Nonlinear Derivative-free Constrained Optimization with a Mixed Penalty-Logarithmic Barrier Approach 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-logarithmic barrier approach and generating set search, for addressing nonlinearly constrained derivative-free optimization problems. A merit function is considered, wherein the set of 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, initially proposed in the framework of LOG-DFL [12], is adapted and incorporated into SID-PSM [14, 15] algorithm, a generalized pattern search method, allowing to effectively handle general nonlinear constraints. Under reasonable assumptions regarding the smoothness of the functions, convergence is established, without any convexity assumptions. Using CUTEst test problems, numerical experiments demonstrate the robustness, efficiency, and overall effectiveness of the proposed method, when compared with state-of-art solvers and with the original SID-PSM and LOG-DFL implementations.

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 the derivative-free optimization problem with general constraints (linear and nonlinear), defined by:

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & g(\mathbf{x}) \leq 0 \\ & h(\mathbf{x}) = 0 \\ & \mathbf{x} \in X \end{array} \tag{1}$$

where $f : X \subseteq \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, $g : X \subseteq \mathbb{R}^n \to \{\mathbb{R} \cup \{+\infty\}\}^m$, $h : X \subseteq \mathbb{R}^n \to \{\mathbb{R} \cup \{+\infty\}\}^p$, and $X = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$, with $\mathbf{A} \in \mathbb{R}^{q \times n}$ and $\mathbf{b} \in \mathbb{R}^q$. Furthermore, we assume that the functions involved (f, g, and h) are continuously differentiable, but derivatives 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, 13].

Considering an initial point $\mathbf{x}_0 \in X$ we define two sets: $\mathcal{G}^{\mathtt{log}} = \{\ell \mid g_\ell(\mathbf{x}_0) < 0\}$, and $\mathcal{G}^{\mathtt{ext}} = \{\ell \mid g_\ell(\mathbf{x}_0) \geq 0\}$. We denote the sets defined by the inequality constraints as $\Omega_{\mathcal{G}^{\mathtt{log}}}$ and $\Omega_{\mathcal{G}^{\mathtt{ext}}}$, respectively. The first corresponds to the inequality constraints, to be addressed with a logarithmic barrier approach, while the constraints in the second will be handled using a penalization term. Further explanation will be given in Section 2. Additionally, Ω_h corresponds to the set defined by the equality constraints, and X is a polyhedron, defining a set that we assume to be compact. Therefore, the feasible region $\mathcal{F}$, which is assumed to be nonempty, is given by

$$\mathcal{F} = X \cap \Omega_{\mathcal{G}^{\mathtt{log}}} \cap \Omega_{\mathcal{G}^{\mathtt{ext}}} \cap \Omega_h \neq \emptyset,$$

and it is also a compact set.

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 [23] or linearly constrained problems [24], and after for general nonlinear constraints [22]. When the constraints are only linear, inspired by the work of May [29], 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 [21, 24, 28], including specific strategies to address the degenerated case [1]. In the presence of nonlinear constraints, augmented Lagrangian approaches have been proposed [22], 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 [18] 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 [25], 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 [27] 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.* [12] 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.

In this work, the strategy of [12] will be adapted and incorporated into generalized pattern search. Starting from the SID-PSM algorithm [14, 15], an implementation of a generalized pattern search method, where polynomial models are used for both the search and the poll steps to improve the numerical performance of the code, LOG-DS has been developed, a direct search method able to explicitly address nonlinear constraints by a mixed penalty-logarithmic barrier approach.

The manuscript is organized as follows. Section 2 details the proposed algorithmic structure and the related convergence properties are analyzed in Section 3. Details of the numerical implementation are provided in Section 4 and numerical results are reported in Section 5. We summarize our conclusions in Section 6, also discussing some future avenues of research. An appendix completes the paper, including some auxiliary technical results.

1.1 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 $\Omega \subset \mathbb{R}^n$, we use the notation $T_\Omega(\mathbf{x})$ to denote the tangent cone to Ω at $\mathbf{x}$.

2 A Direct Search Algorithm for Nonlinear Constrained Optimization

This section is devoted to the introduction of a new DS algorithm, based on the sequential minimization of an adequate merit function, to solve nonlinearly constrained derivative-free optimization problems. Sequential minimization consecutively solves linearly constrained subproblems, differing from each other in a penalty parameter ρ.

Following the original idea of [12], to solve Problem (1), a merit function is considered:

$$Z(\mathbf{x};\rho) = f(\mathbf{x}) - \rho \sum_{\ell \in \mathcal{G}^{\log}} \log(-g_\ell(\mathbf{x})) + \frac{1}{\rho^{\nu-1}} \left(\sum_{\ell \in \mathcal{G}^{\text{ext}}} (\max\{g_\ell(\mathbf{x}), 0\})^\nu + \sum_{j=1}^{p} |h_j(\mathbf{x})|^\nu \right), \quad (2)$$

where $\rho > 0$ and $\nu \in (1,2]$. Therefore, $Z(\mathbf{x};\rho) = +\infty$, for all $\mathbf{x} \in X$ such that $g_\ell(\mathbf{x}) \not< 0$, for $\ell \in \mathcal{G}^{\log}$.

For the linear constraints, defining the set X, strategies that conform the search directions to the geometry of the nearby feasible region are implemented, using the construction procedure detailed in [1, 21, 24].

The following problem

$$\begin{aligned} &\min \ Z(\mathbf{x};\rho) \\ &\text{s.t. } \mathbf{x} \in X \cap \mathring{\Omega}_{\mathcal{G}^{\log}} \end{aligned} \quad (3)$$

will be considered at each iteration, with an adequate strategy for updating the parameter $\rho > 0$. We assume that a point $\mathbf{x}_0 \in X \cap \mathring{\Omega}_{\mathcal{G}^{\log}}$ is known, so that the set $X \cap \mathring{\Omega}_{\mathcal{G}^{\log}}$ is nonempty.

Lemma 2.1. *Let $\rho > 0$, $\nu \in (1,2]$, and $\alpha \in \mathbb{R}$ be given parameters. If $X \cap \Omega_{\mathcal{G}^{\log}}$ is compact, then,*

$$L(\alpha) = \{\mathbf{x} \in X \cap \mathring{\Omega}_{\mathcal{G}^{\log}} : \ Z(\mathbf{x};\rho) \le \alpha\}$$

is compact.

Proof. The set $L(\alpha)$ is bounded since, by definition, it is a subset of $X \cap \Omega_{\mathcal{G}^{\log}}$ which is compact. It remains to prove that $L(\alpha)$ is closed. To this end, we will show that for any sequence $\{\mathbf{x}_k\} \subset L(\alpha)$ such that $\lim\limits_{k\to+\infty} \mathbf{x}_k = \bar{\mathbf{x}}$, it results $\bar{\mathbf{x}} \in L(\alpha)$.

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

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

Taking the limit for $k \to +\infty$ in the above relation we get

$$\lim_{k\to+\infty} Z(\mathbf{x}_k;\rho) \le \alpha. \quad (4)$$

Then, $\bar{\mathbf{x}} \notin \partial\mathring{\Omega}_{\mathcal{G}^{\log}}$, otherwise we would get $\lim_{k\to+\infty} Z(\mathbf{x}_k;\rho) = +\infty > \alpha$. Thus, $\bar{\mathbf{x}} \in X \cap \mathring{\Omega}_{\mathcal{G}^{\log}}$. We know that $Z(\mathbf{x};\rho)$ is continuous on $X \cap \mathring{\Omega}_{\mathcal{G}^{\log}}$. Thus, by (4), we have

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

which means that $\bar{\mathbf{x}} \in L(\alpha)$. That concludes the proof. ☐

One might notice that nonlinear constraints still appear, due to the presence of the inequalities in $\mathcal{G}^{\log}$ which still need to be satisfied. Nevertheless, considering the properties of the merit function and the structure of the proposed scheme, it will be clear that any point $\mathbf{x} \notin \mathring{\Omega}_{\mathcal{G}^{\log}}$ will be rejected. That is, Problem 3 can be reformulated using only linear constraints, i.e.,

$$\begin{aligned} &\min \ Z(\mathbf{x};\rho) \\ &\text{s.t. } \mathbf{x} \in X. \end{aligned} \quad (5)$$

In the proposed approach, the acceptance of new points will rely on the notion of sufficient decrease, justifying the classification of the algorithm as a Generating Set Search (GSS) method. The next definition adjusts the concept of forcing function (see [20]) and is used to define the sufficient decrease condition required to accept new points in LOG-DS.

