Proposition 2.4, and using $\varepsilon \in (0, \min\{\bar{\varepsilon}, \varepsilon^*\}]$, for sufficiently large $k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{a},\mathtt{D}}$ we have $\mathcal{T}_X(\mathbf{x}^*) = \mathcal{T}_X(\mathbf{x}_k, \varepsilon) = \text{cone}(\mathcal{D}_k \cap \mathcal{T}_X(\mathbf{x}_k, \varepsilon)) = \text{cone}(\mathcal{D}^*)$.

Taking the limit for $k \to +\infty$ and $k \in \mathcal{K}_\rho^{\mathtt{x},\mathtt{a},\mathtt{D}}$, and using (48), (49), and (53), we obtain

$$\left(\sum_{\ell=1}^m z_\ell \nabla g_\ell(\mathbf{x}^*) + \sum_{j=1}^p z_{m+j} \nabla h_j(\mathbf{x}^*) \right)^\top \bar{\mathbf{d}}^i \geq 0, \quad \forall \bar{\mathbf{d}}^i \in \mathcal{D}^*. \tag{56}$$

We recall that $\mathbf{x}^*$ satisfies the MFCQ conditions. Let $\mathbf{d}$ be the direction satisfying condition (b) of Definition 3.1. For every $\mathbf{d} \in \mathcal{T}_{\mathcal{X}}(\mathbf{x}^*)$, there exist nonnegative numbers β_i such that

$$\mathbf{d} = \sum_{\bar{\mathbf{d}}^i \in \mathcal{D}^*} \beta_i \bar{\mathbf{d}}^i. \tag{57}$$

Thus, from (56) and (57), we obtain

$$\begin{aligned} &\left(\sum_{\ell=1}^{m} z_\ell \nabla g_\ell(\mathbf{x}^*) + \sum_{j=1}^{p} z_{m+j} \nabla h_j(\mathbf{x}^*) \right)^\top \mathbf{d} = \\ &= \sum_{\bar{\mathbf{d}}^i \in \mathcal{D}^*} \beta_i \left(\sum_{\ell=1}^{m} z_\ell \nabla g_\ell(\mathbf{x}^*) + \sum_{j=1}^{p} z_{m+j} \nabla h_j(\mathbf{x}^*) \right)^\top \bar{\mathbf{d}}^i \\ &= \sum_{\bar{\mathbf{d}}^i \in \mathcal{D}^*} \beta_i \sum_{\ell=1}^{m} z_\ell \nabla g_\ell(\mathbf{x}^*)^\top \bar{\mathbf{d}}^i + \sum_{\bar{\mathbf{d}}^i \in \mathcal{D}^*} \beta_i \sum_{j=1}^{p} z_{m+j} \nabla h_j(\mathbf{x}^*)^\top \bar{\mathbf{d}}^i \geq 0. \end{aligned} \tag{58}$$

Considering Definition 3.1, from (58) it follows

$$\sum_{\ell=1}^{m} z_\ell \nabla g_\ell(\mathbf{x}^*)^\top \mathbf{d} \geq 0. \tag{59}$$

Theorem 2.7 and the definition of z_ℓ for $\ell \in \{1, \dots, m\}$, guarantee

$$z_\ell = 0, \qquad \text{for all} \qquad \ell \notin \mathcal{I}_+(\mathbf{x}^*). \tag{60}$$

Since $\mathbf{x}^*$ satisfies the MFCQ conditions, (59) implies

$$z_\ell = 0, \qquad \text{for all } \ell \in \mathcal{I}_+(\mathbf{x}^*). \tag{61}$$

Applying (60) and (61) to (56) we get

$$\left(\sum_{j=1}^{p} z_{m+j} \nabla h_j(\mathbf{x}^*) \right)^\top \mathbf{d}^* \geq 0, \qquad \text{for all} \qquad \mathbf{d}^* \in D^*, \tag{62}$$

using again Definition 3.1 and (62), we obtain

$$z_{m+j} = 0, \qquad \text{for all} \qquad j \in \{1, \dots, p\}. \tag{63}$$

In conclusion, we get (60), (61), and (63), contradicting (54) and this concludes the proof. □

C Appendix

In the following, we show that the results of Theorem 2.7 still hold if we have two parameters $\rho_k^{\mathtt{ext}}$ and $\rho_k^{\mathtt{log}}$ as considered in the implementation details of Section 4. To this aim, let us define

$$\begin{aligned}\mathcal{K}^{\mathtt{log}} &= \{k \in \mathcal{K}_u \mid \text{ (29) is satisfied}\}\\ \mathcal{K}^{\mathtt{ext}} &= \{k \in \mathcal{K}_u \mid \text{ (30) is satisfied}\}.\end{aligned}$$

Recall that when $k \in \mathcal{K}^{\mathtt{log}}$, then $\rho_{k+1}^{\mathtt{log}} < \rho_k^{\mathtt{log}}$, and when $k \in \mathcal{K}^{\mathtt{ext}}$, then $\rho_{k+1}^{\mathtt{ext}} < \rho_k^{\mathtt{ext}}$.

