Smoothing ℓ_1-exact penalty method for intrinsically constrained Riemannian optimization problems*

Ernesto G. Birgin[†] Orizon P. Ferreira [‡] Gabriel Haeser[§] Nelson Maculan[¶]
Lennin Mallma Ramirez[¶] Leandro F. Prudente[‡]

January 8, 2025

Abstract

This paper deals with the Constrained Riemannian Optimization (CRO) problem, which involves minimizing a function subject to equality and inequality constraints on Riemannian manifolds. The study aims to advance optimization theory in the Riemannian setting by presenting and analyzing a penalty-type method for solving CRO problems. The proposed approach is based on techniques that involve smoothing the classical ℓ_1-exact penalty function. This penalty-type method extends previous research by incorporating different smoothing functions, refining the penalty multipliers, and relaxing the constraints qualifications necessary for convergence. The method uses the extended Mangasarian-Fromovitz constraint qualification to ensure boundedness of Lagrange multipliers and global convergence to feasible and optimal solutions. In addition, under the assumption that the limit points are feasible, it is shown that these points satisfy the Approximate KKT (AKKT) conditions. Furthermore, when AKKT is combined with a weak constraint qualification, it is proved that the limit points satisfy the KKT conditions. Preliminary numerical experiments are conducted to demonstrate the effectiveness of the proposed method, which indicates that the method effectively addresses the complexity associated with CRO problems.

Key words: Nonlinear optimization, penalty methods, convergence, numerical experiments.

AMS subject classifications: 90C30, 65K05.

1 Introduction

This paper addresses the Constrained Riemannian Optimization (CRO) problem, defined as:

$$\underset{p\in\mathcal{M}}{\text{Minimize}}\, f(p) \text{ subject to } h(p)=0 \text{ and } g(p)\leq 0, \tag{1}$$

*This work has been supported by FAPERJ (grant E-26/205.684/2022), FAPESP (grants 2013/07375-0, 2022/05803-3 and 2023/08706-1), COPPETEC Foundation, and CNPq (grants 302073/2022-1 and 304666/2021-1).

[†]Department of Computer Science, Institute of Mathematics and Statistics, University of São Paulo, Rua do Matão, 1010, Cidade Universitária, 05508-090, São Paulo, SP, Brazil. e-mail: egbirgin@ime.usp.br

[‡]Institute of Mathematics and Statistics, Federal University of Goiás, 74001-970, Goiânia, Brazil. e-mail: {orizon,lfprudente}@ufg.br

[§]Department of Applied Mathematics, Institute of Mathematics and Statistics, University of São Paulo, Rua do Matão, 1010, Cidade Universitária, 05508-090, São Paulo, SP, Brazil. e-mail: ghaeser@ime.usp.br

[¶]Systems Engineering and Computer Science Program-Applied Mathematics (COPPE/IM), Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil. email: {maculan,lenninmr}@cos.ufrj.br

where $\mathcal{M}$ is an n-dimensional smooth and complete Riemannian manifold, the functions $f\colon \mathcal{M} \to \mathbb{R}$, $h = (h_1, \ldots, h_s)\colon \mathcal{M} \to \mathbb{R}^s$ and $g = (g_1, \ldots, g_m)\colon \mathcal{M} \to \mathbb{R}^m$ are continuously differentiable. Optimization methods for the CRO problem, as well as the theoretical foundations supporting them in the Riemannian context, are still in early stages of development. Despite considerable progress in recent years, this field remains nascent, with many challenges and open questions yet to be explored. The objective here is to advance the CRO field by presenting and analyzing a penalty-type method for solving Problem (1), inspired by [33, Algorithm 2]. Penalty techniques have a well-established history in Euclidean spaces, addressing numerous practical problems (see, for example, [11, 18]). It is reasonable to expect that the success of penalty techniques in the Euclidean context will extend to the Riemannian setting, providing the primary motivation for this research.

In contrast to the significant advancements in unconstrained Riemannian optimization, the development of theory and methods for CRO remains comparatively limited. The extension of the Karush-Kuhn-Tucker (KKT) optimality conditions to the Riemannian setting was first introduced in 2014 [54]. Subsequently, an intrinsic approach to KKT conditions and the analysis of various constraint qualifications were explored in [5]. More recent developments include the introduction of the Approximate KKT (AKKT) conditions by [53] and an extensive investigation of strict constraint qualifications by [3], providing a comprehensive framework for algorithmic convergence theory in CRO. These foundational studies have been instrumental in supporting the global convergence of algorithms, such as the augmented Lagrangian method in [3], which builds upon the initial work of [33] that introduced penalty methods for CRO. In response to these foundational results, several approaches for CRO have emerged. For example, an exact penalty method for constrained problems on Stiefel manifolds was proposed in [27], addressing constraint qualifications and optimality conditions. Similarly, [15] developed an inexact augmented Lagrangian framework for nonsmooth optimization on Riemannian submanifolds, while [40] introduced a Riemannian sequential quadratic optimization algorithm with an ℓ_1-penalty function. The primal-dual interior point method for Riemannian manifolds, presented in [30], generalizes the classical primal-dual framework for nonlinear programming. In [26], the projection robust Wasserstein distance is reformulated as a CRO problem over the Cartesian product of Stiefel manifolds, where a Riemannian version of an exponential augmented Lagrangian method is introduced with established global convergence. Furthermore, [20] addresses the minimization of smooth functions under smooth equality constraints using an algorithm based on Fletcher's ℓ_1-exact penalty function. Additional studies complementing these methods are found in [33, 30].

Considering the inherent complexities of designing projection methods within the Riemannian context, it is expected that future research on the CRO problem will largely focus on penalty-type and augmented Lagrangian-type methods. These approaches are promising due to their straightforward design and analysis, as well as their efficacy in managing the complexities associated with Riemannian geometry. This prediction is supported by the preceding discussion, highlighting the advantages of advanced optimization methods in addressing practical problems within the Riemannian framework that are not adequately solved by existing Euclidean theories. In accordance with current trends, this study proposes a penalty-type method for solving Problem (1). The method builds upon techniques involving the smoothing of the exact penalty function, which has been extensively explored in Euclidean spaces. To the best of our knowledge, this methodology originates from [49] and was further developed in [6, 29, 50]. Additional contributions to this topic include [19, 21, 32, 34, 36, 37, 41, 29, 43, 50, 51, 52, 56]. The penalty-type method presented in this paper represents an advancement over [33, Algorithm 2] in three aspects. First, it accommodates a broader range of functions for smoothing the exact penalty function, enhancing both its flexibility and applicability. Second, unlike previous methods that uniformly adjusted all penalty multipliers across constraints, the proposed approach employs a more selective strategy. It adjusts the penalty multiplier based on the feasibility measure of the violated constraint, offering a more nuanced and adaptive treatment of constraint handling. Third, and most importantly, the analysis of the method is conducted under weaker assumptions than those used in prior work, which relied on the linear independence constraint qualification (LICQ). Instead, it employs the extended Mangasarian-Fromovitz constraint qualification (EMFCQ), broadening the scope of applicability and enriching the analytical framework. Additionally, assuming the feasibility of the limit points, it is shown that these points satisfy the Approximate KKT (AKKT) conditions. Furthermore, when AKKT points are combined with either the Relaxed Constant Positive Linear Dependence (Relaxed-CPLD) condition or the Constant Rank of the Subspace Component (CRSC) condition—both recently introduced in [3]—it follows that the limit points satisfy the KKT conditions. The method, along with its respective designs and analyses, are distinguished for their simplicity. We will delve into further details on this matter below.

The penalty-type approaches are well-established methodologies for addressing constrained optimization problems. These techniques transform the original problem into a sequence of unconstrained subproblems by incorporating penalty functions, which enforce constraints indirectly. By penalizing constraint violations, these methods enable the application of comprehensive unconstrained optimization algorithms and theoretical frameworks to the transformed problem. This approach facilitates the efficient handling of the complexities introduced by constraints, allowing for the utilization of robust, well-established unconstrained optimization techniques, as the ones [1, 9, 17, 22, 23, 25, 46]. In this paper, we focus on one of the most popular exact penalty functions for the proposed method, the so-called ℓ_1-*exact penalty function*, which is defined as

$$E(p, \rho) := f(p) + \rho \sum_{i=1}^{s} |h_i(p)| + \rho \sum_{j=1}^{m} \max\{g_j(p), 0\}. \tag{2}$$

This function was introduced in [55], where it was proven that if $E(\cdot, \xi)$ has a minimizer for some $\xi > 0$, then the same will be true to $E(\cdot, \rho)$, for any $\rho > \xi$. Moreover, it is widely acknowledged that, under mild assumptions, if ρ is sufficiently large, the minimizers of $E(\cdot, \rho)$ that are feasible to Problem (1) also satisfy the KKT optimality conditions, as demonstrated in [39, Theorem 17.4]. The primary challenge associated with the exact penalty function lies in its non-differentiability, which impedes the application of efficient minimization algorithms to solve the subproblems. To circumvent this drawback, one of the important tools is the smoothing approach, which is based on creating a differentiable approximation of the non-differentiable term in (2). Thus, instead of solving a sequence of non-differentiable problems, a sequence of differentiable problems is solved. The first studies on the smoothing approach in Euclidean space are due to [6, 29, 49] and in the Riemannian setting to [33]. In this paper, we follow this approach but with a slight modification to the function (2). Instead of penalizing the entire sum with a single parameter ρ, we penalize each function in the sum with a different parameter, resulting in a function like

$$E(p, \rho_1, \ldots, \rho_s, \sigma_1, \ldots, \sigma_m) = f(p) + \sum_{i=1}^{s} \rho_i |h_i(p)| + \sum_{j=1}^{m} \sigma_j \max\{g_j(p), 0\}, \tag{3}$$

and then we apply the smoothing approach to each functions $|h_i(p)|$ and $\max\{g_j(p), 0\}$ individually. The proposed method involves solving inner subproblems by minimizing a smooth approximation of the ℓ_1-exact penalty function and updating the penalty parameters based on the improvement of an infeasibility measure of the violated constraints. The smoothness parameters are also updated in each iteration.

The organization of this paper is as follows: In Section 2, we recall the notations and fundamental concepts of Riemannian manifolds used throughout the paper and revisit the topic of smooth approximations of the absolute value function. Section 3 introduces a class of penalty functions designed to address Problem (1) and discusses their properties. In Section 4, we analyze a penalty-type method proposed to solve Problem (1). Preliminary numerical experiments are presented in Section 5, and final remarks are offered in Section 6.

2 Basics concepts and terminology

In this section, we recall some notations and basic concepts used throughout the paper. The concepts of Riemannian manifolds can be found, for example, in [13, 31, 44, 47, 48].

Let $\mathcal{M}$ be an n-dimensional smooth Riemannian manifold. Denote the *tangent space* at a point p by $T_p\mathcal{M}$. Assume also that $\mathcal{M}$ has a *Riemannian metric* denoted by $\langle \cdot, \cdot \rangle$ and the corresponding *norm* by $\|\cdot\|$. For $f\colon U \to \mathbb{R}$ a differentiable function with derivative $df(\cdot)$, where U is an open subset of the manifold $\mathcal{M}$, the Riemannian metric induces the mapping $f \mapsto \operatorname{grad} f$ which associates its *gradient vector field* via the following rule $\langle \operatorname{grad} f(p), X(p)\rangle := df(p)X(p)$, for all $p \in U$and all vector field X in $\mathcal{M}$. The *Riemannian distance* between p and q is denoted by $d(p, q)$. This distance induces the original topology on $\mathcal{M}$, namely $(\mathcal{M}, d)$ is a complete metric space and the bounded and closed subsets are compact. The *open* and *closed balls* of radius $r > 0$, centered at p, are respectively defined by $B_r(p) := \{q \in \mathcal{M} :\ d(p, q) < r\}$ and $B_r[p] := \{q \in \mathcal{M} :\ d(p, q) \leq r\}$. A Riemannian manifold is *complete* if its geodesics $\zeta(t)$ are defined for any value of $t \in \mathbb{R}$. From now on, $\mathcal{M}$ *denotes an n-dimensional smooth and complete Riemannian manifold.* For that, given two points $p, q \in \mathcal{M}$, β_{pq} denotes the set of all geodesic segments $\zeta\colon [0, 1] \to \mathcal{M}$ with $\zeta(0) = p$ and $\zeta(1) = q$. A function $f\colon \mathcal{M} \to \mathbb{R}$ is said to be *convex* if, for any $p, q \in \mathcal{M}$ and $\zeta \in \beta_{pq}$, the composition $f \circ \zeta\colon [0, 1] \to \mathbb{R}$ is convex, i.e., $(f \circ \zeta)(t) \leq (1 - t)f(p) + tf(q)$, for all $t \in [0, 1]$ and f is concave, if $-f$ is convex. If f is convex and differentiable, then critical points $\bar{p}$ of f are global minimizers if and only if $\operatorname{grad} f(\bar{p}) = 0$.

In the following, we introduce the positive-linearly dependent condition, an important concept in understanding constraint qualifications, which will be explored further in the following section; for more details see [3].

Definition 1. *Let $U = \{u_1, \ldots, u_s\}$ and $V = \{v_1, \ldots, v_m\}$ be finite multisets on $T_p\mathcal{M}$, that is, repetition of the same element is allowed. Then $U \cup V$ is said to be positive-linearly dependent if there exist $a = (a_1, \ldots, a_s) \in \mathbb{R}^s$ and $b = (b_1, \ldots, b_m) \in \mathbb{R}^m_+$ such that $(a, b) \neq 0$ and $\sum_{i=1}^{s} a_i u_i + \sum_{j=1}^{m} b_j v_j = 0$. Otherwise, $U \cup V$ is said to be positive-linearly independent.*

Next, we recall a general version of Farkas' Lemma, also known as Motzkin's transposition theorem, with the statement adapted from [45] to suit our specific application of interest.