Definition 2.2. *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* $t \downarrow 0$;
- *if* $\xi(t) \to 0$ *then* $t \to 0$.

The proposed algorithmic structure is detailed below.

LOG-DS

Data. $\mathbf{x}_0 \in X$ such that $g_\ell(\mathbf{x}_0) < 0$ for all $\ell \in \mathcal{G}^{\log}$, $\mathcal{D}$ a family of sets of directions, $\alpha_0 > 0$, $\rho_0 > 0$, $\nu \in (1, 2]$, $\theta_\alpha, \theta_\rho \in (0, 1)$, $\phi \geq 1$, and $\beta > 1$.

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

Step 1. (Search Step, optional)
If $\mathbf{z}_k \in X$ can be computed such that $Z(\mathbf{z}_k; \rho_k) \leq Z(\mathbf{x}_k; \rho_k) - \xi(\alpha_k)$,
Then set $\mathbf{x}_{k+1} = \mathbf{z}_k$, $\alpha_{k+1} = \phi\alpha_k$, and go to **Step 3.**

Step 2. (Poll Step)

Select a set $\mathbb{D}_k \in \mathcal{D}$
If $\exists\ \mathbf{d}_k^i \in \mathbb{D}_k$: $\mathbf{x}_k + \alpha_k \mathbf{d}_k^i \in X$ and $Z(\mathbf{x}_k + \alpha_k \mathbf{d}_k^i; \rho_k) \leq Z(\mathbf{x}_k; \rho_k) - \xi(\alpha_k)$,
Then set $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k^i$ and $\alpha_{k+1} = \phi\alpha_k$.

Else set $\mathbf{x}_{k+1} = \mathbf{x}_k$ and $\alpha_{k+1} = \theta_\alpha \alpha_k$.

Step 3. Set $(g_{\min})_k = \min_{\ell \in \mathcal{G}^{\log}} \{|g_\ell(\mathbf{x}_{k+1})|\}$

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

Endfor

Following the general structure proposed by Audet and Dennis [6] for generalized pattern search, each iteration of LOG-DS is organized into two main steps, plus an additional one related to the novel approach:

- Step 1 and Step 2, namely the Search Step and the Poll Step, which are part of the basic structure of a generalized pattern search method, here applied to the solution of (5);

- Step 3, the penalty parameter updating step, which is the original feature of the proposed approach.

As usual, the search step 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 will be detailed in Section 4, the original SID-PSM algorithm uses quadratic interpolation 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 will be omitted.

For the theoretical analysis, we will consider general assumptions on the directions used by the algorithm, formalized below.

Definition 2.3 (Active constraints and tangent related sets)**.** *For every* $\mathbf{x} \in X$*, i.e., such that* $\mathbf{A}\mathbf{x} \leq \mathbf{b}$*:*

$$
\begin{aligned}
I_X(\mathbf{x}) &= \{i \mid \mathbf{a}_i^\top \mathbf{x} = b_i\} \quad \textit{(set of indices of active constraints)} \\
T_X(\mathbf{x}) &= \{\mathbf{d} \in \mathbb{R}^n \mid \mathbf{a}_i^\top \mathbf{d} \leq 0,\ i \in I_X(\mathbf{x})\} \quad \textit{(tangent cone at } \mathbf{x}\textit{)} \\
T_X^\circ(\mathbf{x}) &= \left\{ \mathbf{v} \in \mathbb{R}^n \mid \mathbf{v} = \sum_{i \in I_X(\mathbf{x})} \lambda_i \mathbf{a}_i,\ \lambda_i \geq 0 \right\} \quad \textit{(polar of the tangent cone at } \mathbf{x}\textit{)}
\end{aligned}
$$

Given an iterate $\mathbf{x}_k \in X$ (possibly not belonging to the boundary of X), it is important to be able to capture the geometry of the set X near $\mathbf{x}_k$. Hence, the previous sets are approximated by the following ones, depending on a parameter $\varepsilon > 0$:

Definition 2.4 (ε-Active constraints and tangent related sets)**.**

$$
\begin{aligned}
I_X(\mathbf{x}_k, \varepsilon) &= \{i \mid \mathbf{a}_i^\top \mathbf{x}_k \geq b_i - \varepsilon\} \quad \textit{(set of indices of } \varepsilon\textit{-active constraints)} \\
T_X(\mathbf{x}_k, \varepsilon) &= \{\mathbf{d} \in \mathbb{R}^n \mid \mathbf{a}_i^\top \mathbf{d} \leq 0,\ i \in I_X(\mathbf{x}_k, \varepsilon)\} \quad \textit{(}\varepsilon\textit{-tangent cone at } \mathbf{x}_k\textit{)} \\
T_X^\circ(\mathbf{x}_k, \varepsilon) &= \left\{ \mathbf{v} \in \mathbb{R}^n \mid \mathbf{v} = \sum_{i \in I_X(\mathbf{x}_k, \varepsilon)} \lambda_i \mathbf{a}_i,\ \lambda_i \geq 0 \right\} \quad \textit{(polar of the } \varepsilon\textit{-tangent cone at } \mathbf{x}_k\textit{)}
\end{aligned}
$$

A relation between the two groups of sets of Definition 2.3 and Definition 2.4 has been established in [28]. We recall the result in the following proposition.

Proposition 2.5. *Let* $\{\mathbf{x}_k\}_{k \in \mathbb{N}}$ *be a sequence of points in* X*, converging to* $\mathbf{x}^* \in X$*. 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*

$$
\begin{aligned}
I_X(\mathbf{x}^*) &= I_X(\mathbf{x}_k, \varepsilon) \\
T_X(\mathbf{x}^*) &= T_X(\mathbf{x}_k, \varepsilon)
\end{aligned}
$$

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

We are now in position of specifying the requests on the sets of directions, $\mathbb{D}_k$, used by the algorithm (see [26, Assumption 2]).

Assumption 2.6. *Let $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$ be a sequence of points in X. The sequence $\{\mathbb{D}_k\}$ of poll directions satisfies:*

$$\mathbb{D}_k = \{\mathbf{d}_k^i \mid \|\mathbf{d}_k^i\| = 1, i = 1, \ldots, |\mathbb{D}_k|\}$$

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

$$\text{cone}(\mathbb{D}_k \cap T_X(\mathbf{x}_k, \varepsilon)) = T_X(\mathbf{x}_k, \varepsilon), \quad \forall\, \varepsilon \in (0, \bar{\varepsilon}].$$

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

One problem still to be addressed is related to the fact that the merit function is defined and differentiable only in the interior of the feasible region, i.e. $\mathring{\Omega}_{\mathcal{G}^{\log}}$. An analysis based on differentiability assumptions seems not to be applicable. It turns out that the structure of the merit function can be exploited to extract information that allows the infeasible points to be treated as feasible failures, as it will be stated in Proposition 3.6.

Finally, some comments on the penalty parameter updating rule. This approach was originally proposed in [12], in a clear difference with respect to interior point methods when applied to nonlinear optimization problems (even when in presence of first or second-order information). In [17], the convergence of sequential penalization methods with a mixed interior-exterior strategy for nonlinear optimization problems has been established considering the sequence of exact minimizers of the subproblems. In general, finding the exact solution of any subproblem is practically unviable. In [9], inspired by [17], the authors proposed a method using Newton-type directions and the same mixed penalty approach, but allowing a monotonically decreasing sequence of errors on the unconstrained minimizers. More precisely, the algorithm decreases the penalty parameter when the Euclidean norm of the gradient of the merit function is below the penalty parameter itself. In the present work, the updating rule plays a crucial role in the convergence properties of the algorithm, allowing to establish the boundedness of the sequence of associated Lagrange multipliers.

3 Convergence Analysis

In order to prove the global convergence of the method, we will establish that the sequences of stepsizes and penalty parameters converge to zero. Initially, we derive an auxiliary result analyzing the behavior of the algorithm for fixed values of the penalty parameter.