Lemma C.1. *Let Assumption 2 hold. Let $\{\rho_k^{\mathtt{log}}\}_{k\in\mathbb{N}}$, $\{\rho_k^{\mathtt{ext}}\}_{k\in\mathbb{N}}$, and $\{\alpha_k\}_{k\in\mathbb{N}}$ be the sequences of penalty parameters and stepsizes generated by LOG-DS. Then, $\mathcal{K}^{\mathtt{log}}$ and $\mathcal{K}^{\mathtt{ext}}$ are both infinite and it results $\mathcal{K}^{\mathtt{ext}} \subseteq \mathcal{K}^{\mathtt{log}}$. Furthermore,*

$$\lim_{k\to+\infty} \rho_k^{\mathtt{log}} = 0, \tag{64}$$

$$\lim_{k\to+\infty} \rho_k^{\mathtt{ext}} = 0, \tag{65}$$

$$\lim_{k\to+\infty, k\in\mathcal{K}^{\mathtt{ext}}} \alpha_k = 0. \tag{66}$$

Proof. By definition, of $\mathcal{K}^{\mathtt{ext}}$ and $\mathcal{K}^{\mathtt{log}}$, since

$$\min\{(\rho_k^{\mathtt{log}})^\beta, (\rho_k^{\mathtt{ext}})^\beta, (g_{\min})_k^2\} \le \min\{(\rho_k^{\mathtt{log}})^\beta, (g_{\min})_k^2\},$$

we have that if $k \in \mathcal{K}^{\mathtt{ext}}$ then $k \in \mathcal{K}^{\mathtt{log}}$, that is $\mathcal{K}^{\mathtt{ext}} \subseteq \mathcal{K}^{\mathtt{log}}$. Hence, if $\mathcal{K}^{\mathtt{ext}}$ is infinite, then $\mathcal{K}^{\mathtt{log}}$ is infinite too. By using the same arguments used in Theorem 2.7, we can show that $\mathcal{K}^{\mathtt{ext}}$ and $\mathcal{K}^{\mathtt{log}}$ cannot be both finite. Then, it remains to consider the case where $\mathcal{K}^{\mathtt{ext}}$ is finite and $\mathcal{K}^{\mathtt{log}}$ is infinite. In this case, we have $\rho_k^{\mathtt{log}} \to 0$ and $\rho_k^{\mathtt{ext}} = \bar{\rho}$ for k sufficiently large. Since, $0 < \min\{(\rho_k^{\mathtt{log}})^\beta, (g_{\min})_k^2\} \le (\rho_k^{\mathtt{log}})^\beta$, we also have that

$$\min\{(\rho_k^{\mathtt{log}})^\beta, (g_{\min})_k^2\} \to 0.$$

But then, for k sufficiently large

$$\min\{\bar{\rho}^\beta, (\rho_k^{\mathtt{log}})^\beta, (g_{\min})_k^2\} = \min\{(\rho_k^{\mathtt{log}})^\beta, (g_{\min})_k^2\}$$

so that also $\rho_k^{\mathtt{ext}}$ would get updated thus contradicting that $\rho_k^{\mathtt{ext}} = \bar{\rho}$ for k sufficiently large.

Hence, concerning the two sets $\mathcal{K}^{\mathtt{log}}$ and $\mathcal{K}^{\mathtt{ext}}$ it can only happen that they are both infinite. Thus we have that $\rho_k^{\mathtt{log}} \to 0$ and $\rho_k^{\mathtt{ext}} \to 0$.

Finally, we can use similar arguments of Theorem 2.7 to prove (66). □

Data availability statement

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflict of interests statement

The authors have no competing interests to declare that are relevant to the content of this article.

References

[1] M. A. Abramson, O. A. Brezhneva, J. E. Dennis Jr., and R. L. Pingel. "Pattern search in the presence of degenerate linear constraints". In: *Optim. Methods Softw.* 23 (2008), pp. 297–319.

[2] C. Audet, S. Le Digabel, and M. Peyrega. "Linear equalities in blackbox optimization". In: *Comput. Optim. Appl.* 61 (2015), pp. 1–23.

[3] C. Audet and W. Hare. *Derivative-Free and Blackbox Optimization.* Cham, Switzerland: Springer, 2017.

[4] C. Audet and J. E. Dennis Jr. "A pattern search filter method for nonlinear programming without derivatives". In: *SIAM J. Optim.* 14 (2004), pp. 980–1010.

[5] C. Audet and J. E. Dennis Jr. "A progressive barrier for derivative-free nonlinear programming". In: *SIAM J. Optim.* 20 (2009), pp. 445–472.

[6] C. Audet and J. E. Dennis Jr. "Analysis of generalized pattern searches". In: *SIAM J. Optim.* 13 (2003), pp. 889–903.

[7] C. Audet and J. E. Dennis Jr. "Mesh adaptive direct search algorithms for constrained optimization". In: *SIAM J. Optim.* 17 (2006), pp. 188–217.