Lemma 1. *Let $v \in T_p\mathcal{M}$ and $u_i, v_j \in T_p\mathcal{M}$, for all $i \in \{1, \ldots, s\}$ and $j \in \mathcal{A} \subset \{1, \ldots, m\}$. The system: $\langle v, d\rangle < 0$, $\langle u_i, d\rangle = 0$, for $i \in \{1, \ldots, s\}$ and $\langle v_j, d\rangle \leq 0$, for $j \in \mathcal{A}$, has no solution $d \in T_p\mathcal{M}$ if and only if there exist $\bar{\lambda}_i$, for all $i \in \{1, \ldots, s\}$, and $\bar{\mu}_j \geq 0$, for $j \in \mathcal{A}$, such that $v + \sum_{i=1}^{s} \bar{\lambda}_i u_i + \sum_{j\in\mathcal{A}} \bar{\mu}_j v_j = 0$.*

Throughout the paper, we define the set of positive integers as $\mathbb{N} := \{1, 2, \ldots\}$.

2.1 Concepts and terminology of constrained optimization problems

In this section, we present the notations and definitions related to the CRO Problem (1) that are used throughout the paper. The *feasible set* Ω *of Problem* (1) is defined by

$$\Omega := \{p \in \mathcal{M} : \ h(p) = 0, \ g(p) \leq 0\},$$

which is closed. For a given $\lambda \in \mathbb{R}^s$ and $\mu \in \mathbb{R}^m_+$, the *Lagrangian function* $L(\cdot, \lambda, \mu) \colon \mathcal{M} \to \mathbb{R}$ associated with Problem (1) is defined by

$$L(p, \lambda, \mu) := f(p) + \sum_{i=1}^{s} \lambda_i h_i(p) + \sum_{j=1}^{m} \mu_j g_j(p). \tag{4}$$

Since f and g are continuously differentiable, the gradient[1] of $L(\cdot, \lambda, \mu) \colon \mathcal{M} \to \mathbb{R}$ is given by

$$\operatorname{grad} L(p, \lambda, \mu) = \operatorname{grad} f(p) + \sum_{i=1}^{s} \lambda_i \operatorname{grad} h_i(q) + \sum_{i=1}^{m} \mu_j \operatorname{grad} g_j(p). \tag{5}$$

For a given point $\bar{p} \in \Omega$, let $\mathcal{A}(\bar{p})$ be the set of *indexes of active inequality constraints*, that is,

$$\mathcal{A}(\bar{p}) := \{j \in \{1, \ldots, m\} : \ g_j(\bar{p}) = 0\} .$$

We say that the *Karush/Kuhn-Tucker (KKT) conditions* are satisfied at $\bar{p} \in \mathcal{M}$ when there exist so-called *Lagrange multipliers* $\bar{\lambda} \in \mathbb{R}^s$ and $\bar{\mu} \in \mathbb{R}^m_+$ such that the following three conditions hold:

(i) $\operatorname{grad} L(\bar{p}, \bar{\lambda}, \bar{\mu}) = 0$, i.e., $\operatorname{grad} f(\bar{p}) + \sum_{j=1}^{s} \bar{\lambda}_j \operatorname{grad} h_j(\bar{p}) + \sum_{j=1}^{m} \bar{\mu}_j \operatorname{grad} g_j(\bar{p}) = 0$;

(ii) $\bar{\mu}_j = 0$, for all $j \notin \mathcal{A}(\bar{p})$, i.e, $\bar{\mu}_j g_j(\bar{p}) = 0$, for all $j \in \{1, \ldots, m\}$;

(iii) $\bar{p} \in \Omega$, i.e., $h(\bar{p}) = 0$ and $g(\bar{p}) \leq 0$.

In this case, we refer to $\bar{p} \in \mathcal{M}$ as a KKT point. A constraint qualification (CQ) is a condition regarding the structure of the feasible set, ensuring that every local minimum is a KKT point. It is important to note that without a CQ at a local optimum $\bar{p}$, the existence of Lagrange multipliers satisfying the KKT conditions cannot be guaranteed; for further details, refer to [3, 5, 53]. Below we introduce a useful CQ for studying Problem (1). To simplify the notation, we first present the following definitions: For a given $\bar{p} \in \mathcal{M}$ define

$$\mathcal{A}_+(\bar{p}) := \{j \in \{1, \ldots, m\} : \ g_j(\bar{p}) \geq 0\} ,$$

and, for each $q \in \mathcal{M}$, $\mathcal{I} \subseteq \{1, \ldots, s\}$ and $\mathcal{J} \subseteq \{1, \ldots, m\}$, define the set $\mathcal{B}(q, \mathcal{I}, \mathcal{J}) \subset T_q\mathcal{M}$ as

$$\mathcal{B}(q, \mathcal{I}, \mathcal{J}) := \{\operatorname{grad} h_i(q) : \ i \in \mathcal{I}\} \cup \{\operatorname{grad} g_j(q) : \ j \in \mathcal{J}\}.$$

Definition 2. *The point $\bar{p} \in \mathcal{M}$ is said to satisfy the extended Mangasarian-Fromovitz constraint qualification (EMFCQ) if the set $\mathcal{B}(\bar{p}, \{1, \ldots, s\}, \mathcal{A}_+(\bar{p}))$ is positive-linearly independent.*

If in the above definition we consider $\bar{p} \in \Omega$, then it becomes the Mangasarian-Fromovitz constraint qualification (MFCQ) considered in Riemannian setting in [5]. For more details and applications of EMFCQ in Euclidean context, see, for example, [28]. Next we state strict constraint qualifications for the CRO Problem (1), which were introduced in [3]. We begin by recalling the *relaxed Constant Positive Linear Dependence Condition* (Relaxed-CPLD), which is stated as follows:

[1] Although the Lagrangian L is a function of three variables, to simplify the notation, we denote by $\operatorname{grad} L$ the gradient with respect to the first variable.

Definition 3. *The point $\bar{p} \in \Omega$ is said to satisfy Relaxed-CPLD (RCPLD), if there exists $\epsilon > 0$ such that the following two conditions hold:*

(a) the rank of $\{\operatorname{grad} h_i(\bar{p}) : \ i = 1, \ldots, s\}$ *is constant for all* $p \in B_\epsilon(\bar{p})$;

(b) Let $\mathcal{K} \subset \{1, \ldots, s\}$, *such that* $\{\operatorname{grad} h_i(\bar{p}) : \ i \in \mathcal{K}\}$ *is a basis for the subspace generated by* $\{\operatorname{grad} h_i(\bar{p}) : \ i = 1, \ldots, s\}$. *For all* $\mathcal{J} \subset \mathcal{A}(\bar{p})$, *if* $\mathcal{B}(\bar{p}, \mathcal{K}, \mathcal{J})$ *is positive-linearly dependent, then* $\mathcal{B}(p, \mathcal{K}, \mathcal{J})$ *is linearly dependent, for all* $p \in B_\epsilon(\bar{p})$.

To state the *Constant Rank of the Subspace Component* (CRSC), for $\bar{p} \in \Omega$, we define the *linearized cone* $\mathcal{L}(\bar{p})$ as follows

$$\mathcal{L}(\bar{p}) := \big\{ v \in T_{\bar{p}}\mathcal{M} : \ \langle \operatorname{grad} h_i(\bar{p}), v \rangle = 0, \ i = 1, \ldots, s; \ \langle \operatorname{grad} g_j(\bar{p}), v \rangle \leq 0, \ j \in \mathcal{A}(\bar{p}) \big\},$$

and its polar $\mathcal{L}(\bar{p})^\circ$ is given by

$$\mathcal{L}(\bar{p})^\circ = \Big\{ v \in T_{\bar{p}}\mathcal{M} : \ v = \sum_{i=1}^{s} \lambda_i \operatorname{grad} h_i(\bar{p}) + \sum_{j \in \mathcal{A}(\bar{p})} \mu_j \operatorname{grad} g_j(p), \ \mu_j \geq 0, \lambda_i \in \mathbb{R} \Big\}.$$

Definition 4. *Let $\bar{p} \in \Omega$ and define the index set $\mathcal{J}_-(\bar{p}) := \{ j \in \mathcal{A}(\bar{p}) : \ -\operatorname{grad} g_j(\bar{p}) \in \mathcal{L}(\bar{p})^\circ \}$. The point $\bar{p}$ is said to satisfy CRSC if there exists $\epsilon > 0$ such that the rank of $\mathcal{B}(p, \{1, \ldots, s\}, \mathcal{J}_-(\bar{p}))$ is constant for all $p \in B_\epsilon(\bar{p})$.*

In [4], it was proven that the CRSC condition is strictly weaker than RCPLD in the Euclidean setting. While we do not yet have a similar proof in the Riemannian setting, [3] presents an example demonstrating that CRSC does not imply RCPLD in any Riemannian manifold $\mathcal{M}$ of dimension $n \geq 2$. We conclude this section by revisiting the concept of the Approximate-KKT condition introduced in [53], which will be instrumental in the analysis of the smooth ℓ_1-exact penalty method.

Definition 5. *A point $\bar{p} \in \Omega$ is said to be an Approximate KKT (AKKT) point for Problem* (1) *if there exist sequences $(p^k)_{k \in \mathbb{N}} \subset \mathcal{M}$, $(\lambda^k)_{k \in \mathbb{N}} \subset \mathbb{R}^s$ and $(\mu^k)_{k \in \mathbb{N}} \subset \mathbb{R}^m_+$ such that*

(i) $\lim_{k \to \infty} p^k = \bar{p}$,

(ii) $\lim_{k \to \infty} \operatorname{grad} L(p^k, \lambda^k, \mu^k) = 0$,

(iii) $\lim_{k \to \infty} \mu_j^k = 0$, *for all* $j \notin \mathcal{A}(\bar{p})$.

2.2 Smooth approximation to absolute value function

In this section, we revisit the topic of smooth approximations to the absolute value function. We focus on presenting some well-known examples from the literature. We denote the *max function* by $t_+ := \max\{0, t\}$ and the *absolute value function* by $|t| := \max\{t, -t\}$, for $t \in \mathbb{R}$. Considering that

$$t_+ = \frac{1}{2}(t + |t|),$$

the problem of approximating the max function is equivalent to approximating the absolute value function. We begin with the following definition.

Definition 6. *Let $\phi \colon \mathbb{R} \times \mathbb{R}_{++} \to \mathbb{R}$ be a continuously differentiable function. The function $\phi_\tau \colon \mathbb{R} \to \mathbb{R}$ defined by $\phi_\tau(t) := \phi(t, \tau)$ is called a smooth absolute value function if it satisfies the following conditions:*

(i) $\lim_{(t,\tau)\to(\bar{t},+\infty)} \phi_\tau(t) = |\bar{t}|$, *for all* $\bar{t} \in \mathbb{R}$;

(ii) $\phi_\tau(t) > 0$, *for all* $t \neq 0$;

(iii) ϕ'_τ *is not decreasing,* ϕ'_τ *is strictly increasing in* $(-\delta_\tau, \delta_\tau)$, *for some* $\delta_\tau > 0$, *and* $\phi'_\tau(0) = 0$;

(iv) $0 < \phi'_\tau(t) \leq 1$, *for all* $t > 0$;

(v) $-1 \leq \phi'_\tau(t) < 0$, *for all* $t < 0$;

(vi) ϕ_τ *is convex, for all* $\tau > 0$;

(vii) $\lim_{t\to+\infty} \phi'_\tau(t) = 1$ *and* $\lim_{t\to-\infty} \phi'_\tau(t) = -1$, *for all* $\tau > 0$;

(viii) *There hold* $\lim_{(t,\tau)\to(\bar{t},+\infty)} \phi'_\tau(t) = -1$, *for* $\bar{t} < 0$ *and* $\lim_{(t,\tau)\to(\bar{t},+\infty)} \phi'_\tau(t) = 1$, *for* $\bar{t} > 0$.

Here are several well-known examples of smooth absolute value functions from the literature.

Example 1. *In the following, we present several examples of smooth approximations for the absolute value function:*

(i) $\phi_{1,\tau}(t) := (|t|^r + \frac{1}{\tau^{r/2}})^{\frac{1}{r}}$, *for any* $r > 1$

(ii) $\phi_{2,\tau}(t) := \frac{1}{\tau} \ln(e^{\tau t} + e^{-\tau t})$

(iii) $\phi_{3,\tau}(t) = \begin{cases} t & \text{if } t \geq \frac{1}{2\tau}, \\ \tau t^2 + \frac{1}{4\tau} & \text{if } -\frac{1}{2\tau}, < t < \frac{1}{2\tau}, \\ -t & \text{if } t \leq -\frac{1}{2\tau} \end{cases}$

(iv) $\phi_{4,\tau}(t) = \begin{cases} \frac{\tau}{2} t^2 & \text{if } |t| \leq \frac{1}{\tau}, \\ |t| - \frac{1}{2\tau} & \text{if } |t| > \frac{1}{\tau} \end{cases}$

(v) $\phi_{5,\tau}(t) := (|t|^r + \frac{1}{\tau^{r/2}})^{\frac{1}{r}} - \frac{1}{\tau^{1/2}}$, *for any* $r > 1$

(vi) $\phi_{6,\tau}(t) := \frac{1}{\tau} \ln(e^{\tau t} + e^{-\tau t}) - \frac{1}{\tau} \ln(2) = \frac{1}{\tau} \ln(\cosh(\tau t))$.