Lemma 3.1. *Let $\{\rho_k\}_{k\in\mathbb{N}}$ and $\{\alpha_k\}_{k\in\mathbb{N}}$ be the sequences of penalty parameters and stepsizes, respectively, generated by algorithm LOG-DS. Assume that*

$$\lim_{k\to+\infty} \rho_k = \bar{\rho} > 0 \tag{6}$$

Then,

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

Proof. From the updating rule of the penalty parameter, i.e. $\rho_{k+1} = \theta_\rho \rho_k$, we have that $\{\rho_k\}$ is a monotone non-increasing sequence. Furthermore, if ρ_k is updated infinitely many times we would have $\bar{\rho} = 0$. Hence, we have that $\rho_{k+1} = \rho_k = \bar{\rho}$ for all k sufficiently large. Let us split the iteration sequence into the following two sets

$$K_s = \{k : \mathbf{x}_{k+1} \neq \mathbf{x}_k\},$$
$$K_u = \{k : \mathbf{x}_{k+1} = \mathbf{x}_k\}.$$

At every iteration k of the algorithm, for k sufficiently large, we have either $Z(\mathbf{x}_{k+1}; \bar{\rho}) = Z(\mathbf{x}_k; \bar{\rho})$ (when $k \in K_u$) or $Z(\mathbf{x}_{k+1}; \bar{\rho}) \leq Z(\mathbf{x}_k; \bar{\rho}) - \xi(\alpha_k)$ (when $k \in K_s$). Hence, the sequence of function values $\{Z(\mathbf{x}_k; \bar{\rho})\}$ is monotonically nonincreasing. By Lemma 2.1, $Z(\mathbf{x}; \bar{\rho})$ has compact level sets, thus it is bounded from below. Hence,

$$\lim_{k \to +\infty} Z(\mathbf{x}_k; \bar{\rho}) = \bar{Z}. \tag{7}$$

If K_s is infinite, from (7) we get

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

Recalling Definition 2.2, we get

$$\lim_{\substack{k \to +\infty \\ k \in K_s}} \alpha_k = 0. \tag{8}$$

If K_u is infinite, for every $k \in K_u$, let us define m_k to be the largest index such that $m_k \in K_s$ and $m_k < k$ (the result is immediate if K_s is empty). Then, we can write

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

When $k \to +\infty$, $k \in K_u$, we have that either $m_k \to +\infty$ as well (when K_s is infinite) or $k - m_k - 1 \to +\infty$ (when K_s is finite). Thus, by (8) and the fact that $\theta_\alpha \in (0, 1)$, we have

$$\lim_{\substack{k \to +\infty \\ k \in K_u}} \alpha_k = \lim_{\substack{k \to +\infty \\ k \in K_u}} \alpha_{m_k} \phi \theta_\alpha^{k - m_k - 1} = 0. \tag{9}$$

The proof is concluded considering (8) and (9). □

Lemma 3.1 is used to show that the sequence of penalty parameters converges to zero, which is required to ensure that in the limit the algorithm will solve the original problem. As a consequence, we are able to establish that the sequence of stepsizes will also converge to zero in the general case. Since it is well-known that the stepsize is related to some measures of stationarity of the problem [20], that property is also relevant.

Theorem 3.2. *Let $\{\rho_k\}_{k \in \mathbb{N}}$ and $\{\alpha_k\}_{k \in \mathbb{N}}$ be the sequences of penalty parameters and stepsizes generated by LOG-DS. Then,*

$$\lim_{k \to +\infty} \rho_k = 0, \tag{10}$$

$$\lim_{k \to +\infty} \alpha_k = 0. \tag{11}$$

Proof. We first prove (10). The algorithmic structure implies that $\{\rho_k\}_{k\in\mathbb{N}}$ is a monotone nonincreasing sequence of positive numbers. Hence, we have that

$$\lim_{k\to+\infty} \rho_k = \bar{\rho} \geq 0.$$

By contradiction, let us assume that $\bar{\rho} > 0$. If this is the case, there must exist an integer $\bar{k}$ such that

$$\rho_{k+1} = \rho_k = \bar{\rho} > 0, \quad \forall k \geq \bar{k}.$$

Again, the instructions of the algorithm imply that, for all $k \geq \bar{k}$,

$$\alpha_{k+1} > \min\{\bar{\rho}^\beta, (g_{\min})_k^2\} \text{ or } \alpha_{k+1} \geq \alpha_k$$

and

$$Z(\mathbf{x}_{k+1}; \bar{\rho}) \leq Z(\mathbf{x}_k; \bar{\rho}) \leq Z(\mathbf{x}_{\bar{k}}; \bar{\rho}) = \bar{Z} < +\infty. \tag{12}$$

Since $\rho_k = \bar{\rho}$, for all $k \geq \bar{k}$, by Lemma 3.1 we have $\lim_{k\to+\infty} \alpha_k = 0$. Then, there must exist an infinite index set K such that $\alpha_{k+1} < \alpha_k$, for all $k \in K$. Hence, for every $k \in K$, $k \geq \bar{k}$, we also have that

$$\alpha_{k+1} > \min\{\bar{\rho}^\beta, (g_{\min})_k^2\} \tag{13}$$

(otherwise, the algorithm would have updated ρ). Taking the limit in relation (13), we obtain

$$\lim_{k\to+\infty} (g_{\min})_k = 0.$$

Thus, we have

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

However, this limit is in contradiction with (12), thus proving (10).

Now, we prove (11). We proceed by contradiction and assume that a subset of iterations $\bar{K}$ exists such that, for all $k \in \bar{K}$,

$$\alpha_k \geq \bar{\alpha} > 0.$$

Since $\rho_k \to 0$, we also have $\lim_{k\to+\infty, k\in\bar{K}} \rho_k = 0$. Then, since $\{\rho_k\}$ is a monotone nonincreasing sequence, a further subset $\tilde{K} \subseteq \bar{K}$ exists such that

$$\lim_{k\to+\infty, k\in\tilde{K}} \rho_k = 0,$$

$$\rho_{k+1} < \rho_k, \quad \text{for all } k \in \tilde{K}.$$

If $k \in \tilde{K}$ is a successful iteration, then $\alpha_{k+1} = \phi\alpha_k \geq \phi\bar{\alpha}$; otherwise, $\alpha_{k+1} = \theta_\alpha \alpha_k \geq \theta_\alpha \bar{\alpha}$. Hence, for $k \in \tilde{K}$, we can write

$$\alpha_{k+1} \geq \theta_\alpha \bar{\alpha}.$$

Thus, for all $k \in \tilde{K}$, we have

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

which contradicts $\rho_k \to 0$, thus concluding the proof. □

Let us define the index set

$$K = \{k \in \mathbb{N} : \rho_{k+1} < \rho_k\}, \tag{14}$$

i.e., the set of iteration indices where the penalty parameter is updated. Note that, by the instructions of the algorithm, $\alpha_{k+1} < \alpha_k$ for all $k \in K$, i.e., every iteration $k \in K$ is an unsuccessful iteration. Recall that, by Theorem 3.2, K is an infinite index set.

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

Definition 3.3. *The point $\mathbf{x} \in X$ is said to satisfy the MFCQ for Problem* (1) *if the two following conditions are satisfied:*

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

$$\left(\sum_{i=1}^{q} \alpha_i \nabla h_i(\mathbf{x})\right)^\top \mathbf{d} \geq 0, \qquad \forall \mathbf{d} \in T_X(\mathbf{x}); \tag{15}$$

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

$$\nabla g_\ell(\mathbf{x})^\top \mathbf{d} < 0, \quad \forall \ell \in I_+(\mathbf{x}), \quad \nabla h_j(\mathbf{x})^\top \mathbf{d} = 0, \quad \forall j = 1, \dots, p \tag{16}$$

where $I_+(\mathbf{x}) = \{\ell \mid g_\ell(\mathbf{x}) \geq 0\}$.

Consider the Lagrangian function $L(\mathbf{x}, \lambda, \mu)$, associated with the nonlinear constraints of Problem (1), 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 [10, Prop. 3.3.8]), which states necessary optimality conditions for Problem (1).

Proposition 3.4. *Let $\mathbf{x}^* \in \mathcal{F}$ be a local minimum of Problem* (1) *that satisfies the MFCQ. Then, there exist vectors $\lambda^* \in \mathbb{R}^m$, $\mu^* \in \mathbb{R}^p$ such that*

$$\nabla_x L(\mathbf{x}^*, \lambda^*, \mu^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0, \qquad \forall \mathbf{x} \in X \tag{17}$$

$$(\lambda^*)^\top g(\mathbf{x}^*) = 0, \quad \lambda^* \geq 0. \tag{18}$$

Therefore, we consider the following definition of stationarity.