[8] D. P. Bertsekas. *Nonlinear Programming.* Belmont, Massachusetts: Athena Scientific, 1999.

[9] D. M. Bortz and C. T. Kelley. "The simplex gradient and noisy optimization problems". In: *Computational Methods in Optimal Design and Control, Progress in Systems and Control Theory.* Ed. by J. T. Borggaard, J. Burns, E. Cliff, and S. Schreck. Vol. 24. Birkhäuser, Boston, 1998, pp. 77–90.

[10] A. Brilli, G. Liuzzi, and S. Lucidi. *An interior point method for nonlinear constrained derivative-free optimization.* 2025. DOI: https://doi.org/10.1080/10556788.2025.2453110.

[11] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization.* MPS-SIAM Series on Optimization. Philadelphia: Society for Industrial and Applied Mathematics, 2009.

[12] I. D. Coope and C. J. Price. "Positive bases in numerical optimization". In: *Comput. Optim. Appl.* 21 (2002), pp. 169–175.

[13] A. L. Custódio, H. Rocha, and L. N. Vicente. "Incorporating minimum Frobenius norm models in direct search". In: *Comput. Optim. Appl.* 46 (2010), pp. 265–278.

[14] A. L. Custódio and L. N. Vicente. "Using sampling and simplex derivatives in pattern search methods". In: *SIAM J. Optim.* 18 (2007), pp. 537–555.

[15] R. De Leone, M. Gaudioso, and L. Grippo. "Stopping criteria for linesearch methods without derivatives". In: *Math. Program.* 30 (1984), pp. 285–300.

[16] E. D. Dolan and J. J. Moré. "Benchmarking optimization software with performance profiles". In: *Math. Program.* 91 (2002), pp. 201–213.

[17] R. Fletcher and S. Leyffer. "Nonlinear programming without a penalty function". In: *Math. Program.* 91 (2002), pp. 239–269.

[18] N. I. M. Gould, D. Orban, and Ph. L. Toint. "CUTEst: A constrained and unconstrained testing environment with safe threads for mathematical optimization". In: *Comput. Optim. Appl.* 60 (2015), pp. 545–557.

[19] S. Gratton, C. Royer, L. N. Vicente, and Z. Zhang. "Direct search based on probabilistic feasible descent for bound and linearly constrained problems". In: *Comput. Optim. Appl.* 72 (2019), pp. 525–559.

[20] W. Hare and G. Jarry-Bolduc. "A deterministic algorithm to compute the cosine measure of a finite positive spanning set". In: *Optim. Lett.* 14 (2020), pp. 1305–1316.

[21] W. Hare, G. Jarry-Bolduc, and C. Planiden. "Nicely structured positive bases with maximal cosine measure". In: *Optim. Lett.* 17 (2023), pp. 1495–1515.

[22] T. G. Kolda, R. M. Lewis, and V. Torczon. "Optimization by direct search: new perspectives on some classical and modern methods". In: *SIAM Rev.* 45 (2003), pp. 385–482.

[23] T. G. Kolda, R. M. Lewis, and V. Torczon. "Stationarity results for generating set search for linearly constrained optimization". In: *SIAM J. Optim.* 17 (2006), pp. 943–968.

[24] 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 (2002), pp. 1075–1089.

[25] R. M. Lewis and V. Torczon. "Pattern search algorithms for bound constrained minimization". In: *SIAM J. Optim.* 9 (1999), pp. 1082–1099.

[26] R. M. Lewis and V. Torczon. "Pattern search methods for linearly constrained minimization". In: *SIAM J. Optim.* 10 (2000), pp. 917–941.

[27] G. Liuzzi and S. Lucidi. "A derivative-free algorithm for inequality constrained nonlinear programming via smoothing of an ℓ_∞ penalty function". In: *SIAM J. Optim.* 20 (2009), pp. 1–29.

[28] G. Liuzzi, S. Lucidi, and M. Sciandrone. "A derivative-free algorithm for linearly constrained finite minimax problems". In: *SIAM J. Optim.* 16 (2006), pp. 1054–1075.

[29] G. Liuzzi, S. Lucidi, and M. Sciandrone. "Sequential penalty derivative-free methods for nonlinear constrained optimization". In: *SIAM J. Optim.* 20 (2010), pp. 2614–2635.

[30] S. Lucidi, M. Sciandrone, and P. Tseng. "Objective-derivative-free methods for constrained optimization". In: *Math. Program.* 92 (2002), pp. 37–59.

[31] J. H. May. “Linearly Constrained Nonlinear Programming: A Solution Method That Does Not Require Analytic Derivatives”. PhD thesis. Yale University, 1974.

[32] J. J. Moré and S. M. Wild. “Benchmarking derivative-free optimization algorithms”. In: *SIAM J. Optim.* 20 (2009), pp. 172–191.

[33] W. Rudin. *Functional Analysis*. 2nd. International Series in Pure and Applied Mathematics. New York: McGraw-Hill, Inc., 1991. ISBN: 0-07-054236-8.