The functions in items (i) with $r = 2$, (ii), (iv), and (v) with $r = 2$ in Example 1 have been used to study absolute value equations, as discussed in [42], where their properties are presented, see also [36, 37, 43]. In particular,

$$0 = \phi_{4,\tau}(0) < \phi_{4,\tau}(t) < |t| < \phi_{3,\tau}(t), \qquad 0 < |t| - \phi_{4,\tau}(t) < \frac{1}{2\tau}, \qquad \forall t \in \mathbb{R}, t \neq 0. \tag{6}$$

The function in item (ii) in Example 1 was used in the context of linear complementarity problems, see [14]. For future reference, we also note the following properties of the function $\phi_{6,\tau}$ in Example 1

$$0 = \phi_{6,\tau}(0) < \phi_{6,\tau}(t) < |t|, \qquad 0 < |t| - \phi_{6,\tau}(t) < \frac{\ln 2}{\tau}, \qquad \forall t \in \mathbb{R}, t \neq 0. \tag{7}$$

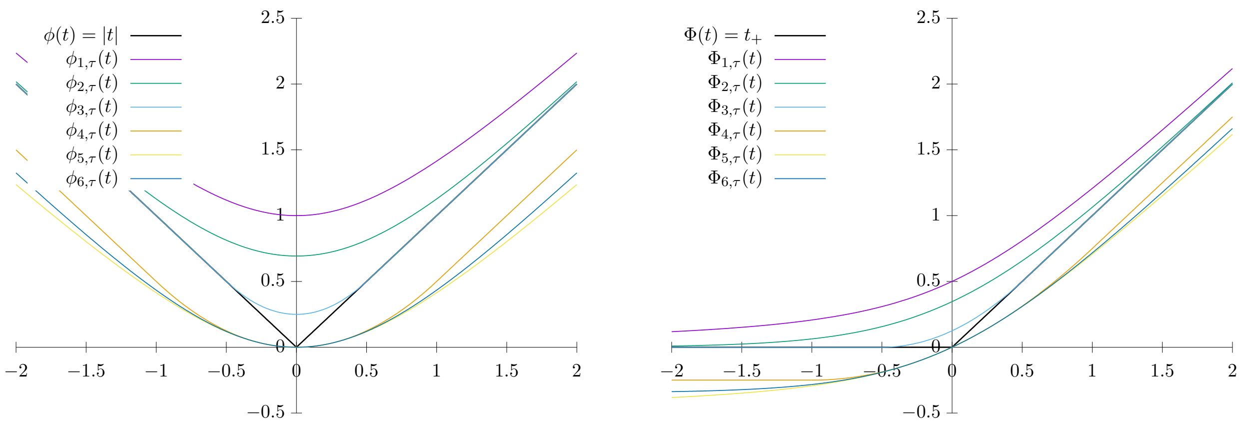
Some results, such as those in Section 3.1, require additional assumptions about the smoothing function, which are stated as follows: For a given function ϕ_τ, consider the following condition:

$$0 \leq \phi_\tau(t) \leq |t|, \qquad 0 \leq |t| - \phi_\tau(t) \leq \frac{\kappa}{\tau}, \qquad \forall t \in \mathbb{R}, \quad \tau > 0, \tag{8}$$

for some $\kappa > 0$. The functions $\phi_{4,\tau}$ and $\phi_{6,\tau}$ satisfy (8). For future reference, we state the following lemma, whose proof is straightforward.

Lemma 2. *If ϕ_τ satisfies* (8)*, then $0 < t_+ - (1/2)(t + \phi_\tau(t)) \leq \frac{\kappa}{2\tau}$, for all $t \in \mathbb{R}$ and $\tau > 0$.*

We conclude this section with Figure 1, which illustrates the functions $\phi(t) = |t|$, $\Phi(t) = t_+$, and the corresponding smoothed functions $\phi_{i,\tau}(t)$ and $\Phi_{i,\tau}(t) = (t + \phi_{i,\tau}(t))/2$ from Example 1 for $i = 1, \ldots, 6$ with a smoothing parameter $\tau = 1$ and $r = 2$ in $\phi_{1,\tau}$ and $\phi_{5,\tau}$. The functions $\phi_{i,\tau}$ provide an upper approximation to ϕ for $i = 1, 2, 3$ and a lower approximation for $i = 4, 5, 6$. Consequently, the functions $\Phi_{i,\tau}$ approximate Φ from above for $i = 1, 2, 3$ and from below for $i = 4, 5, 6$.



(a) Approximations of $\phi(t) = |t|$

(b) Approximations of $\Phi(t) = t_+$

Figure 1: Approximations of $\phi(t) = |t|$ and $\Phi(t) = t_+$ as described in Example 1.

3 Smooth ℓ_1-exact penalty function

In this section, we introduce a class of penalty functions designed to address Problem (1), by smoothing the ℓ_1-exact penalty function in (3). To achieve this, we use functions that serve as smooth approximations of the absolute value function, as in Definition 6. Then, for a given $(\rho, \sigma, \tau, \theta) \in \mathbb{R}^s_{++} \times \mathbb{R}^m_{++} \times \mathbb{R}_{++} \times \mathbb{R}_{++}$, the *smoothing ℓ_1-exact penalty Lagrangian function* associated with Problem (1), denoted by $\mathbb{L}_{(\tau,\theta)}(\cdot, \rho, \sigma) \colon \mathcal{M} \to \mathbb{R}$ is defined as follows:

$$\mathbb{L}_{(\tau,\theta)}(p, \rho, \sigma) := f(p) + \sum_{i=1}^{s} \rho_i \varphi_\tau(h_i(p)) + \sum_{j=1}^{m} \sigma_j \left(g_j(p) + \psi_\theta(g_j(p)) \right), \tag{9}$$

where φ_τ *and* ψ_θ *are smooth absolute value functions satisfying Definition 6.* At this point, it is important to note that we could assign in (9) distinct smooth absolute value functions to each function h_i and g_j individually, resulting in a total of $s + m$ smooth absolute value functions. However, for the sake of simplifying both the notation and analysis, we will instead use a single function φ_τ for all $h_{i's}$ and a single function ψ_θ for all $g_{j's}$. Since φ_τ and ψ_θ are nonnegative and increasing functions for positive parameters, the function $\mathbb{L}_{(\tau,\theta)}$ acts as a penalty for the constraint h_i and g_j violated. Since the functions f, h and g are continuously differentiable on $\mathcal{M}$, φ_τ and ψ_θ in Definition 6 are continuously differentiable, we conclude that $\mathbb{L}_{(\tau,\theta)}(\cdot, \rho, \sigma)$ is also continuously differentiable. It is worth noting that the terms inside the sums in (9) approximate the ℓ_1-exact penalty terms $p \mapsto \rho_i |h_i(p)|$ and $p \mapsto 2\sigma_j (g_j(p))_+$, respectively. Indeed, $\lim_{\tau \to +\infty} \varphi_\tau(h_i(p)) =$

$|h_i(p)|$ and $\lim_{\theta\to+\infty}(g_j(p)+\psi_\theta(g_j(p))) = 2(g_j(p))_+$, which were introduced in Euclidean space for the first time in [55] to define the ℓ_1-exact penalty function. For a comprehensive study on exact penalty functions in Euclidean setting, see, for example, [16]. By selecting suitable scalars ρ_i and σ_j, and smoothing functions φ_τ and ψ_θ, such as those demonstrated in Example 1, the function (9) incorporates those outlined in [33]. Note that, in the Euclidean context, if the approximation functions φ_τ and ψ_θ are the ones described in item (i) of Example 1 with $r = 2$, then (9) corresponds to the smooth ℓ_1-exact penalty function introduced and analyzed in [29]. This smooth ℓ_1-exact penalty function has also been studied in [50], where it was termed the hyperbolic penalty function. For additional developments on this topic, see also [10, 36, 37, 51, 52].

For future reference in the upcoming sections, the gradient[2] of the smoothing ℓ_1-exact penalty function $\mathbb{L}_{(\tau,\theta)}(\cdot,\rho,\sigma) : \mathcal{M} \to \mathbb{R}$ is given by

$$\operatorname{grad}\mathbb{L}_{(\tau,\theta)}(p,\rho,\sigma) = \operatorname{grad} f(p) + \sum_{i=1}^{s} \rho_i \varphi'_\tau(h_i(p)) \operatorname{grad} h_i(p) + \sum_{j=1}^{m} \sigma_j \left(1 + \psi'_\theta(g_j(p))\right) \operatorname{grad} g_j(p). \quad (10)$$

3.1 Properties of smoothing ℓ_1-exact penalty function

In this section, we present some properties of the smoothing exact penalty function (9). Note that due to item (i) of Definition 6 we have $\lim_{\tau,\theta\to+\infty}\mathbb{L}_{(\tau,\theta)}(p,\rho,\sigma) = \mathbb{E}(p,\rho,\sigma)$, where $\mathbb{E}$ is defined as the *classic ℓ_1-exact penalty Lagrangian function*

$$\mathbb{E}(p,\rho,\sigma) := f(p) + \sum_{i=1}^{s} \rho_i |h_i(p)| + \sum_{j=1}^{m} \sigma_j \left(g_j(p) + |g_j(p)|\right). \quad (11)$$

Therefore, we conclude that $\mathbb{L}_{(\tau,\theta)}$ smooths to $\mathbb{E}$. The following lemma presents a straightforward yet valuable property that establishes a relationship between the penalty function given in (9) and (11), for the special case when φ_τ and ψ_θ satisfy (8). To state the next lemma we set

$$\rho := (\rho_1,\ldots,\rho_s) \in \mathbb{R}^s_{++}, \qquad \tilde{\xi} := \sum_{i=1}^{s} \rho_i, \qquad \sigma := (\sigma_1,\ldots,\sigma_m) \in \mathbb{R}^m_{++}, \qquad \hat{\xi} := \sum_{j=1}^{m} \sigma_j$$

Lemma 3. *Assume that φ_τ and ψ_θ satisfy* (8)*. Then, the following inequalities hold*

$$\mathbb{E}(p,\rho,\sigma) - (\tilde{\xi}/\tau + \hat{\xi}/\theta)\kappa \le \mathbb{L}_{(\tau,\theta)}(p,\rho,\sigma) \le \mathbb{E}(p,\rho,\sigma), \quad (12)$$

for all $(\rho,\sigma,\tau,\theta) \in \mathbb{R}^s_{++} \times \mathbb{R}^m_{++} \times \mathbb{R}_{++} \times \mathbb{R}_{++}$.

Proof. By combining (9) and (11) with (8), we have

$$0 \le \mathbb{E}(p,\rho,\sigma) - \mathbb{L}_{(\tau,\theta)}(p,\rho,\sigma) = \sum_{i=1}^{s} \rho_i \big(|h_i(p)| - \varphi_\tau(h_i(p))\big) + \sum_{j=1}^{m} \sigma_j \big(|g_j(p)| - \psi_\theta(g_j(p))\big) \le \Big(\frac{\tilde{\xi}}{\tau} + \frac{\hat{\xi}}{\theta}\Big)\kappa,$$

for all $(p,\rho,\sigma,\tau,\theta) \in \mathcal{M} \times \mathbb{R}^{s+m+2}_{++}$, which implies the desired inequalities. □

To continue our discussion, for a given $(\rho,\sigma,\tau,\theta) \in \mathbb{R}^s_{++} \times \mathbb{R}^m_{++} \times \mathbb{R}_{++} \times \mathbb{R}_{++}$, consider the smoothed optimization problem:

$$\underset{p\in\mathcal{M}}{\text{Minimize}}\, \mathbb{L}_{(\tau,\theta)}(p,\rho,\sigma). \quad (13)$$

[2] Although the Lagrangian $\mathbb{L}_{(\tau,\theta)}$ is a function of three variables, to simplify the notation, we denote by $\operatorname{grad}\mathbb{L}$ the gradient with respect to the first variable.

In Euclidean space, it is known that for appropriately chosen penalty parameters ρ and σ, the solutions of the non-differentiable ℓ_1-exact penalized problem also serve as solutions to the original problem (see, for example, [7, Chapter 4]). Consequently, our next task is to investigate this matter for the smoothed ℓ_1-exact penalty optimization problem in (13). Specifically, we will study the conditions under which the solutions of Problem (13) are approximated solutions for Problem (1), for the special case when φ_τ and ψ_θ satisfy (8).

Proposition 4. *Assume that φ_τ and ψ_θ satisfy* (8)*. In addition, assume that $\bar{p} \in \mathcal{M}$ is a local solution to optimization Problem* (13)*. If the point $\bar{p}$ is feasible for Problem* (1)*, i.e., $\bar{p} \in \Omega$, then $\bar{p}$ is a local $((\tilde{\xi}/\tau + \hat{\xi}/\theta)\kappa)$-solution to Problem* (1)*, meaning that for all $p \in \Omega \cap B_r(\bar{p})$ and for some $r > 0$, the inequality $f(\bar{p}) \leq f(p) + ((\tilde{\xi}/\tau + \hat{\xi}/\theta)\kappa)$ holds.*

Proof. Assume that $\bar{p} \in \mathcal{M}$ is a local solution to Problem (13). Hence, there exists $r > 0$ such that $\mathbb{L}_{(\tau,\theta)}(\bar{p}, \rho, \sigma) \leq \mathbb{L}_{(\tau,\theta)}(p, \rho, \sigma)$, for all $p \in B_r(\bar{p})$. Thus, since the first inequality in (8) implies that $\varphi_\tau(h_i(p)) = 0$ and $g_j(p) + \psi_\theta(g_j(p)) \leq 0$, for all $p \in \Omega$, we obtain that

$$f(\bar{p}) + \sum_{i=1}^{s} \rho_i \varphi_\tau(h_i(\bar{p})) + \sum_{i=1}^{m} \sigma_j \left(g_j(\bar{p}) + \psi_\theta(g_j(\bar{p}))\right) \leq f(p), \quad \forall p \in \Omega \cap B_r(\bar{p}).$$

Due to $\bar{p} \in \Omega$, we have $|h_i(\bar{p})| = 0$ and $g_j(\bar{p}) + |g_j(\bar{p})| = 0$. Thus, the last inequality yelds

$$f(\bar{p}) + \sum_{i=1}^{s} \left(\rho_i \varphi_\tau(h_i(\bar{p})) - \rho_i |h_i(\bar{p})|\right) + \sum_{j=1}^{m} \left(\sigma_j \psi_\theta(g_j(\bar{p})) - \sigma_j |g_j(\bar{p})|\right) \leq f(p), \quad \forall p \in \Omega \cap B_r(\bar{p}).$$

Therefore, since φ_τ and ψ_θ satisfy (8), we conclude that $f(\bar{p}) - ((\xi/\tau + m/\theta)\kappa) \leq f(p)$, for all $p \in \Omega \cap B_r(\bar{p})$, which concludes the proof. □

In the following proposition, we investigate a type of converse of Proposition 4, focusing specifically on the convex case.