Definition 3.5 (Stationary point)**.** *A point $\mathbf{x}^* \in \mathcal{F}$ is said to be a stationary point for Problem* (1) *if vectors $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ exist such that* (17) *and* (18) *are satisfied.*

Before proving the main convergence result, since the merit function is not defined in $\mathbb{R}^n \setminus \mathring{\Omega}_{\mathcal{G}}^{\mathtt{log}}$ and the trial points corresponding to failures might not be feasible, we present the following proposition that allows us to use the mean value theorem in the proof of Theorem 3.7.

Proposition 3.6. *Given any $\bar{\rho} > 0$, $\mathbf{x} \in X \cap \mathring{\Omega}_{\mathcal{G}^{\log}}$, $\mathbf{d} \in \mathbb{R}^n$, and $\bar{\alpha} \in \mathbb{R}_+$ such that $\mathbf{x}+\bar{\alpha}\mathbf{d} \in X$ and $\mathbf{x} + \bar{\alpha}\mathbf{d} \notin \mathring{\Omega}_{\mathcal{G}^{\log}}$, there exists $\hat{\alpha} < \bar{\alpha}$ such that:*

$$\mathbf{x} + \alpha\mathbf{d} \in X \cap \mathring{\Omega}_{\mathcal{G}^{\log}} \quad \textit{for all} \quad \alpha \in (0, \hat{\alpha}],$$
$$Z(\mathbf{x} + \hat{\alpha}\mathbf{d}; \bar{\rho}) > Z(\mathbf{x}; \bar{\rho}) - \xi(\hat{\alpha}).$$

Proof. By definition X is convex, then $\mathbf{x} + \alpha d \in X$ for all $\alpha \in [0, \bar{\alpha}]$.

Since $\mathbf{x} + \bar{\alpha}\mathbf{d} \notin \mathring{\Omega}_{\mathcal{G}^{\log}}$ there exists an index $\ell \in \mathcal{G}^{\log}$ such that

$$\min_{i \in \mathcal{G}^{\log}, g_i(\mathbf{x}+\bar{\alpha}\mathbf{d}) \geq 0} \{g_i(\mathbf{x} + \bar{\alpha}\mathbf{d})\} = g_\ell(\mathbf{x} + \bar{\alpha}\mathbf{d}) \geq 0.$$

The continuity of g and the fact that $x \in \mathring{\Omega}_{\mathcal{G}^{\log}}$ allow us to conclude that there exists at least one scalar $\tilde{\alpha} \in (0, \bar{\alpha}]$ such that $g_\ell(\mathbf{x} + \tilde{\alpha}\mathbf{d}) = 0$. If such a scalar is not unique, let $\tilde{\alpha}$ be the smallest one. This implies that $x + \alpha\mathbf{d} \in \mathring{\Omega}_{\mathcal{G}^{\log}}$ for all $\alpha \in [0, \tilde{\alpha})$. Recall that, by definition of $Z(\mathbf{x}; \rho)$, for all $\mathbf{y} \in \partial\mathring{\Omega}_{\mathcal{G}^{\log}}$, it results

$$\lim_{\substack{\mathbf{x} \to \mathbf{y}, \\ \mathbf{x} \in \mathring{\Omega}_{\mathcal{G}^{\log}}}} Z(\mathbf{x}; \bar{\rho}) = +\infty.$$

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

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

Theorem 3.7. *Let $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$ be the sequence of iterates generated by LOG-DS and recall definition* (14) *of K. Assume that the sets of directions $\{\mathbb{D}_k\}_{k\in\mathbb{N}}$, used by the algorithm, satisfy Assumption 2.6 and define $J_k = \{i \in \{1, 2, \ldots, |\mathbb{D}_k|\} : \mathbf{d}_k^i \in \mathbb{D}_k \cap T_X(\mathbf{x}_k; \varepsilon)\}$, with $\varepsilon \in (0, \min\{\bar{\varepsilon}, \varepsilon^*\}]$ where ε^* and $\bar{\varepsilon}$ are the constants appearing in Proposition 2.5 and Assumption 2.6, respectively. Then, any limit point of $\{\mathbf{x}_k\}_{k\in K}$ that satisfies the MFCQ is a stationary point of Problem* (1)*.*

Proof. First note that, by Theorem 3.2, we have

$$\lim_{\substack{k \to +\infty \\ k \in K}} \rho_k = 0,$$
$$\lim_{\substack{k \to +\infty \\ k \in K}} \alpha_k = 0.$$

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

$$\lim_{\substack{k \to +\infty \\ k \in \hat{K}}} \rho_k = 0,$$
$$\lim_{\substack{k \to +\infty \\ k \in \hat{K}}} \alpha_k = 0,$$
$$\lim_{\substack{k \to +\infty \\ k \in \hat{K}}} \mathbf{x}_k = \mathbf{x}^*,$$

with $\alpha_{k+1} < \alpha_k$, for all $k \in \hat{K}$. Recall that $\mathbb{D}_k = \{\mathbf{d}_k^1, \mathbf{d}_k^2, \dots, \mathbf{d}_k^{r_k}\}$. Then, for all $k \in \hat{K}$ sufficiently large, we know that $\mathbf{x}_k + \alpha_k \mathbf{d}_k^i \in X$ for all $i \in J_k$. For every $i \in J_k$, if $\mathbf{x}_k + \alpha_k \mathbf{d}_k^i \in \mathring{\Omega}_{\mathcal{G}^{\texttt{log}}}$, by the instructions of the algorithm we have

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

Otherwise, i.e. when $\mathbf{x}_k + \alpha_k \mathbf{d}_k^i \notin \mathring{\Omega}_{\mathcal{G}^{\texttt{log}}}$, Proposition 3.6 allows us to ensure the existence of a scalar $\hat{\alpha}_k^i \le \alpha_k$ such that

$$Z(\mathbf{x}_k + \hat{\alpha}_k^i \mathbf{d}_k^i; \rho_k) > Z(\mathbf{x}_k; \rho_k) - \xi(\hat{\alpha}_k^i). \tag{19}$$

Applying the mean value theorem to (19), we can write

$$-\xi(\hat{\alpha}_k^i) \le Z(\mathbf{x}_k + \hat{\alpha}_k^i \mathbf{d}_k^i; \rho_k) - Z(\mathbf{x}_k; \rho_k) = \hat{\alpha}_k^i \nabla Z(\mathbf{y}_k^i; \rho_k)^\top \mathbf{d}_k^i, \tag{20}$$

for all $k \in \hat{K}$ sufficiently large and all $i \in J_k$, 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

$$\nabla Z(\mathbf{y}_k^i; \rho_k)^\top \mathbf{d}_k^i \ge -\frac{\xi(\hat{\alpha}_k^i)}{\hat{\alpha}_k^i}, \quad \forall i \in J_k. \tag{21}$$

By considering the expression of $Z(\mathbf{x}; \rho_k)$, we can write

$$\nabla Z(\mathbf{y}_k^i; \rho_k)^\top \mathbf{d}_k^i = \left[\nabla f(\mathbf{y}_k^i) + \sum_{\ell \in \mathcal{G}^{\texttt{log}}} \frac{\rho_k}{-g_\ell(\mathbf{y}_k^i)} \nabla g_\ell(\mathbf{y}_k^i) + \nu \left(\sum_{\ell \in \mathcal{G}^{\texttt{ext}}} \left(\frac{\max\{g_\ell(\mathbf{y}_k^i), 0\}}{\rho_k} \right)^{\nu-1} \nabla g_\ell(\mathbf{y}_k^i) + \right.\right.$$
$$\left.\left. \sum_{j=1}^p \left(\frac{|h_j(\mathbf{y}_k^i)|}{\rho_k} \right)^{\nu-1} \nabla h_j(\mathbf{y}_k^i) \right)\right]^\top \mathbf{d}_k^i \ge -\frac{\xi(\hat{\alpha}_k^i)}{\hat{\alpha}_k^i}, \quad \forall i \in J_k \text{ and } k \in \hat{K} \text{ sufficiently large.} \tag{22}$$

By Assumption 2.6, we can extract a further subset of indices $\tilde{K} \subseteq \hat{K}$ such that, $\alpha_{k+1} < \alpha_k$, for all $k \in \tilde{K}$ and

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

and $D^* = \{\bar{\mathbf{d}}^i\}_{i \in J}$. When $k \in \tilde{K}$ is sufficiently large, for all $i \in J$, 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 3.2, we have that, $\lim_{\substack{k \to +\infty \\ k \in \tilde{K}}} \mathbf{y}_k^i = \mathbf{x}^*$.

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

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

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

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

$$\lim_{\substack{k \to +\infty \\ k \in K'}} \lambda_\ell(\mathbf{x}_k; \rho_k) = \lambda_\ell^*, \qquad \ell = 1, \dots, m \tag{23}$$

$$\lim_{\substack{k \to +\infty \\ k \in K'}} \mu_j(\mathbf{x}_k; \rho_k) = \mu_j^*, \qquad j = 1, \dots, p \tag{24}$$

and define $\lambda_\ell^* = 0$ for $\ell \notin I_+(\mathbf{x}^*)$.

Multiplying (22) by $\rho_k^{\nu-1}$ and taking the limit for $k \to +\infty, k \in K'$, 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 2.5 and Assumption 2.6, we know that there is $\varepsilon > 0$ such that for all $k \in K'$ sufficiently large

$$T_X(\mathbf{x}^*) = T_X(\mathbf{x}_k; \varepsilon) = \text{cone}(\mathbb{D}_k \cap T_X(\mathbf{x}_k; \varepsilon)) = \text{cone}(D^*).$$

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

$$\mathbf{d} = \sum_{i \in J} \beta_i \bar{\mathbf{d}}^i, \text{with } \bar{\mathbf{d}}^i \in D^*. \tag{25}$$

Let us recall that, by assumption, $\mathbf{x}^*$ satisfies MFCQ conditions, and let $\mathbf{d}$ be the direction satisfying (16) in point (b). Then we have,

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

Again by (16), $\nabla g_\ell(\mathbf{x}^*)^T \mathbf{d} < 0$, for all $\ell \in I_+(\mathbf{x}^*)$, 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 $\ell \in \ell \in I_+(\mathbf{x}^*) \cap \mathcal{G}^{\mathtt{ext}}$, so that $g_\ell(\mathbf{x}^*) \leq 0$ for all $\ell \in \mathcal{G}^{\mathtt{ext}}$.

Furthermore, that implies

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

$$\left(\sum_{j=1}^{p} \nu |h_j(\mathbf{x}^*)|^{\nu-1} \nabla h_j(\mathbf{x}^*) \right)^\top \bar{\mathbf{d}} \geq 0, \text{for all } \bar{\mathbf{d}} \in T_X(\mathbf{x}^*).$$

Using (15), we get $h_j(\mathbf{x}^*) = 0$ for all $j = 1, \ldots, p$. Therefore, the point $\mathbf{x}^*$ is feasible.

By simple manipulations, inequality (22) can be rewritten as

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

Taking limits for $k \to +\infty$, $k \in K'$ and considering (48) and (49), we get:

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

Again, by (25) and (27), we have, for all $\mathbf{d} \in T_X(\mathbf{x}^*)$,

$$\left(\nabla f(\mathbf{x}^*) + \sum_{\ell=1}^{m} \nabla g_\ell(\mathbf{x}^*) \lambda_\ell^* + \sum_{j=1}^{p} \nabla h_j(\mathbf{x}^*) \mu_j^* \right)^\top \mathbf{d} =$$

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

Since $\mathbf{x}^*$ is feasible and considering the definition of λ_ℓ^* for $\ell \in I_+(\mathbf{x}^*)$, $\lambda_\ell^* \, g_\ell(\mathbf{x}^*) = 0$, for all $\ell = 1, \ldots, m$, and the proof is concluded. □

4 Implementation Details

In this section, we describe a practical implementation for LOG-DS, based on the original implementation of SID-PSM [14, 15].

4.1 LOG-DS vs SID-PSM

LOG-DS enhances SID-PSM with the capability of handling general constraints through a mixed penalty log-barrier approach. Thus, the original structure and algorithmic options of SID-PSM implementation are kept. In this subsection, we will provide a general overview of the main features of SID-PMS, highlighting the differences with LOG-DS. For more details on SID-PSM, the original references [14, 15] 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 infeasibility regarding the nonlinear constraints. The merit function also replaces the original objective function through the different algorithmic steps. So, at the Search step, quadratic polynomial models are built for the objective and constraints functions and, after being aggregated into the merit function, are minimized inside a ball with radius directly related to the stepsize 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 [14] to compute the model coefficients.

After the model minimization, LOG-DS needs to make a decision on accepting or rejecting the new trial point. Differently from SID-PSM, where only simple decrease is required for accepting new points, in LOG-DS points are accepted if they satisfy the sufficient decrease condition

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

where $\gamma = 10^{-9}$. The use of a sufficient decrease condition for the acceptance of new points changes the type of globalization strategy used by the algorithm, which is no longer classified as a Generalized Pattern Search method, being now a Generating Set Search (GSS) method.

The algorithm proceeds with an opportunistic Poll Step, accepting the first poll point that satisfies the sufficient decrease condition. Before initiating the polling procedure, previously evaluated points are again used to build a simplex gradient [11], which will be used as an ascent indicator. Poll directions are reordered according to the largest angle made with this ascent indicator, before initiating polling (see [15]). In LOG-DS, the simplex gradient is built for the merit function, while in SID-PSM the original objective function is considered.

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 $\mathbf{x}_0 \in X$:

$$\mathcal{G}^{\texttt{log}} = \{i \mid g_\ell(\mathbf{x}_0) < 0\} \text{ and } \mathcal{G}^{\texttt{ext}} = \{i \mid g_\ell(\mathbf{x}_0) \geq 0\}.$$

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 ρ^{ext} as $\rho^{\nu-1}$ in equation (2). Therefore, we can write the merit function as

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

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

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

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

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

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

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

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

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

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 very capable of attaining feasibility. While scaling the initial exterior penalty parameter with respect to the initial value of the objective function improved the 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 many 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 [12], part of the CUTEst collection [19]. 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 $\mathbf{x}_0 \in \mathring{\Omega}_{\mathcal{G}^{\texttt{log}}}$.

Table 1 details the test set, providing the name, the number of variables, n_p, the number of inequality constraints, $m_p = |\mathcal{G}^{\texttt{log}} \cup \mathcal{G}^{\texttt{ext}}|$, the number of inequality constraints treated by the logarithmic barrier $\bar{m}_p = |\mathcal{G}^{\texttt{log}}|$, and the number of nonlinear equalities, meq_p, for each problem. Additional information can be found in [12, 19].

5.1 Performance and data profiles

Numerical experiments will be analyzed using performance [16] and data [30] 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 $c(x) > 10^{-4}$, where

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

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:

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

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 tolerance τ) with κ estimates of simplex gradients and is defined by:

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

where n_p represents the dimension of problem p.

Problem	n_p	m_p	$\bar{m}_p$	meq_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_p	m_p	$\bar{m}_p$	meq_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_p, m_p, $\bar{m}_p$, and meq_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 comparison

This subsection aims to demonstrate the good numerical performance of the LOG-DS algorithm.

5.2.1 Comparison between strategies for 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 addresses each linear inequality constraint as a general nonlinear inequalitiy, i.e. through the penalty approach discussed previously. The second approach entails addressing the linear inequality constraints via an extreme barrier method, adjusting the directions in accordance with the geometry of the feasible region.

The works by Lucidi et al. [28] and Lewis and Torczon [24] 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 the set of desired directions, regardless of whether the constraints are degenerate or not.

In order to use Definition 2.4, of ε-active constraints, we assume a preliminary scaling of the constraints. For this purpose, we multiply each ith constraint by $\|\mathbf{a}_i\|^{-1}$, since the vectors $\mathbf{a}_i$, $i \in I_X$, are not null (we have that $\|\mathbf{a}_i\| \neq 0$ for all $i \in I_X$). Therefore, we consider

$$\overline{\mathbf{a}}_i = \frac{\mathbf{a}_i}{\|\mathbf{a}_i\|}, \quad \overline{b}_i = \frac{b_i}{\|\mathbf{a}_i\|}, i \in I_X. \tag{34}$$

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

To compute the set of directions $\mathbb{D}_k$ that conform 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 ε-active non-redundant constraints, and the second the set of directions $\mathbb{D}_k$, which include the generators of the cone $T_X(\mathbf{x}_k, \varepsilon)$.