Proposition 5. *Assume that φ_τ and ψ_θ satisfy* (8)*. Moreover, assume that $f\colon \mathcal{M} \to \mathbb{R}$ and $g_j\colon \mathcal{M} \to \mathbb{R}$, for all $j = 1, \ldots, m$, are convex functions and $h_i\colon \mathcal{M} \to \mathbb{R}$, for all $i = 1, \ldots, s$, are both convex and concave functions. In addition, assume $\bar{p} \in \mathcal{M}$, $\bar{\lambda} = (\bar{\lambda}_1, \ldots, \bar{\lambda}_s) \in \mathbb{R}^s$ and $\bar{\mu} = (\bar{\mu}_1, \ldots, \bar{\mu}_m) \in \mathbb{R}^m$ satisfy KKT conditions for Problem* (1)*. Then, for all $\rho \in \mathbb{R}^s_{++}$ and $\sigma \in \mathbb{R}^m_{++}$ such that $\rho_i \geq |\bar{\lambda}_i|$, for all $i = 1, \ldots, s$, and $\sigma_j \geq \bar{\mu}_j$, for all $j = 1, \ldots, m$, the point $\bar{p}$ is a global $((\tilde{\xi}/\tau + \hat{\xi}/\theta)\kappa)$-solution to Problem* (13)*, i.e., $\mathbb{L}_{(\tau,\theta)}(\bar{p}, \rho, \sigma) \leq \mathbb{L}_{(\tau,\theta)}(p, \rho, \sigma) + ((\tilde{\xi}/\tau + \hat{\xi}/\theta)\kappa)$.*

Proof. Let $\eta_{(\bar{\lambda},\bar{\mu})}\colon \mathcal{M} \to \mathbb{R}$ be defined by $\eta_{(\bar{\lambda},\bar{\mu})}(p) := L(p, \bar{\lambda}, \bar{\mu})$. Since $\bar{\lambda} \in \mathbb{R}^s$ and $\bar{\mu} \in \mathbb{R}^m_+$ and $f\colon \mathcal{M} \to \mathbb{R}$ and $g_j\colon \mathcal{M} \to \mathbb{R}$, for all $j = 1, \ldots, m$, are convex functions and $h_i\colon \mathcal{M} \to \mathbb{R}$, for all $i = 1, \ldots, s$, are both convex and concave functions, it follows from (4) that the function $\eta_{(\bar{\lambda},\bar{\mu})}$ is also convex and differentiable. Moreover, due to $\bar{p}$, $\bar{\lambda}$ and $\bar{\mu}$ satisfying KKT conditions for Problem (1), we have $\operatorname{grad} f(\bar{p}) + \sum_{i=1}^{s} \bar{\lambda}_i \operatorname{grad} h_i(\bar{p}) + \sum_{j=1}^{m} \bar{\mu}_j \operatorname{grad} g_j(\bar{p}) = 0$. Thus, $\operatorname{grad} \eta_{(\bar{\lambda},\bar{\mu})}(\bar{p}) = 0$, which due to $\eta_{(\bar{\lambda},\bar{\mu})}$ being convex implies that $\bar{p}$ is a global minimizer to $\eta_{(\bar{\lambda},\bar{\mu})}$. Therefore, we conclude that

$$L(\bar{p}, \bar{\lambda}, \bar{\mu}) \leq L(p, \bar{\lambda}, \bar{\mu}), \qquad \forall p \in \mathcal{M}. \tag{14}$$

Since φ_τ and ψ_θ satisfy (8) and $\bar{p} \in \Omega$, it follows from (9), by taking into account that $\bar{\mu} := (\bar{\mu}_1, \ldots, \bar{\mu}_m) \in \mathbb{R}^m_+$and $\bar{\mu}_j g_j(\bar{p}) = 0$, for all $j \in \{1, \ldots, m\}$, that

$$\mathbb{L}_{(\tau,\theta)}(\bar{p}, \rho, \sigma) = f(\bar{p}) + \sum_{i=1}^{s} \rho_i \varphi_\tau(h(\bar{p})) + \sum_{j=1}^{m} \sigma_j \left(g_j(\bar{p}) + \psi_\theta(g_j(\bar{p}))\right) \leq f(\bar{p}) = L(\bar{p}, \bar{\lambda}, \bar{\mu}).$$

Hence, taking into account (14), we obtain that

$$\mathbb{L}_{(\tau,\theta)}(\bar{p},\rho,\sigma) \le L(p,\bar{\lambda},\bar{\mu}), \qquad \forall p \in \mathcal{M}. \tag{15}$$

On the other hand, for any $\rho \in \mathbb{R}^s_{++}$ and $\sigma \in \mathbb{R}^m_{++}$ such that $\rho_i \ge |\bar{\lambda}_i|$, for all $i = 1,\dots,s$, and $\sigma_j \ge \bar{\mu}_j$, for all $j = 1,\dots,m$, we have $\bar{\lambda}_i h_i(p) \le \rho_i |h_i(p)|$ and $\bar{\mu}_i(g_j(p)+|g_j(p)|) \le \sigma_i(g_j(p)+|g_j(p)|)$. Hence, using (15) we conclude that

$$\mathbb{L}_{(\tau,\theta)}(\bar{p},\rho,\sigma) \le L(p,\bar{\lambda},\bar{\mu}) = f(p) + \sum_{i=1}^{s} \bar{\lambda}_i h_i(p) + \sum_{j=1}^{m} \bar{\mu}_j g_j(p) \le \mathbb{E}(p,\rho,\sigma), \qquad \forall p \in \mathcal{M}.$$

Therefore, the first inequality in (12) implies the desired inequality, which concludes the proof. □

In the next proposition, we demonstrate that $\mathbb{E}$ is indeed exact for the Problem (1). To this end, for a given $\rho \in \mathbb{R}^s$ and $\sigma \in \mathbb{R}^m_{++}$, define the following optimization problem

$$\underset{p\in\mathcal{M}}{\text{Minimize}}\, \mathbb{E}(p,\rho,\sigma) \tag{16}$$

Proposition 6. *Assume that $\bar{p} \in \mathcal{M}$ is a local solution to Problem* (16)*. If the point $\bar{p}$ is feasible for Problem* (1)*, i.e., $\bar{p} \in \Omega$, then the following two statements hold:*

(i) $\bar{p}$ is a local solution to Problem (1)*;*

(ii) $\bar{p}$ is a KKT point, i.e., there exist $\bar{\lambda} \in \mathbb{R}^s$ and $\bar{\mu} \in \mathbb{R}^m_+$ such that $\bar{p}$, $\bar{\lambda}$ and $\bar{\mu}$ satisfy KKT conditions.

Proof. To prove item (i), assume that $\bar{p} \in \mathcal{M}$ is a local solution to Problem (16). Hence, there exists $r > 0$ such that $\mathbb{E}(\bar{p},\rho,\sigma) \le \mathbb{E}(p,\rho,\sigma)$, for all $p \in B_r(\bar{p})$. Since $\bar{p} \in \Omega$ and $\mathbb{E}(p,\rho,\sigma) = f(p)$, for all $p \in \Omega$, we conclude that $f(\bar{p}) \le f(p)$, for all $p \in B_r(\bar{p}) \cap \Omega$, which proves item (i). We proceed to prove that item (ii) holds. For that, define $\psi_{\rho,\sigma}(p) := \mathbb{E}(p,\rho,\sigma)$. Since $\bar{p} \in \mathcal{M}$ is a local solution to Problem (16), we have $0 \le \psi'_{\rho,\sigma}(\bar{p},d)$, for all $d \in T_p\mathcal{M}$. Thus, taking into account that $\bar{p} \in \Omega$, we have

$$0 \le \langle \operatorname{grad} f(\bar{p}), d\rangle + \sum_{i=1}^{s} \rho_i |\langle \operatorname{grad} h_i(\bar{p}), d\rangle| + \sum_{j\in\mathcal{A}(\bar{p})} \sigma_j \langle \operatorname{grad} g_j(\bar{p}), d\rangle + \sum_{j\in\mathcal{A}(\bar{p})} \sigma_j |\langle \operatorname{grad} g_j(\bar{p}), d\rangle|,$$

for all $d \in T_{\bar{p}}\mathcal{M}$. Hence, considering that $\rho \in \mathbb{R}^s_{++}$ and $\sigma \in \mathbb{R}^m_{++}$, the last inequality implies that there is no $d \in T_{\bar{p}}\mathcal{M}$ satisfying

$$\langle \operatorname{grad} f(\bar{p}), d\rangle < 0, \qquad \langle \operatorname{grad} h_i(\bar{p}), d\rangle = 0, \quad \forall i \in \{i,\dots,s\}, \qquad \langle \operatorname{grad} g_j(\bar{p}), d\rangle \le 0, \quad \forall j \in \mathcal{A}(\bar{p}).$$

Therefore, by applying Lemma 1 in $T_{\bar{p}}\mathcal{M}$, we conclude that there exist $\bar{\lambda}_i$, for $i = i,\dots,s$, and $\bar{\mu}_j \ge 0$, for all $i \in \mathcal{A}(\bar{p})$, such that

$$\operatorname{grad} f(\bar{p}) + \sum_{i=1}^{s} \bar{\lambda}_i \operatorname{grad} h_i(\bar{p}) + \sum_{j\in\mathcal{A}(\bar{p})} \bar{\mu}_j \operatorname{grad} g_j(\bar{p}) = 0. \tag{17}$$

Taking into account that $\bar{p} \in \Omega$, defining $\bar{\mu}_j = 0$, for all $j \in \{1,\dots,m\}\backslash\mathcal{A}(\bar{p})$, we have $\bar{\mu} \in \mathbb{R}^m_+$ and $\bar{\mu}_j g_j(\bar{p}) = 0$, for all $j \in \{1,\dots,m\}$, which together with (17) imply that $\bar{p}$, $\bar{\lambda}$ and $\bar{\mu}$ satisfy KKT conditions and item (ii) is also proved. □

In the following proposition, we present a counterpart to Proposition 6 for Problem (16).

Proposition 7. *Assume that $f\colon \mathcal{M} \to \mathbb{R}$ and $g_j\colon \mathcal{M} \to \mathbb{R}$, for all $j = 1,\dots,m$, are convex functions and $h_i\colon \mathcal{M} \to \mathbb{R}$, for all $i = 1,\dots,s$, are both convex and concave functions. In addition, assume that $\bar{p} \in \mathcal{M}$, $\bar{\lambda} = (\bar{\lambda}_1,\dots,\bar{\lambda}_s) \in \mathbb{R}^n$ and $\bar{\mu} = (\bar{\mu}_1,\dots,\bar{\mu}_m) \in \mathbb{R}^m_+$ satisfy KKT conditions for Problem* (1)*. Then, for all $\rho \in \mathbb{R}^s_{++}$ and $\sigma \in \mathbb{R}^m_{++}$ such that $\rho_i \geq |\bar{\lambda}_i|$, for all $i = 1,\dots,s$, and $\sigma_j \geq \bar{\mu}_j$, for all $j = 1,\dots,m$, the point $\bar{p}$ is a global solution to Problem* (16).

Proof. Let $\eta_{(\bar{\lambda},\bar{\mu})} : \mathcal{M} \to \mathbb{R}$ be defined by $\eta_{(\bar{\lambda},\bar{\mu})}(p) := L(p,\bar{\lambda},\bar{\mu})$. Since $\bar{\mu} \in \mathbb{R}^m_+$ and the functions f and g_j, for all $i = 1,\dots,m$, are convex and differentiable and $h_i\colon \mathcal{M} \to \mathbb{R}$, for all $i = 1,\dots,s$, are both convex and concave functions and also differentiable, it follows from (4) that the function $\eta_{(\bar{\lambda},\bar{\mu})}$ is also convex and differentiable. Moreover, due to $\bar{p}$, $\bar{\lambda}$ and $\bar{\mu}$ satisfying KKT conditions for Problem (1), we have

$$\operatorname{grad} f(\bar{p}) + \sum_{j=1}^{s} \bar{\lambda}_i \operatorname{grad} h_i(\bar{p}) + \sum_{j=1}^{m} \bar{\mu}_j \operatorname{grad} g_j(\bar{p}) = 0.$$

Thus, we conclude that $\operatorname{grad} \eta_{(\bar{\lambda},\bar{\mu})}(\bar{p}) = 0$, which implies that $\bar{p}$ is a global minimizer to $\eta_{(\bar{\lambda},\bar{\mu})}$. Hence, we obtain that $L(\bar{p},\bar{\lambda},\bar{\mu}) \leq L(p,\bar{\lambda},\bar{\mu})$, for all $p \in \mathcal{M}$. Due to $\bar{p}$, $\bar{\lambda}$ and $\bar{\mu}$ satisfying KKT conditions for Problem (1), we have $h_i(p) = 0$, for all $i \in \{1,\dots,s\}$ and $\bar{\mu}_j g_j(\bar{p}) = 0$, for all $j \in \{1,\dots,m\}$. Therefore, we conclude that

$$\mathbb{E}(\bar{p},\bar{\lambda},\bar{\mu}) = f(\bar{p}) = f(\bar{p}) + \sum_{i=1}^{s} \bar{\lambda}_i h_i(\bar{p}) + \sum_{j=1}^{m} \bar{\mu}_j g_j(\bar{p}) = L(\bar{p},\bar{\lambda},\bar{\mu}) \leq L(p,\bar{\lambda},\bar{\mu}),$$

for all $p \in \mathcal{M}$. On the other hand, taking into account that $\rho_i \geq |\bar{\lambda}_i|$, for all $i = 1,\dots,s$, and $\sigma_j \geq \bar{\mu}_j$, for all $j = 1,\dots,m$, $L(p,\bar{\lambda},\bar{\mu}) \leq f(p) + \sum_{i=1}^{s} \rho_i |h_i(p)| + \sum_{j=1}^{m} \sigma_j \max\{0, g_j(p)\} = \mathbb{E}(p,\rho,\sigma)$, for all $p \in \mathcal{M}$. Therefore, combining two previous inequalities, we conclude that $\mathbb{E}(\bar{p},\bar{\lambda},\bar{\mu}) \leq \mathbb{E}(p,\rho,\sigma)$, for all $p \in \mathcal{M}$, and the proof is concluded. □

Propositions 6 and 7 suggest that instead of solving Problem (1), we can alternatively solve Problem (16) for sufficiently large $\rho_{i's} > 0$ and $\sigma_{j's} > 0$, as these multipliers penalize the constraints $h_{i's}$ and $g_{j's}$. In the Euclidean case, it is well-established that only finite penalty weights $\rho_{i's} > 0$ and $\sigma_{j's} > 0$ are needed to achieve exact satisfaction of constraints (see [7, Chapter 4]). A similar property for the Riemannian case, under a second-order condition, is demonstrated in [33]. Additionally, Problem (1) can alternatively be solved by addressing the smoothed version of Problem (16), namely Problem (13). In this context, differentiable optimization methods can be employed. The following section introduces a smooth penalization algorithm to address this problem.