It is important to understand why we are comparing the two strategies. First, linear constraints might not be explicitly given for a black-box type problem, making it impossible to conform the directions to the linear constraints. In such cases, we would be forced to treat the linear inequalities with a mixed penalty. Furthermore, the logarithmic barrier approach is well-known for handling linear constraints very efficiently, especially in the presence of a large number of constraints. Finally, 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.

For identifying a constraint as being ε-active, we considered $\varepsilon = 10^{-5}$ and we used $\mathbb{D}_k = [\mathbf{1} \; -\mathbf{1} \quad \mathbb{I}_n \; -\mathbb{I}_n]$ 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}.

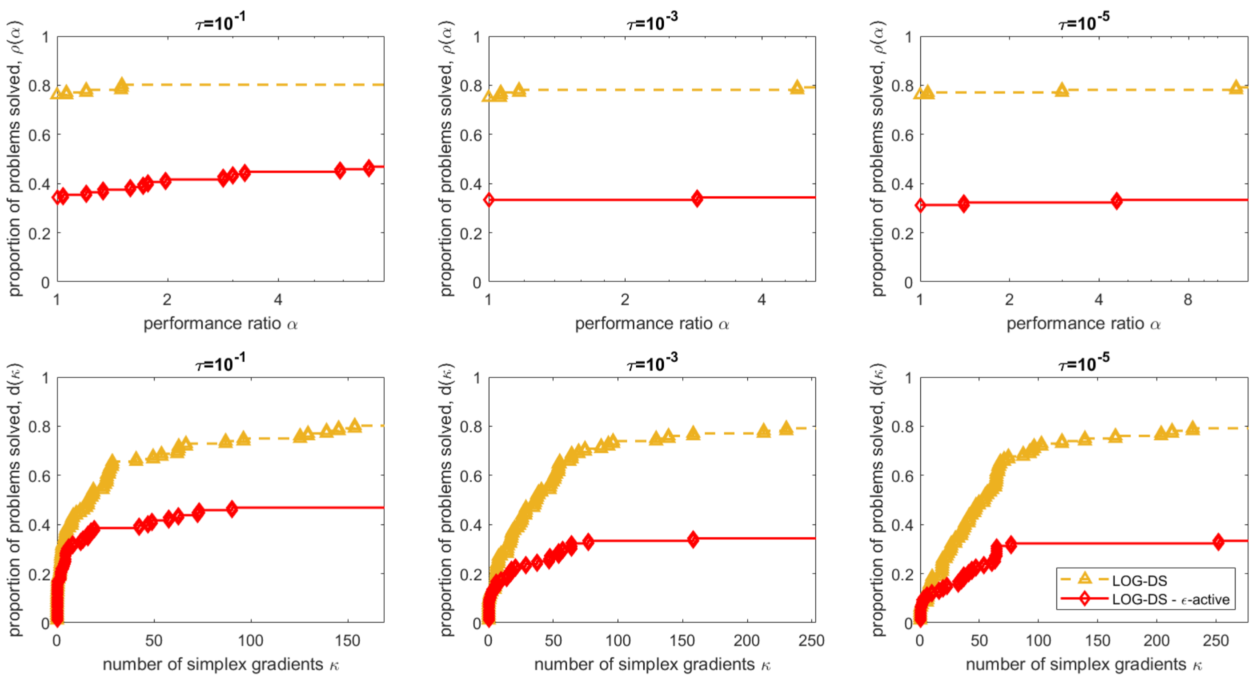

Figure 1: Performance (on top) and data (on bottom) profiles comparing LOG-DS using two different approaches to address linear inequality constraints.

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 Extreme Barrier

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 can only be adopted for problems without equality constraints and for which a strictly 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 tolerance equal to 10^{-8}.

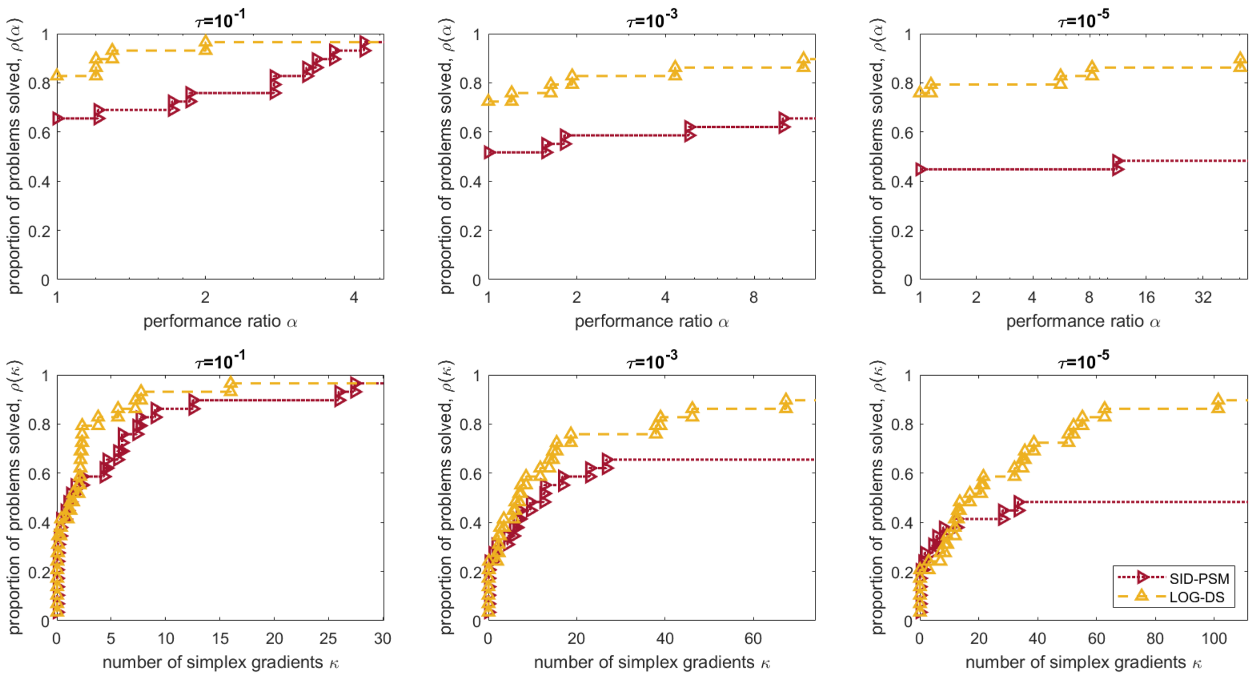

Figure 2: Performance (on top) and data (on bottom) profiles comparing LOG-DS and SID-PSM.

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 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` [8]. Additionally, the X-LOG-DFL algorithm [12], available through the DFL library as the LOGDFL package at `https://github.com/DerivativeFreeLibrary/LOGDFL`, was also tested. Comparison with LOG-DFL [12] 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 budget of 2000 function evaluations.

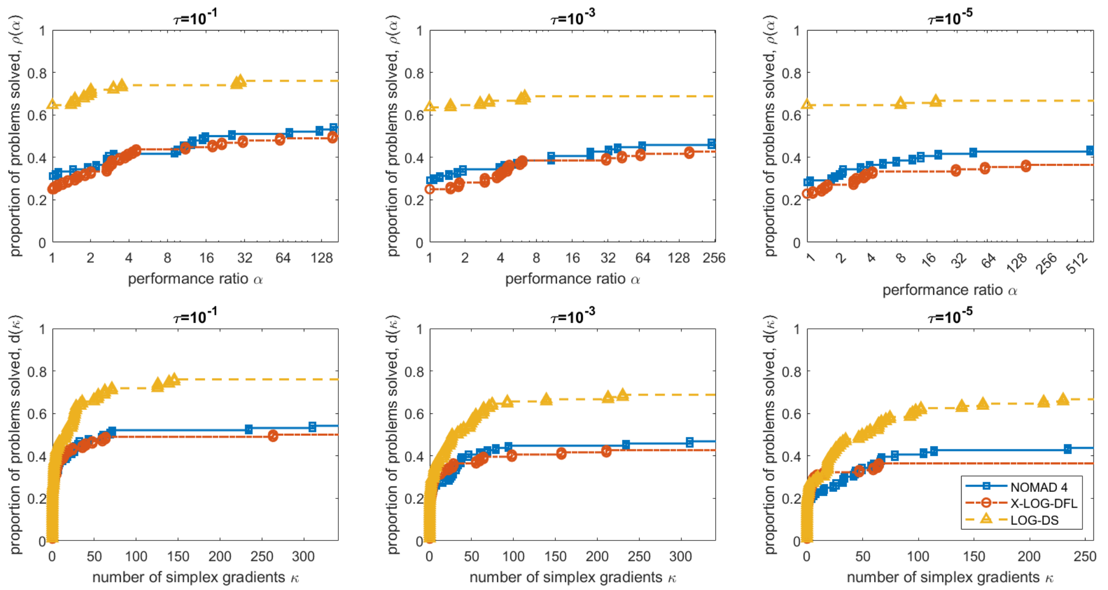

Figure 3: Performance (on top) and data (on bottom) comparing LOG-DS, NOMAD, and X-LOG-DFL, on the complete problem collection.

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 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.

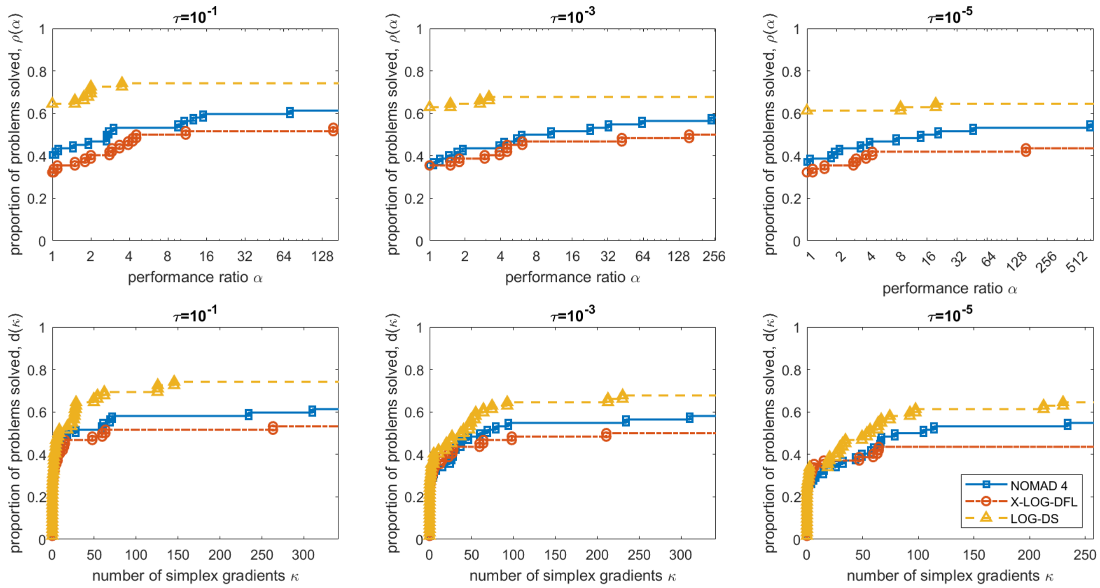

Figure 4: Performance (on top) and data (on bottom) profiles comparing LOG-DS, NOMAD, and LOG-DFL, on the subset of problems with only inequality constraints.

6 Conclusions

The primary objective of this work was to extend the approach introduced in [12] 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-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.

A Appendix

Lemma A.1. *Let $a \in \mathbb{R}_+$, $b \in \mathbb{R}_+$, $a + b > 0$ and $p \in [0, 1]$, it results*

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

Proof. We have $\dfrac{a}{a+b} \leq 1$ and $\dfrac{b}{a+b} \leq 1$, so

$$\begin{aligned}(a+b)^p &= \frac{(a+b)}{(a+b)^{1-p}} = \frac{a}{(a+b)^{1-p}} + \frac{b}{(a+b)^{1-p}} \\ &= a^p \left(\frac{a}{a+b}\right)^{1-p} + b^p \left(\frac{b}{a+b}\right)^{1-p} \\ &\leq a^p + b^p.\end{aligned}$$

□

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

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

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

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* $K_j \subseteq \{0, 1, \dots\}$ *and a positive scalar* $\bar{a}_j$ *exist such that*

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

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* $K \subseteq \mathbb{N}$ *such that:*

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

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

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

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

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

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

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

This is (obviously) true when $\ell = r$. Indeed, when $\ell = r$, there is K_r such that

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

Furthermore, we have

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

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

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

- $\lim_{k\to+\infty} a_k^1 = \lim_{k\to+\infty} a_k^2 = \cdots = \lim_{k\to+\infty} a_k^{r-1} = 0$;

- $\{a_k^i\}$, $i = r, \dots, \ell-1$, are not convergent to zero,

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

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

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

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

$$\lim_{\substack{k\to+\infty,\\ k\in\hat{K}}} \frac{a_k^\ell}{|a_k^{\hat{\imath}}|} = 0;$$

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

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

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

   (b) an infinite index set $K_1 \subseteq \hat{K}$ exists such that $\left\{\frac{a_k^\ell}{a_k^{\hat{\imath}}}\right\}_{k\in K_1}$ is unbounded.

   In the first case, an infinite index set $K_2 \subseteq \hat{K}$ exists such that $\left\{\frac{a_k^\ell}{a_k^{\hat{\imath}}}\right\}_{k\in K_2}$ is convergent. Then, the thesis follows by taking $s = \hat{\imath}$ and $K = K_2$.

   In the second case, we have

$$\lim_{\substack{k\to+\infty,\\ k\in K_1}} \frac{a_k^{\hat{\imath}}}{a_k^\ell} = 0.$$

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

$$\lim_{\substack{k\to+\infty,\\ k\in K_1}} \frac{a_k^i}{a_k^\ell} = \lim_{\substack{k\to+\infty,\\ k\in K_1}} \frac{a_k^i}{a_k^{\hat{\imath}}}\frac{a_k^{\hat{\imath}}}{a_k^\ell} = 0,$$

   and again the thesis is proved with $s = \ell$ and $K = K_1$.

Thus, the proof is concluded. □

Theorem B.2. *In the conditions of Theorem 3.7,*

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

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

the subsequences $\{\lambda_\ell(\mathbf{x}_k;\rho_k)\}_{k\in\hat{K}}$, $\ell = 1, \dots, m$ *and* $\{\mu_j(\mathbf{x}_k;\rho_k)\}_{k\in\hat{K}}$, $j = 1, \dots, p$ *are bounded.*

Proof. Recalling the expressions of $\lambda_\ell(\mathbf{x};\rho)$, $\ell=1,\dots,m$, and of $\mu_j(\mathbf{x};\rho)$, $j=1,\dots,p$, we can rewrite inequality (22) as

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

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

We will start by establishing that

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

In fact,

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

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

Then, there is $c_1 > 0$ such that

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

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

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

Suppose, in order to arrive to a contradiction, that c_2 does not exist. This would imply that there exists $K' \subseteq \hat{K} \subseteq K$ such that

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

Let us consider the case where

$$\lim_{\substack{k\to+\infty\\ k\in K'}} |g_\ell(\mathbf{x}_k)| = 0.$$

Since $g_\ell(\mathbf{x}_k) < 0$ and $g_\ell(\mathbf{y}_k^i) < 0$ for all $k \in K'$, by (45) there exists $\overline{k} \in \mathbb{N}$ such that, for all $k \geq \overline{k}$, $k \in K'$, 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 $i \in J_k$, 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 K 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

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

Thus,

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

which leads to a contradiction, proving (44).