The following theorem provides a version of a classical result in exact penalty methods, specifically applied to the smoothed ℓ_1-exact penalty optimization problem in (13). It demonstrates that, under appropriate conditions on the penalty and smoothing parameters, any cluster point of a sequence of global solutions to the penalized problem is a global solution to the original problem. The exact conditions for this convergence are detailed below.

Theorem 8. *Assume that φ_τ and ψ_θ satisfy* (8) *and that $-\infty < \bar{c} = \inf\{f(p) : \ p \in \Omega\}$. Take sequences $(\rho^k)_{k\in\mathbb{N}} \subset \mathbb{R}^s_{++}$ and $(\sigma^k)_{k\in\mathbb{N}} \subset \mathbb{R}^m_{++}$, such that $\lim_{k\to+\infty} \rho^k_i = +\infty$, for all $i = 1,\dots,s$ and $\lim_{k\to+\infty} \sigma^k_j = +\infty$, for all $j = 1,\dots,m$, and strictly increasing sequences $(\tau_k)_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ and*

$(\theta_k)_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ *such that* $\lim_{k\to+\infty}\tau_k = +\infty$ *and* $\lim_{k\to+\infty}\theta_k = +\infty$*. Let* p^k *be a global solution of the penalty problem*

$$\underset{p\in\mathcal{M}}{\text{Minimize}}\, \mathbb{L}_{(\tau_k,\theta_k)}(p,\rho^k,\sigma^k). \tag{18}$$

If $\lim_{k\to+\infty}\sigma_j^k/\theta_k = 0$*, for all* $j = 1,\dots,m$*, then all cluster points of the sequence* $(p^k)_{k\in\mathbb{N}}$ *are global solutions of Problem* (1)*.*

Proof. Let $(p^k)_{k\in\mathbb{N}}$ be the sequence of global minimizers of the penalized problem (18), i.e.,

$$p^k \in \arg\min_{p\in\mathcal{M}} \mathbb{L}_{(\tau_k,\theta_k)}(p,\rho^k,\sigma^k).$$

Let $\bar{p}$ be any cluster point of $(p^k)_{k\in\mathbb{N}}$ and $(p^{k_\ell})_{\ell\in\mathbb{N}}$ be a subsequence such that $\lim_{\ell\to+\infty} p^{k_\ell} = \bar{p}$. Out task is to prove that $\bar{p}\in\Omega$ and that $f(\bar{p}) = \bar{c}$. We will first show that $\bar{p}\in\Omega$. Since p^k is global solution of (18), we have

$$\mathbb{L}_{(\tau_k,\theta_k)}(p^k,\rho^k,\sigma^k) \leq \mathbb{L}_{(\tau_k,\theta_k)}(p,\rho^k,\sigma^k), \qquad \forall p\in\mathcal{M}. \tag{19}$$

Let $p\in\Omega$ be any feasible point. For such p, due to φ_τ and ψ_θ satisfy (8), we obtain that $\varphi_{\tau_k}(h_i(p)) = 0$ and $g_j(p) + \psi_{\theta_k}(g_j(p)) \leq 0$. Thus, the penalty term is non-positive and does not increase the objective value, which implies that $\mathbb{L}_{(\tau_k,\theta_k)}(p,\rho^k,\sigma^k) \leq f(p)$, for all $p\in\Omega$. Hence, it follows from (19) that $\mathbb{L}_{(\tau_k,\theta_k)}(p^k,\rho^k,\sigma^k) \leq f(p)$, for all $p\in\Omega$. Taking the infimum over all feasible $p\in\Omega$, we conclude that

$$\mathbb{L}_{(\tau_k,\theta_k)}(p^k,\rho^k,\sigma^k) \leq \bar{c}. \tag{20}$$

On the other hand, considering that ψ_θ satisfy (8) and $\rho^k\in\mathbb{R}^s_{++}$ and $\sigma_k\in\mathbb{R}^m_{++}$, we conclude that

$$f(p^k) + \sum_{i=1}^{s}\rho_i^k\,\varphi_{\tau_k}(h_i(p^k)) + \sum_{j=1}^{m}\sigma_j^k\left(g_j(p^k) + |g_j(p^k)|\right) + \kappa\sum_{j=1}^{m}\frac{\sigma_j^k}{\theta_k} \leq \bar{c}. \tag{21}$$

Considering that $\rho_i^k\,\varphi_{\tau_k}(h_i(p^k)) \geq 0$ and $\sigma_j^k\left(g_j(p^k) + |g_j(p^k)|\right) \geq 0$, (20) with (21) imply that

$$f(p^k) + \kappa\sum_{j=1}^{m}\frac{\sigma_j^k}{\theta_k} \leq f(p^k) + \sum_{i=1}^{s}\rho_i^k\,\varphi_{\tau_k}(h_i(p^k)) + \sum_{j=1}^{m}\sigma_j^k\left(g_j(p^k) + |g_j(p^k)|\right) \leq \bar{c}. \tag{22}$$

Moreover, since $\lim_{k\to+\infty}\sigma_i^k/\theta_k = 0$, for all $i = 1,\dots,m$, and $\lim_{\ell\to+\infty}p^{k_\ell} = \bar{p}$, it follows from (22) that $\lim_{\ell\to+\infty}\rho_i^{k_\ell}\,\varphi_{\tau_{k_\ell}}(h_i(p^{k_\ell})) = 0$ and $\lim_{\ell\to+\infty}\sigma_j^{k_\ell}(g_j(p^{k_\ell}) + |g_j(p^{k_\ell})|) = 0$, for all $i = 1,\dots,s$ and for all $j = 1,\dots,m$. Given that $\lim_{k\to+\infty}\rho_i^k = +\infty$ and $\lim_{k\to+\infty}\sigma_j^k = +\infty$, we conclude that $\lim_{\ell\to+\infty}\varphi_{\tau_{k_\ell}}(h_i(p^{k_\ell})) = 0$ and $\lim_{\ell\to+\infty}(g_j(p^{k_\ell}) + |g_j(p^{k_\ell})|) = 0$, for all $i = 1,\dots,s$ and for all $j = 1,\dots,m$. Therefore, due to h_i being continuous and $\lim_{\ell\to+\infty}p^{k_\ell} = \bar{p}$, by using item (i) of Definition 6 we conclude that $h_i(\bar{p}) = 0$ for all $i = 1,\dots,s$. In addition, considering that g_j is continuous, we have $g_j(\bar{p}) + |g_j(\bar{p})| = 0$, or equivalently, $g_j(\bar{p}) \leq 0$, for all $j = 1,\dots,m$. Therefore, $\bar{p}\in\Omega$. Thus, $\bar{p}$ is a feasible point of Problem (1).

We proceed to prove the optimality of $\bar{p}$. For that, we first note that (22) gives

$$f(p^k) + \kappa\sum_{j=1}^{m}\frac{\sigma_j^k}{\theta_k} \leq \bar{c}.$$

Given that $\lim_{k\to+\infty}\sigma_i^k/\theta_k = 0$ and $\lim_{\ell\to+\infty}p^{k_\ell} = \bar{p}$ along with the continuity of f we obtain that $f(\bar{p}) = \lim_{\ell\to+\infty}f(p^{k_\ell}) \leq \bar{c}$. But since $\bar{p}$ is feasible, we have $f(\bar{p}) \geq \bar{c}$, because $\bar{c}$ is the infimum of f over all feasible points. Thus, $f(\bar{p}) = \bar{c}$. Therefore, $\bar{p}$ is a global minimizer of f over the feasible set Ω. □

4 Smooth ℓ_1-exact penalty algorithm

This section introduces a smooth ℓ_1-exact penalty algorithm designed to solve Problem (1). It is based on an adaptation of the method in [33, Algorithm 2], modified to incorporate general smoothing functions as defined in Definition 6. The algorithm adjusts both smoothing and penalty parameters to handle nonlinear constraints effectively. By iteratively refining these parameters, it penalizes constraint violations while updating the smoothing parameters. Under EMFCQ, if the algorithm converges, the limit point must satisfy the KKT conditions for Problem (1). Additionally, it is shown that, assuming feasibility of the limit point, the sequence converges to an AKKT point, which is also a KKT point for Problem (1) under the weak constraint qualifications RCPLD or CRSC. If the penalty parameters $(\rho^k)_{k\in\mathbb{N}}$ and $(\sigma^k)_{k\in\mathbb{N}}$ are bounded and the smoothing parameter sequences $(\tau_k)_{k\in\mathbb{N}}$ and $(\theta_k)_{k\in\mathbb{N}}$ grow unbounded, any cluster point satisfies the KKT conditions. The smooth ℓ_1-exact penalty algorithm can be conceptually outlined as follows:

Algorithm 1 : Smooth ℓ_1-exact penalty algorithm (Sℓ_1-EPA)

Step 0. (Initialization) Take forcing parameters $\tilde{\beta}, \hat{\beta} \in (0,1)$ and constants $\tilde{\nu}, \hat{\nu} > 1$. Take $\rho^1 \in \mathbb{R}^s_{++}$ and $\sigma^1 \in \mathbb{R}^m_{++}$ initial penalty vectors estimate and an initial point $p^0 \in \mathcal{M}$. Choose smoothing parameter sequences $(\tau_k)_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ and $(\theta_k)_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ and a sequence of tolerance parameters $(\epsilon_k)_{k\in\mathbb{N}} \subset \mathbb{R}_+$ such that $\lim_{k\to+\infty} \epsilon_k = 0$. Set $k \leftarrow 1$.

Step 1. (Solve the subproblem) Compute (if possible) $p^k \in \mathcal{M}$ such that

$$\left\|\operatorname{grad} \mathbb{L}_{(\tau_k,\theta_k)}(p^k, \rho^k, \sigma^k)\right\| \leq \epsilon_k. \tag{23}$$

If it is not possible, stop the execution of the algorithm, declaring failure;

Step 2. (Updating penalty parameters) For $i = 1, \ldots, s$, if

$$|h_i(p^k)| \leq \tilde{\beta}|h_i(p^{k-1})|, \tag{24}$$

then set $\rho_i^{k+1} = \rho_i^k$, otherwise, set $\rho_i^{k+1} = \tilde{\nu}\rho_i^k$, and for $j = 1, \ldots, m$, if

$$g_j(p^k)_+ \leq \hat{\beta} g_j(p^{k-1})_+, \tag{25}$$

then set $\sigma_j^{k+1} = \sigma_j^k$, otherwise, set $\sigma_j^{k+1} = \hat{\nu}\sigma_j^k$;

Step 3. (Begin a new iteration) Set $k \leftarrow k+1$ and go to **Step 1**.

Before proceeding with the analysis of Algorithm 1, we first highlight its main aspects. One fundamental aspect to address from the outset is the choice of the smoothing functions φ_τ and ψ_θ used in the definition of the smoothed exact penalty Lagrangian function, as indicated in (9). It is noteworthy that each choice of the functions φ_τ and ψ_θ leads to distinct algorithms. It is worth noting that, due to $\varphi_\tau(t) > 0$ and $\psi_\theta(t) > 0$, for $t \neq 0$, (see item (ii) of Definition 6), it follows from (9) that ρ_i and σ_j serve as a penalty parameter when the constraints h_i and g_j are violated, respectively. In **Step 1**, to compute p^k satisfying (23), any unconstrained optimization algorithm can be employed to approximately solve the following subproblem

$$\underset{p\in\mathcal{M}}{\text{Minimize}}\, \mathbb{L}_{(\tau_k,\theta_k)}(p, \rho^k, \sigma^k), \tag{26}$$

which could involve first-order or second-order algorithms, accompanied by a stopping criterion that satisfies (23). Various algorithms addressing this subproblem have been proposed, such as those discussed in [1, 17, 22, 23, 25, 46]. In **Step 2**, we update the penalty multipliers ρ_i^k and σ_j^k based on the feasibility measure (24) and (25) for the violated constraints h_i and g_j, respectively.

This selective updating approach contrasts with [33, Algorithm 2], which uniformly updates all penalty multipliers by assuming uniformity across all constraints h_i and g_j. On the other hand, the smoothing parameter sequences $(\tau_k)_{k\in\mathbb{N}} \subset \mathbb{R}{++}$ and $(\theta_k)_{k\in\mathbb{N}} \subset \mathbb{R}{++}$ are chosen exogenously. However, if both sequences $(\tau_k)_{k\in\mathbb{N}}$ and $(\theta_k)_{k\in\mathbb{N}}$ are selected to be increasing, then the smooth ℓ_1-exact penalty function $\mathbb{L}_{(\tau,\theta)}$ progressively converges toward the ℓ_1-exact penalty function $\mathbb{E}$ with each iteration. For further insights and discussions on penalty methods in the Euclidean setting, refer to [7, 39]. It is important to note that the absence of a stopping criterion in Algorithm 1 permits it to generate an infinite sequence. Therefore, practical implementation requires an appropriate stopping criterion, which will be thoroughly addressed in the implementation section. Let $(p^k)_{k\in\mathbb{N}}$ be a sequence generated by Algorithm 1, which we assume to be infinite and, $(\rho^k)_{k\in\mathbb{N}}$ and $(\sigma^k)_{k\in\mathbb{N}}$ denote the associated penalty sequences.

In the following sections, the behavior of the sequence $(p^k)_{k\in\mathbb{N}}$ is analyzed. For simplicity, the gradient of the smoothed ℓ_1-exact penalty function, as defined in equation (10), is used to introduce two sequences of associated Lagrange multipliers:

$$\lambda_i^k := \rho_i^k \varphi'_{\tau_k}\big(h_i(p^k)\big), \qquad \mu_j^k := \sigma_j^k\left(1 + \psi'_{\theta_k}\big(g_j(p^k)\big)\right), \qquad \forall k \in \mathbb{N}, \tag{27}$$

for each $i \in \{1, \ldots, s\}$ and $j \in \{1, \ldots, m\}$. Hence, by combining (5), (10) and (27), we obtain

$$\operatorname{grad} L(p^k, \lambda^k, \mu^k) = \operatorname{grad} \mathbb{L}_{(\tau_k,\theta_k)}(p^k, \rho^k, \sigma^k), \tag{28}$$

which establishes a fundamental equality essential for the analysis of the sequence $(p^k)_{k\in\mathbb{N}}$.

4.1 Optimality guarantees of Sℓ_1-EPA assuming convergence

In this section, we examine the optimality properties of the limit points of the sequence $(p^k)_{k\in\mathbb{N}}$. To this end, we assume the following condition throughout the section:

(A1) The smoothing parameter sequences $(\tau_k)_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ and $(\theta_k)_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ are chosen such that $\lim_{k\to\infty} \tau_k = +\infty$ and $\lim_{k\to\infty} \theta_k = +\infty$.

It is important to note that for large enough k, condition (A1) guarantees that subproblem (26) closely approximates the ℓ_1-exact penalized problem (16). This condition helps control the penalty parameters, preventing them from growing too large while ensuring that conditions (24) and (25) are satisfied. Specifically, we show that *if condition (A1) holds and $(p^k)_{k\in\mathbb{N}}$ converges, its limit satisfies the KKT conditions for Problem* (1). Moreover, under EMFCQ, the limit of $(p^k)_{k\in\mathbb{N}}$ satisfies the optimality conditions for Problem (1). We begin by establishing the necessary properties of the Lagrange multipliers.

Lemma 9. *Assume that $\lim_{k\to+\infty} p^k = \bar{p}$ and (A1) holds. Then, the following statements are true:*

(i) $(\rho^k)_{k\in\mathbb{N}} \subset \mathbb{R}^s_{++}$ *and* $(\sigma^k)_{k\in\mathbb{N}} \subset \mathbb{R}^m_{++}$.

(ii) If $g_j(\bar{p}) < 0$ for some $j \in \{1, \ldots, m\}$, then $\lim_{k\to+\infty} \mu_j^k = 0$.

(iii) If $h_i(\bar{p}) \neq 0$ for some $i \in \{1, \ldots, s\}$, then the sequence $(\lambda_i^k)_{k\in\mathbb{N}}$ is unbounded.

(iv) If $g_j(\bar{p}) > 0$ for some $j \in \{1, \ldots, m\}$, then the sequence $(\mu_j^k)_{k\in\mathbb{N}}$ is unbounded.

Consequently, if both sequences $(\lambda^k)_{k\in\mathbb{N}}$ and $(\mu^k)_{k\in\mathbb{N}}$ are bounded, then $\bar{p} \in \Omega$.

Proof. Since $\rho^1 \in \mathbb{R}^s_{++}$ and $\sigma^1 \in \mathbb{R}^m_{++}$, it follows from **Step 2** that either $\rho_i^{k+1} = \rho_i^k > 0$ or $\rho_i^{k+1} = \tilde{\nu}\rho_i^k > 0$, for all $i \in \{1, \dots, s\}$ and $k \in \mathbb{N}$. Similarly, either $\sigma_j^{k+1} = \sigma_j^k > 0$ or $\sigma_j^{k+1} = \hat{\nu}\sigma_j^k > 0$, for all $j \in \{1, \dots, m\}$ and $k \in \mathbb{N}$, establishing item (i).

For item (ii), assume that for a given $j \in \{1, \dots, m\}$, it holds that $g_j(\bar{p}) < 0$. By the continuity of g_j, there exist $\delta > 0$ and $k_0 \in \mathbb{N}$ such that $g_j(p^k) < -\delta$, implying $g_j(p^k)_+ = 0$ for all $k \geq k_0$. Consequently, by **Step 2**, (25) holds and $\sigma_j^{k+1} = \sigma_j^{k_0}$ for all $k \geq k_0$. Given that $\lim_{k\to+\infty} \theta_k = +\infty$ and $\lim_{k\to+\infty} g_j(p^k) = g_j(\bar{p}) < 0$, it follows from item (viii) of Definition 6 that $\lim_{k\to+\infty} \psi'_{\theta_k}(g_j(p^k)) = -1$. Consequently, the second equality in (27) implies $\lim_{k\to+\infty} \mu_j^k = 0$, completing the proof of item (ii).

For item (iii), assume $h_i(\bar{p}) \neq 0$ for some $i \in \{1, \dots, s\}$. Given the continuity of h_i, there exist $\delta > 0$ and $k_0 \in \mathbb{N}$ such that $|h_i(p^k)| > \delta$ for all $k \geq k_0$. Since $\tilde{\beta} \in (0, 1)$ and $\lim_{k\to+\infty} p^k = \bar{p}$, the inequality in (24) cannot hold for infinitely many $k \geq k_0$. Therefore, $\rho_i^{k+1} = \tilde{\nu}\rho_i^k$ for all $k \geq k_1$ and some $k_1 > k_0$. Consequently, the sequence $(\rho^k)_{k\in\mathbb{N}}$ is unbounded. Moreover, since $h_i(\bar{p}) \neq 0$, condition $(viii)$ of Definition 6 implies that $\lim_{k\to+\infty} |\varphi'_{\tau_k}(h_i(p^k))| = 1$. Thus, by the first equality in (27) and the unboundedness of $(\rho^k)_{k\in\mathbb{N}}$, it follows that the sequence $(\lambda_i^k)_{k\in\mathbb{N}}$ is also unbounded, proving item (iii).

To prove item (iv), assume $g_j(\bar{p}) > 0$ for a given $j \in \{1, \dots, m\}$. By the continuity of g_j, there exist $\delta > 0$ and $k_0 \in \mathbb{N}$ such that $g_j(p^k) > \delta$ for all $k \geq k_0$. Consequently, $g_j(p^k)_+ > \delta$ for all $k \geq k_0$. Given that $\hat{\beta} \in (0, 1)$ and $\lim_{k\to+\infty} p^k = \bar{p}$, the inequality in (25) cannot hold for infinitely many $k \geq k_0$. Thus, $\sigma_j^{k+1} = \hat{\nu}\sigma_j^k$ for all $k \geq k_1$ and some $k_1 > k_0$. Since $\hat{\nu} > 1$, the sequence $(\sigma^k)_{k\in\mathbb{N}}$ is unbounded. Moreover, with $g_j(\bar{p}) > 0$ and condition $(viii)$ of Definition 6, it follows that $\lim_{\ell\to+\infty} \psi'_{\theta_k}(g_j(p^k)) = 1$. Therefore, the second equality in (27) implies that the sequence $(\mu_i^k)_{k\in\mathbb{N}}$ is unbounded, completing the proof of item (iv).

For the final statement, assume that $(\lambda^k)_{k\in\mathbb{N}}$ and $(\mu^k)_{k\in\mathbb{N}}$ are bounded. Then, by items (iii) and (iv), $h_i(\bar{p}) = 0$ for all $i \in \{1, \dots, s\}$ and $g_j(\bar{p}) \leq 0$ for all $j \in \{1, \dots, m\}$, implying $\bar{p} \in \Omega$. This completes the proof. □

In the following theorem we study convergence properties of $(p^k)_{k\in\mathbb{N}}$ under the EMFCQ. In this case, we establish optimality without assuming that $\bar{p} \in \Omega$.

Theorem 10. *Assume that $\lim_{k\to+\infty} p^k = \bar{p}$ and condition (A1) holds. If $\bar{p}$ satisfies EMFCQ, then the sequences $(\lambda^k)_{k\in\mathbb{N}}$ and $(\mu^k)_{k\in\mathbb{N}}$ are bounded, implying $\bar{p} \in \Omega$. Moreover, any limit points $\bar{\lambda}$ and $\bar{\mu}$ of $(\lambda^k)_{k\in\mathbb{N}}$ and $(\mu^k)_{k\in\mathbb{N}}$, respectively, are Lagrange multipliers associated with $\bar{p}$. Consequently, $\bar{p}$ is a KKT point.*

Proof. First, note that item (i) of Lemma 9, combined with items (iv) and (v) of Definition 6 and the second equality in (27), implies that $(\mu^k)_{k\in\mathbb{N}} \subset \mathbb{R}^m_+$. Given $\lim_{k\to+\infty} \epsilon_k = 0$, Step 1 of Algorithm 1 and (28) ensure that $\lim_{k\to\infty} \operatorname{grad} L(p^k, \lambda^k, \mu^k) = 0$, i.e.,

$$\lim_{k\to\infty} \Big(\operatorname{grad} f(p^k) + \sum_{i=1}^{s} \lambda_i^k \operatorname{grad} h_i(p^k) + \sum_{j=1}^{m} \mu_j^k \operatorname{grad} g_j(p^k) \Big) = 0. \tag{29}$$

Item (ii) of Lemma 9 further implies that if $g_j(\bar{p}) < 0$, then $\lim_{k\to+\infty} \mu_j^k = 0$. Thus, we have

$$\lim_{k\to\infty} \Big(\operatorname{grad} f(p^k) + \sum_{i=1}^{s} \lambda_i^k \operatorname{grad} h_i(p^k) + \sum_{j\in\mathcal{A}_+(\bar{p})} \mu_j^k \operatorname{grad} g_j(p^k) \Big) = 0. \tag{30}$$

Assume, by contradiction, that $((\lambda^k, \mu^k))_{k\in\mathbb{N}} \subset \mathbb{R}^s \times \mathbb{R}^m_+$ is unbounded. Without loss of generality, suppose $\lim_{k\to+\infty} (\lambda^k, \mu^k)/\|(\lambda^k, \mu^k)\|_1 = (\hat{\lambda}, \hat{\mu})$. By the continuity of $\operatorname{grad} f$, $\operatorname{grad} h_i$, and $\operatorname{grad} g_j$,

(30) implies

$$\sum_{i=1}^{s} \hat{\lambda}_i \operatorname{grad} h_i(\bar{p}) + \sum_{j\in\mathcal{A}_+(\bar{p})} \hat{\mu}_j \operatorname{grad} g_j(\bar{p}) = 0.$$

Thus, given $(\hat{\lambda}, \hat{\mu}) \neq 0$, $\hat{\mu} \geq 0$, and $\bar{p}$ satisfying EMFCQ, a contradiction arises. Therefore, $((\lambda^k, \mu^k))_{k\in\mathbb{N}}$ is bounded, and, by Lemma 9, $\bar{p} \in \Omega$, proving the first statement.

To prove the final statements, let $\bar{\lambda}$ and $\bar{\mu}$ be limit points of $(\lambda^k)_{k\in\mathbb{N}}$ and $(\mu^k)_{k\in\mathbb{N}}$, respectively, and consider subsequences $(\lambda^{k_\ell})_{\ell\in\mathbb{N}}$ and $(\mu^{k_\ell})_{\ell\in\mathbb{N}}$ converging to $\bar{\lambda}$ and $\bar{\mu}$. From (29), it follows that

$$\operatorname{grad} f(\bar{p}) + \sum_{i=1}^{s} \bar{\lambda}_i \operatorname{grad} h_i(\bar{p}) + \sum_{j=1}^{m} \bar{\mu}_j \operatorname{grad} g_j(\bar{p}) = 0.$$

Moreover, if $g_j(\bar{p}) < 0$, item (ii) of Lemma 9 ensures $\bar{\mu}_j = 0$. Since $\bar{p} \in \Omega$, it holds that $\bar{\mu}_j g_j(\bar{p}) = 0$ for all $j \in \{1, \ldots, m\}$. Therefore, $\bar{\lambda}$ and $\bar{\mu}$ are Lagrange multipliers associated with $\bar{p}$, concluding the proof. ☐

An important aspect of Theorem 10 is that it does not require the limit point $\bar{p}$ of the sequence $(p^k)_{k\in\mathbb{N}}$ to be feasible, nor does it assume boundedness of the Lagrange multipliers. Nevertheless, the EMFCQ condition guarantees both the boundedness of the Lagrange multiplier sequence and the feasibility of $\bar{p}$. Furthermore, if the limit point $\bar{p}$ of $(p^k)_{k\in\mathbb{N}}$ is feasible, it can be shown to satisfy the AKKT conditions for Problem (1), as established in the following lemma.

Lemma 11. *Assume that* $\lim_{k\to+\infty} p^k = \bar{p}$ *and that condition* (A1) *holds. If* $\bar{p} \in \Omega$*, then* $\bar{p}$ *is an AKKT point for Problem* (1)*.*

Proof. Since $\lim_{k\to+\infty} p^k = \bar{p}$, item (i) of Definition 5 is satisfied. From (28), **Step 1** of Algorithm 1, and the fact that $\lim_{k\to+\infty} \epsilon_k = 0$, it follows that $\lim_{k\to+\infty} \nabla L(p^k, \lambda^k, \mu^k) = 0$. Thus, $\bar{p}$ satisfies item (ii) of Definition 5. Since $\bar{p} \in \Omega$, item (ii) of Lemma 9 implies that for any $j \in \{1, \ldots, m\}$ with $g_j(\bar{p}) < 0$, we have $\lim_{k\to+\infty} \mu_j^k = 0$. Consequently, $\lim_{k\to+\infty} \mu_j^k = 0$ for all $j \notin \mathcal{A}(\bar{p})$. Therefore, $\bar{p}$ also satisfies item (iii) of Definition 5, completing the proof. ☐

To proceed with the analysis of the sequence $(p^k)_{k\in\mathbb{N}}$ generated by Algorithm 1, let us recall an important result, whose proof is detailed in [3, Theorems 3.4 and 3.5].