Now, by considering the other case

$$\lim_{\substack{k\to+\infty\\ k\in K'}} |g_\ell(\mathbf{x}_k)| = c < +\infty,$$

we have

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

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 $(g_{\min})_k = \min_{\ell\in\mathcal{G}^{\log}}\{|g_\ell(\mathbf{x}_{k+1})|\} = \min_{\ell\in\mathcal{G}^{\log}}\{|g_\ell(\mathbf{x}_k)|\}$. Recalling (46), we get

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

Then, recalling Theorem 3.2, (41) is proved. Furthermore, we will establish

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

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

where $\mathbf{u}_k^i = \mathbf{x}_k + \tilde{t}_k^i(\mathbf{y}_k^i - \mathbf{x}_k)$ with $\tilde{t}_k^i\in(0,1)$, and $c_3>0$. Thus, using $\beta>1$, $\nu\in(1,2]$, $\|\mathbf{d}_k^i\| = 1$ for all $i\in J_k$, and recalling Theorem 3.2, (47) is proved. Therefore, we have that

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

We will now establish that

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

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

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

where $\mathbf{u}_k^j = \mathbf{x}_k + \tilde{t}_k^i(\mathbf{y}_k^i - \mathbf{x}_k)$, with $\tilde{t}_k^i \in (0,1)$ 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 $\hat{\alpha}_k^i \leq \alpha_k$ we can write

$$
\begin{aligned}
\left|\nu\left|\frac{h_j(\mathbf{x}_k)}{\rho_k}\right|^{\nu-1} - \nu\left|\frac{h_j(\mathbf{y}_k^i)}{\rho_k}\right|^{\nu-1}\right| &\leq \frac{\nu}{\rho_k^{\nu-1}}c_3\left(t_k^i\hat{\alpha}_k^i\right)^{\nu-1} \\
&\leq c_3\frac{\nu}{\rho_k^{\nu-1}}\left(t_k^i\alpha_k\right)^{\nu-1} \leq c_3\nu\left(\frac{\alpha_k}{\rho_k}\right)^{\nu-1} \\
&\leq c_3\nu\left(\frac{\rho_k^{\beta-1}}{\theta_\alpha}\right)^{\nu-1} = c_3\nu\theta_\alpha^{1-\nu}\rho_k^{(\beta-1)(\nu-1)}.
\end{aligned}
$$

Given that $\beta > 1$, $\nu \in (1,2]$, and recalling Theorem 3.2, we can conclude that (49) holds.

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

In fact, we can rewrite (40) as

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

Let

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

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

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

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

$$\lim_{\substack{k\to+\infty\\ k\in K'}} \frac{a_k^i}{|a_k^s|} = z_i, \quad |z_i| < +\infty, \quad i = 1, \dots, m+p \tag{53}$$

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 $\{a_k^s\}_{k\in K'}$ 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

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

Since $\lim_{\substack{k\to+\infty\\ k\in \hat{K}}} \mathbf{x}_k = \mathbf{x}^*$, Assumption 2.6 and Proposition 2.5 ensure the existence of $\varepsilon > 0$ such that for $k \in \tilde{K} \subseteq \hat{K}$ sufficiently large, $T_X(\mathbf{x}^*) = T_X(\mathbf{x}_k, \varepsilon) = \text{cone}(\mathbb{D}_k \cap T_X(\mathbf{x}_k, \varepsilon)) = \text{cone}(D^*)$.

Taking the limit for $k \to +\infty$ and $k \in \tilde{K}$, and using (48), (49), and (53), we obtain

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

We recall that $\mathbf{x}^*$ satisfies the MFCQ conditions. Let $\mathbf{d}$ be the direction satisfying condition (b) of Definition 3.3. For every $\mathbf{d} \in T_X(\mathbf{x}^*)$, there exist nonnegative numbers β_i such that

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

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

$$\left(\sum_{\ell=1}^{m} z_\ell \nabla g_\ell(\mathbf{x}^*) + \sum_{j=1}^{p} z_{m+j} \nabla h_j(\mathbf{x}^*)\right)^\top \mathbf{d} = \tag{58}$$

$$= \sum_{d^{*i}\in D^*} \beta_i \left(\sum_{\ell=1}^{m} z_\ell \nabla g_\ell(\mathbf{x}^*) + \sum_{j=1}^{p} z_{m+j} \nabla h_j(\mathbf{x}^*)\right)^\top \mathbf{d}^{*i}$$

$$= \sum_{d^{*i}\in D^*} \beta_i \sum_{\ell=1}^{m} z_\ell \nabla g_\ell(\mathbf{x}^*)^\top \mathbf{d}^{*i} + \sum_{d^{*i}\in D^*} \beta_i \sum_{j=1}^{p} z_{m+j} \nabla h_j(\mathbf{x}^*)^\top \mathbf{d}^{*i} \geq 0.$$

Considering Definition 3.3, the relation (58) becomes

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

Theorem 3.2 and the definition of z_ℓ for $\ell \in \{1, \dots, m\}$, guarantee

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

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

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

Therefore equation (56) becomes

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

using again Definition 3.3 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 provide three lemmas to support our implementation choices. In particular the first two results are the equivalent of Lemma 3.1 and Theorem 3.2, which are the fundamental results the theoretical analysis is based on. Furthermore, since the analysis in Section 3 is carried out by exploiting the subsequence of iterates where the penalty parameter is updated, we state one additional result to prove that the subsequence where both penalty parameters are simultaneously updated is infinite.

Lemma C.1. *Let* $\{\rho_k^{\texttt{log}}\}_{k\in\mathbb{N}}$, $\{\rho_k^{\texttt{ext}}\}_{k\in\mathbb{N}}$, *and* $\{\alpha_k\}_{k\in\mathbb{N}}$ *be the sequences of penalty parameters and stepsizes respectively, generated by algorithm LOG-DS. Assume that*

$$\lim_{k\to+\infty} \rho_k^{\texttt{log}} = \bar{\rho}_1 > 0$$

$$\lim_{k\to+\infty} \rho_k^{\texttt{ext}} = \bar{\rho}_2 > 0$$

Then,

$$\lim_{k\to+\infty} \alpha_k = 0. \tag{64}$$

Proof. Using similar arguments to Lemma 3.1. □

Lemma C.2. *Let* $\{\rho_k^{\texttt{log}}\}_{k\in\mathbb{N}}$, $\{\rho_k^{\texttt{ext}}\}_{k\in\mathbb{N}}$, *and* $\{\alpha_k\}_{k\in\mathbb{N}}$ *be the sequences of penalty parameters and stepsizes generated by LOG-DS. Then,*

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

$$\lim_{k\to+\infty} \rho_k^{\texttt{ext}} = 0, \tag{66}$$

$$\lim_{k\to+\infty} \alpha_k = 0. \tag{67}$$

Proof. By (30) we have that $\rho_k^{\texttt{ext}} \to 0$ only if $\rho_k^{\texttt{log}} \to 0$, so we can use the same arguments of Theorem 3.2 to prove (65). Now, by (29) and (30) it is clear that (65) implies (66) Finally, we can use similar arguments of Theorem 3.2 to prove (67), using Lemma C.1 instead of Lemma 3.1. □

Lemma C.3. *Let* $\{\rho_k^{\texttt{log}}\}_{k\in\mathbb{N}}$, $\{\rho_k^{\texttt{ext}}\}_{k\in\mathbb{N}}$, *and* $\{\alpha_k\}_{k\in\mathbb{N}}$ *be the sequences of penalty parameters and stepsizes generated by LOG-DS. Let* $K_{\texttt{log}} = \{k \in \mathbb{N} : \rho_{k+1}^{\texttt{log}} < \rho_k^{\texttt{log}}\}$ *and* $K_{\texttt{ext}} = \{k \in \mathbb{N} : \rho_{k+1}^{\texttt{ext}} < \rho_k^{\texttt{ext}}\}$. *Then* $|K_{\texttt{log}} \cap K_{\texttt{ext}}| = +\infty$.

Proof. Using (66) and (32), we get $|K_{\texttt{ext}}| = +\infty$. Let $k \in K_{\texttt{ext}}$. From (30), we have

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

so that for all $k \in K_{\texttt{ext}}$ we also have $k \in K_{\texttt{log}}$, and $|K_{\texttt{log}} \cap K_{\texttt{ext}}| = |K_{\texttt{ext}}| = +\infty$, concluding the proof. □

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] C. Audet, S. Le Digabel, V. Rochon Montplaisir, and C. Tribes. *The NOMAD project*. URL: `https://www.gerad.ca/nomad`.

[9] A. Benchakroun, J. Dussault, and A. Mansouri. “A two parameter mixed interior-exterior penalty algorithm”. In: *ZOR - Meth. and Models of Oper. Res.* 41 (1995), pp. 25–55.

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

[11] 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.

[12] A. Brilli, G. Liuzzi, and S. Lucidi. *An interior point method for nonlinear constrained derivative-free optimization*. 2021. arXiv: `2108.05157 [math.OC]`.

[13] 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.

[14] 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.

[15] 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.

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

[17] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. Vol. 4. Classics in Applied Mathematics. Philadelphia: Society for Industrial and Applied Mathematics, 1990.

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

[19] 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.

[20] 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.

[21] 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.

[22] 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.

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

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

[25] 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.

[26] 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.

[27] 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.

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

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

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