Theorem 12. *Suppose that* $\bar{p} \in \Omega$ *satisfies RCPLD or CRSC. If* $\bar{p}$ *is an AKKT point, then* $\bar{p}$ *is a KKT point.*

A direct application of Lemma 11 and Theorem 12 yields the following theorem.

Theorem 13. *Assume that* $\lim_{k\to+\infty} p^k = \bar{p}$ *and that condition (A1) holds. Suppose* $\bar{p} \in \Omega$ *satisfies either RCPLD or CRSC. Then,* $\bar{p}$ *is a KKT point.*

In Lemma 11 and Theorem 13, it is assumed that the limit point $\bar{p}$ of the sequence $(p^k)_{k\in\mathbb{N}}$ satisfies $\bar{p} \in \Omega$. Under this assumption, $\bar{p}$ is an AKKT point of Problem (1). If the RCPLD or CRSC conditions are satisfied, then $\bar{p}$ is also a KKT point for Problem (1). Next, consider the case where $\bar{p} \notin \Omega$. In this scenario, it is assumed that the sequences $(\tau_k)_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ and $(\theta_k)_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ are chosen such that

$$\lim_{k\to\infty} \tau_k = \bar{\tau} \in (0, +\infty), \qquad \lim_{k\to\infty} \theta_k = \bar{\theta} \in (0, +\infty). \tag{31}$$

Theorem 14. *Assume that* $\lim_{k\to+\infty} p^k = \bar{p}$ *and* $\bar{p} \notin \Omega$*. Then, there exists a bounded sequence* $((\bar{\rho}^k, \bar{\sigma}^k))_{k\in\mathbb{N}} \subset \mathbb{R}^s_+ \times \mathbb{R}^m_+$ *such that*

$$\lim_{k\to+\infty} \operatorname{grad}\Big(\sum_{i=1}^{s} \bar{\rho}_i^k \varphi_{\tau_k}(h_i(p^k)) + \sum_{j=1}^{m} \bar{\sigma}_j^k \big(g_j(p^k) + \psi_{\theta_k}(g_j(p^k))\big)\Big) = 0. \tag{32}$$

As a consequence, if (31) *holds, then there exists* $(\bar{\rho}, \bar{\sigma}) \in \mathbb{R}^s_+ \times \mathbb{R}^m_+$*, with* $(\bar{\rho}, \bar{\sigma}) \neq 0$*, such that* $\bar{p}$ *is a stationary point of the following optimization problem:*

$$\operatorname*{Minimize}_{p\in\mathcal{M}} \sum_{i=1}^{s} \bar{\rho}_i \varphi_{\bar{\tau}}(h_i(p)) + \sum_{j=1}^{m} \bar{\sigma}_j \left(g_j(p) + \psi_{\bar{\theta}}(g_j(p))\right). \tag{33}$$

Proof. Since $\bar{p} \notin \Omega$, there exists at least one index i or j such that $h_i(\bar{p}) \neq 0$ or $g_j(\bar{p}) > 0$. By items (iii) and (iv) of Lemma 9, the sequence $((\lambda^k, \mu^k))_{k\in\mathbb{N}} \subset \mathbb{R}^s_{++} \times \mathbb{R}^m_{++}$ is unbounded. Considering that φ_{τ_k} and ψ_{θ_k} are continuously differentiable, and using (27), Definition 6 and $\lim_{k\to+\infty} p^k = \bar{p} \notin \Omega$, it follows that $((\rho^k, \sigma^k))_{k\in\mathbb{N}}$ is also unbounded. With $\lim_{k\to+\infty} \epsilon_k = 0$, Step 1 of Algorithm 1 implies

$$\lim_{k\to\infty} \frac{1}{\|(\rho^k, \sigma^k)\|_1} \operatorname{grad} \mathbb{L}_{(\tau_k, \theta_k)}(p^k, \rho^k, \sigma^k) = 0. \tag{34}$$

Set $(\bar{\rho}^k, \bar{\sigma}^k) := (\rho^k, \sigma^k)/\|(\rho^k, \sigma^k)\|_1$. Therefore, using taking into account that $\operatorname{grad} f$ is continuous and $\lim_{k\to+\infty} p^k = \bar{p}$, the equality (32) follows from (34), which prove the fist statement.

As $((\bar{\rho}^k, \bar{\sigma}^k))_{k\in\mathbb{N}}$ is bounded, we can assume that $\lim_{k\to+\infty}((\rho^k, \sigma^k)/\|(\rho^k, \sigma^k)\|_1) = (\bar{\rho}, \bar{\sigma})$, with $(\bar{\rho}, \bar{\sigma}) \neq 0$, it follows from (32) and the continuity of $\operatorname{grad} h_i$, and $\operatorname{grad} g_j$ that

$$\sum_{i=1}^{s} \bar{\rho}_i \varphi'_{\bar{\tau}}(h_i(\bar{p})) \operatorname{grad} h_i(\bar{p}) + \sum_{j=1}^{m} \bar{\sigma}_j \left(1 + \psi'_{\bar{\theta}}(g_j(\bar{p}))\right) \operatorname{grad} g_j(\bar{p}) = 0,$$

proving that $\bar{p}$ is a critical point of problem (33), thus completing the proof. □

Exact penalty methods have their foundation on strong theoretical principles and can effectively solve a wide range of constrained optimization problems by exploring feasible regions. However, their practical application is often limited by numerical instability, especially in high-dimensional settings. This instability typically arises from the need to increase penalty parameters substantially (see [39, Chapter 17], [7]). Nonetheless, for suitably chosen penalty parameters ρ and σ, the solutions to the non-differentiable ℓ_1-exact penalized problem also solve the original problem (see, for example, [7, Chapter 4]). Since problem (26) serves as a smooth approximation of the ℓ_1-exact penalized problem (16), it is expected that increasing penalty parameters will induce less instability in this context. The next section analyzes the convergence properties of Algorithm 1, assuming that the sequences of penalty parameters $(\rho^k)_{k\in\mathbb{N}}$ and $(\sigma^k)_{k\in\mathbb{N}}$ remain bounded.

4.2 Convergence analysis of Sℓ_1-EPA

In this section we analyze the convergence properties of Sℓ_1-EPA, demonstrating its effectiveness in finding optimal solutions when the penalty parameter sequences $(\rho^k)_{k\in\mathbb{N}}$ and $(\sigma^k)_{k\in\mathbb{N}}$ are bounded. Assuming that these parameters remain bounded and the smoothing parameter sequences $(\rho^k)_{k\in\mathbb{N}}$ and $(\theta_k)_{k\in\mathbb{N}}$ grows unbounded, it is shown that the cluster points of the sequence $(p^k)_{k\in\mathbb{N}}$ generated by Sℓ_1-EPA satisfy the KKT conditions for Problem (1). *The formal assumption necessary for this section, which is applied only when explicitly stated, is as follows:*

(A2) The penalty sequences $(\rho^k)_{k\in\mathbb{N}}$ and $(\sigma^k)_{k\in\mathbb{N}}$ are bounded.

Before proceeding with the analysis, some remarks on assumptions (A1) and (A2) are necessary, as they are crucial in the context of the ℓ_1-exact penalized problem (16) and must be considered together. Traditional penalty methods often rely on unboundedly increasing penalty parameters to ensure feasibility (see Theorem 8). However, this requirement can impose unrealistic or computationally prohibitive conditions. Bounded penalty parameters, on the other hand, may prevent strict feasibility in the classical sense, which motivates assumption (A1) to ensure that subproblem (26) sufficiently approximates the ℓ_1-exact penalized problem (16). As discussed in Section 3.1, the parameters ρ_i and σ_j need not diverge for the solutions of the non-differentiable ℓ_1-exact penalized problem (16) to also be solutions of the original problem (1) (see, for example, [7, Chapter 4]). Thus, it is both practical and theoretically meaningful to explore the balance between the growth of penalty parameters and the smoothing parameter. By examining their interaction, we can assess their impact on the accuracy and quality of feasible solutions. In the following analysis, the essential properties of the penalty parameters are established to ensure the existence of Lagrange multipliers, even when these parameters remain bounded.

Lemma 15. *Assume that $\bar{p}$ is a cluster point of the sequence $(p^k)_{k\in\mathbb{N}}$, and let $(p^{k_\ell})_{\ell\in\mathbb{N}}$ be a subsequence such that $\lim_{\ell\to+\infty} p^{k_\ell} = \bar{p}$. Then, the following conditions hold:*

(i) The sequences $(\rho^k)_{k\in\mathbb{N}} \subset \mathbb{R}^s_{++}$ and $(\sigma^k)_{k\in\mathbb{N}} \subset \mathbb{R}^m_{++}$.

(ii) If $(\rho^k_i)_{k\in\mathbb{N}}$ is bounded for some $i \in \{1,\ldots,s\}$, then $h_i(\bar{p}) = 0$. Additionally, the sequence $(\lambda^k_i)_{k\in\mathbb{N}}$ is bounded.

(iii) If the sequence $(\sigma^k_j)_{k\in\mathbb{N}}$ is bounded for some $j \in \{1,\ldots,m\}$, then $g_j(\bar{p}) \le 0$. Moreover, the sequence $(\mu^k_j)_{k\in\mathbb{N}}$ is bounded. If the smoothing parameter sequence $(\theta_k)_{k\in\mathbb{N}}$ satisfies assumption (A1) and $g_j(\bar{p}) < 0$, then $\lim_{\ell\to+\infty} \mu^{k_\ell}_j = 0$.

Consequently, if the sequences $(\rho^k)_{k\in\mathbb{N}}$ and $(\sigma^k)_{k\in\mathbb{N}}$ satisfy assumption (A2), and the smoothing parameter sequence $(\theta_k)_{k\in\mathbb{N}}$ satisfies (A1), then $\bar{p} \in \Omega$.

Proof. The proof of item (i) follows directly from that of item (i) in Lemma 9. We now proceed to prove item (ii). Since $\tilde{\nu} > 1$ and $(\rho^k_i)_{k\in\mathbb{N}}$ is bounded, it follows from **Step 2** that the equality $\rho^{k+1}_i = \tilde{\nu}\rho^k_i$ cannot hold for infinitely many $k \in \mathbb{N}$. Therefore, there exists $k_0 \in \mathbb{N}$ such that inequality (24) holds for all $k > k_0$. In particular, (24) implies that $\lim_{k\to+\infty} |h_i(p^k)| = 0$. Given that $\lim_{\ell\to+\infty} p^{k_\ell} = \bar{p}$ and the continuity of h, it follows that $h_i(\bar{p}) = 0$ for all $i \in \{1,\ldots,s\}$. This completes the proof of the first statement. The second statement follows from the first equality in (27), along with the boundedness of $(\rho^k_i)_{k\in\mathbb{N}}$ and items (iv) and (v) of Definition 6.

To prove item (iii), observe that, due to $\hat{\nu} > 1$ and the boundedness of $(\sigma^k_j)_{k\in\mathbb{N}}$, it follows from **Step 2** that the equality $\sigma^{k+1}_j = \hat{\nu}\sigma^k_j$ cannot hold for infinitely many $k \in \mathbb{N}$. Thus, there exists $k_0 \in \mathbb{N}$ such that (25) holds for all $k > k_0$. In particular, (25) implies that $\lim_{k\to+\infty} g_j(p^k)_+ = 0$. Since $\lim_{\ell\to+\infty} p^{k_\ell} = \bar{p}$, it follows that $g_j(\bar{p})_+ = 0$, proving the first statement in item (iii). The boundedness of $(\mu^k_j)_{k\in\mathbb{N}}$ follows from the boundedness of $(\sigma^k_j)_{k\in\mathbb{N}}$ and items (iv) and (v) of Definition 6. Assuming $g_j(\bar{p}) < 0$, it follows from item $(viii)$ of Definition 6 and assumption (A1) that

$$\lim_{\ell\to+\infty} \psi'_{\theta_{k_\ell}}(g_j(p^{k_\ell})) = -1.$$

Using the second equality in (27) and the boundedness of $(\sigma^k_j)_{k\in\mathbb{N}}$, we conclude that $\lim_{\ell\to+\infty} \mu^{k_\ell}_j = 0$, completing the proof of the last statement in item (iii).

Finally, to prove the last statement of the lemma, assume that $(\rho^k)_{k\in\mathbb{N}}$ and $(\sigma^k)_{k\in\mathbb{N}}$ satisfy assumption (A2). It follows from items *(ii)* and *(iii)* that $h_i(\bar{p}) = 0$ for all $i \in \{1,\dots,s\}$ and $g_j(\bar{p}) \le 0$ for all $j \in \{1,\dots,m\}$. Hence, $\bar{p} \in \Omega$, completing the proof. □

An essential aspect of Lemma 15 is that the cluster point $\bar{p}$ of the sequence $(p^k)_{k\in\mathbb{N}}$ generated by Algorithm 1 is not initially assumed to be feasible. Nevertheless, under assumptions (A2) and (A1), it is guaranteed that the sequences of Lagrange multipliers $(\lambda^k)_{k\in\mathbb{N}}$ and $(\mu^k)_{k\in\mathbb{N}}$ remain bounded, and that the cluster point $\bar{p}$ becomes feasible. With these conditions satisfied, it is possible to demonstrate that the feasible points satisfy the AKKT conditions for Problem (1). This is formally established in the subsequent lemma.

Lemma 16. *Suppose that assumptions (A1) and (A2) hold. Let $\bar{p}$ be a cluster point of the sequence $(p^k)_{k\in\mathbb{N}}$. Then, $\bar{p}$ is an AKKT point for Problem* (1)*.*

Proof. Let $(p^{k_\ell})_{\ell\in\mathbb{N}}$ be a subsequence of $(p^k)_{k\in\mathbb{N}}$ such that $\lim_{\ell\to+\infty} p^{k_\ell} = \bar{p}$. By assumptions (A1) and (A2), and using Lemma 15, it follows that $\bar{p} \in \Omega$, thus satisfying item *(i)* of Definition 5. Furthermore, by combining the equalities in (27) with (5) and (10), we obtain

$$\operatorname{grad} L(p^{k_\ell}, \lambda^{k_\ell}, \mu^{k_\ell}) = \operatorname{grad} \mathbb{L}_{(\tau_{k_\ell},\theta_{k_\ell})}(p^{k_\ell}, \rho^{k_\ell}, \sigma^{k_\ell}),$$

for all $\ell \in \mathbb{N}$. Given that $\lim_{\ell\to+\infty} \epsilon_{k_\ell} = 0$, Step 1 of Algorithm 1 implies

$$\lim_{\ell\to+\infty} \operatorname{grad} L(p^{k_\ell}, \lambda^{k_\ell}, \mu^{k_\ell}) = 0,$$

satisfying item *(ii)* of Definition 5. According to item *(iii)* of Lemma 15, for any $j \in \{1,\dots,m\}$ with $g_j(\bar{p}) < 0$, we have $\lim_{\ell\to+\infty} \mu_j^{k_\ell} = 0$. Consequently, $\lim_{\ell\to+\infty} \mu_j^{k_\ell} = 0$ for all $j \notin \mathcal{A}(\bar{p})$, satisfying item *(iii)* of Definition 5. This completes the proof. □

A direct application of Lemma 15, Lemma 16 and Theorem 12 yields the following theorem.

Theorem 17. *Suppose that assumptions (A1) and (A2) hold. Let $\bar{p}$ be a cluster point of the sequence $(p^k)_{k\in\mathbb{N}}$. If $\bar{p}$ satisfies either RCPLD or CRSC, then for any subsequence $(p^{k_\ell})_{\ell\in\mathbb{N}}$ of $(p^k)_{k\in\mathbb{N}}$ converging to $\bar{p}$, i.e., $\lim_{\ell\to+\infty} p^{k_\ell} = \bar{p}$, the associated sequences $(\lambda^{k_\ell})_{\ell\in\mathbb{N}}$ and $(\mu^{k_\ell})_{\ell\in\mathbb{N}}$ are bounded. Moreover, any cluster points $\bar{\lambda}$ of $(\lambda^{k_\ell})_{\ell\in\mathbb{N}}$ and $\bar{\mu}$ of $(\mu^{k_\ell})_{\ell\in\mathbb{N}}$ are Lagrange multipliers associated with $\bar{p}$. Consequently, $\bar{p}$ is a KKT point.*

In Lemma 15, conditions (A2) and (A1) are assumed to ensure that $\bar{p} \in \Omega$. However, when $\bar{p} \notin \Omega$, both conditions (A2) and (A1) cannot hold simultaneously, leading to unbounded sequences of Lagrange multipliers, $(\lambda^k)_{k\in\mathbb{N}}$ and $(\mu^k)_{k\in\mathbb{N}}$, as implied by Lemma 15. Additionally, let the sequences $(\tau_k)_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ and $(\theta_k)_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ be chosen to satisfy (31). In this case, a result analogous to Theorem 14 is obtained, following a similar proof structure, which is therefore omitted.

Theorem 18. *Let $\bar{p}$ be a cluster point of the sequence $(p^k)_{k\in\mathbb{N}}$ such that $\bar{p} \notin \Omega$ and $(p^{k_\ell})_{\ell\in\mathbb{N}}$ a subsequence of $(p^k)_{k\in\mathbb{N}}$ converging to $\bar{p}$. Then, there exists a bounded sequence $((\bar{\rho}^{k_\ell}, \bar{\sigma}^{k_\ell}))_{\ell\in\mathbb{N}} \subset \mathbb{R}^s_+ \times \mathbb{R}^m_+$ such that*

$$\lim_{\ell\to+\infty} \operatorname{grad} \Big(\sum_{i=1}^{s} \bar{\rho}_i^{k_\ell} \varphi_{\tau_{k_\ell}}(h_i(p^{k_\ell})) + \sum_{j=1}^{m} \bar{\sigma}_j^{k_\ell} \big(g_j(p^{k_\ell}) + \psi_{\theta_{k_\ell}}(g_j(p^{k_\ell})) \big) \Big) = 0.$$

Consequently, if (31) *holds, then there exists $(\bar{\rho}, \bar{\sigma}) \in \mathbb{R}^s_+ \times \mathbb{R}^m_+$ with $(\bar{\rho}, \bar{\sigma}) \neq 0$ such that $\bar{p}$ is a stationary point of the following optimization problem*

$$\underset{p\in\mathcal{M}}{\text{Minimize}} \sum_{i=1}^{s} \bar{\rho}_i \varphi_{\bar{\tau}}(h_i(p)) + \sum_{j=1}^{m} \bar{\sigma}_j \left(g_j(p) + \psi_{\bar{\theta}}(g_j(p)) \right).$$

As discussed earlier, increasing the penalty parameter enforces feasibility but can lead to numerical instability. In contrast, increasing the smoothing parameter provides better approximations of the ℓ_1-exact penalized problem (16), potentially reducing the need for a high penalty parameter. However, when the smoothing parameter becomes too large, problem (16) approaches a non-differentiable form, making it more challenging to solve. The next section presents numerical experiments that illustrate the performance of Algorithm 1 across different problem classes, providing insights into the behavior of the method under various parameter settings.

5 Numerical experiments

In this section, we report numerical experiments to illustrate the performance of Algorithm 1. We implemented Algorithm 1 in Matlab (R2024b, version 24.2.0.2712019). The code is available for download at `https://github.com/lfprudente/Sl1EPA.git`. In the numerical experiments, we considered three families of problems. The first family corresponds to the non-negative principal component analysis problem, already considered in [33]. The second family of problems corresponds to the problem of packing circles in an ellipse, introduced in [8] and also considered in [17]. The last family of problems corresponds to the four classification problems introduced in [12]. All tests were conducted on a computer with a 3.7 GHz Intel Core i5 6-Core processor and 8GB 2667MHz DDR4 RAM, running macOS Sequoia 15.0.

In the implementation, following basic ideas of augmented Lagrangian methods in Euclidean spaces (see [11, p.116]), given tolerances $\varepsilon_{\rm feas}, \varepsilon_{\rm compl}, \varepsilon_{\rm opt} > 0$, we defined as stopping criterion

$$\max\left\{\max_{i=1,\ldots,s}\left\{|h_i(p^k)|\right\}, \max_{j=1,\ldots,m}\left\{\max\{0, g_j(p^k)\}\right\}\right\} \leq \varepsilon_{\rm feas}, \tag{35}$$

$$\max_{j=1,\ldots,m}\left\{\left|\min\{-g_j(p^k), \mu_j^k\}\right|\right\} \leq \varepsilon_{\rm compl}, \tag{36}$$

$$\left\|\operatorname{grad}\mathbb{L}_{(\tau_k,\theta_k)}(p^k, \rho^k, \sigma^k)\right\|_\infty \leq \varepsilon_{\rm opt}. \tag{37}$$

It is worth recalling that, due to (5), (10), and (27), $\operatorname{grad}\mathbb{L}_{(\tau_k,\theta_k)}(p^k, \rho^k, \sigma^k) = \operatorname{grad} L(p^k, \lambda^k, \mu^k)$. Therefore, the stopping criterion does not depend on the smoothing function, the penalty parameter, or the smoothing parameter, but only on the primal iterate p^k and the estimates λ_i^k and μ_j^k of the Lagrange multipliers. As tolerance for the subproblem of iteration k, we considered $\epsilon_1 = \sqrt{\varepsilon_{\rm opt}}$ and $\epsilon_k = \max\{\epsilon_{k-1}/10, \varepsilon_{\rm opt}\}$ for $k \geq 1$. For the initial value of the penalty parameters (see [11, p.153]), we considered

$$\rho_i^1 = \max\left\{10^{-8}, \min\left\{10\frac{\max\{1, |f(p^0)|\}}{\max\{1, \varphi_{\tau_1}(h_i(p^0))\}}, 10^8\right\}\right\} \text{ for } 1, \ldots, s$$

and

$$\sigma_j^1 = \max\left\{10^{-8}, \min\left\{10\frac{\max\{1, |f(p^0)|\}}{\max\{1, [g_j(p^0) + \psi_{\theta_1}(g_j(p^0))]/2\}}, 10^8\right\}\right\} \text{ for } j = 1, \ldots, m,$$

as well as $\tilde{\beta} = \hat{\beta} = \frac{1}{2}$ and $\tilde{\nu} = \hat{\nu} = 10$ for their update. For the smoothing parameters, we considered $\tau_k = \theta_k = 10^{k-1}$. For the computation of $p^k \in \mathcal{M}$, an approximate solution to (26), we employed the RLBFGS routine, which implements a Riemannian limited-memory BFGS algorithm for unconstrained optimization [24], available in the Manopt toolbox (version 7.1). In the implementation of the smoothing functions $\phi_{2,\tau}$ and $\phi_{6,\tau}$, to avoid a possible overflow caused by the calculation of e^x with $x > 0$, we used

$$\ln\left(e^{\tau t} + e^{-\tau t}\right) = \tau|t| + \ln(1 + e^{-2\tau|t|})$$

for all $t \in \mathbb{R}$.

In the numerical experiments, we also considered a variant of Algorithm 1 in which a single penalty parameter is used for all of the constraints. In this variant, we considered $\rho_i^1 = \zeta$ for all i and $\sigma_j^1 = \zeta$ for all j with

$$\zeta = \max\left\{10^{-8}, \min\left\{10\frac{\max\left\{1, |f(p^0)|\right\}}{\max\left\{1, \sum_{i=1}^{s}\varphi_{\tau_1}(h_i(p^0)) + \sum_{j=1}^{m}[g_j(p^0) + \psi_{\theta_1(g_j(p^0))}]/2\right\}}, 10^8\right\}\right\}.$$

Moreover, Step 2 of Algorithm 1 was replaced by

Step 2'. (Updating penalty parameters) If

$$\max\{\|h(p^k)\|_\infty, \|g(p^k)_+\|_\infty\} \leq \tilde{\beta}\max\{\|h(p^{k-1})\|_\infty, \|g(p^{k-1})_+\|_\infty\},$$

then set $\rho_i^{k+1} = \rho_i^k$ for all i and $\sigma_j^{k+1} = \sigma_j^k$ for all j. Otherwise, set $\rho_i^{k+1} = \tilde{\nu}\rho_i^k$ for all i and $\sigma_j^{k+1} = \tilde{\nu}\sigma_j^k$ for all j.

The convergence theory of this variant, which was not taken into account in the previous sections, is analogous to that of the analyzed algorithm.

In all numerical experiments, we considered the same smooth approximation of the absolute value function to handle both equality and inequality constraints. Since we have six different options $\phi_{1,\tau}, \ldots, \phi_{6,\tau}$, and for each option we can consider a penalty parameter for each constraint or a single penalty parameter for all constraints, we considered a total of twelve variations of Algorithm 1. We have arbitrarily set $r = 2$ for the functions $\phi_{1,\tau}$ and $\phi_{5,\tau}$.

5.1 Non-negative principal component analysis problem

The problem of non-negative principal component analysis (non-negative PCA) arises in the context of the *spiked model* [38], in which the data matrix $A \in \mathbb{R}^{n\times n}$ is constructed as

$$A = \sqrt{\beta}\, v_0 v_0^T + B,$$

where $v_0 \in \mathbb{R}_+^n$ is the principal signal vector with $\|v_0\| = 1$, $\beta > 0$ is the signal-to-noise ratio, and $B \in \mathbb{R}^{n\times n}$ is a symmetric noise matrix. The goal is to recover the signal v_0 by solving

$$\underset{v\in\mathbb{R}^n}{\text{Minimize}} -v^T A v \text{ subject to } \|v\| = 1 \text{ and } v \geq 0.$$

The constraint $\|v\| = 1$ ensures that v lies on the unit sphere $\mathbb{S}^{n-1}$. Therefore, the problem can be rewritten in the format of (1) as

$$\underset{v\in\mathbb{S}^{n-1}}{\text{Minimize}} -v^T A v \text{ subject to } v \geq 0.$$

To generate synthetic instances of the problem, we followed the methodology outlined in [33]. The matrix A was constructed varying the problem parameters:

- the dimension $n \in \{10, 50, 200, 500, 1000, 2000\}$,

- the signal-to-noise ratio $\beta \in \{0.05, 0.1, 0.25, 0.5, 1.0, 2.0\}$, and
- the sparsity level $\delta \in \{0.1, 0.3, 0.7, 0.9\}$, which determines the proportion of nonzero entries in v_0.

We therefore generated $6 \times 6 \times 4 = 144$ different instances. For each combination of n, β, and δ, we randomly generated the support $S \subset \{1, \ldots, n\}$ of cardinality $|S| = \lfloor \delta n \rfloor$, and defined the entries of v_0 as

$$[v_0]_i = \begin{cases} \frac{1}{\sqrt{|S|}}, & \text{if } i \in S, \\ 0, & \text{otherwise.} \end{cases}$$

The noise matrix B was constructed independently for each problem instance, where its off-diagonal entries follow a Gaussian distribution $\mathcal{N}(0, 1/n)$, and its diagonal entries follow a Gaussian distribution $\mathcal{N}(0, 2/n)$.

In the experiments with the non-negative PCA, we considered $\varepsilon_{\text{feas}} = \varepsilon_{\text{opt}} = \varepsilon_{\text{compl}} = 10^{-4}$. Lower values lead to failures of the method used to solve the subproblems. Detailed tables with the results of applying the twelve variations of Algorithm 1 on the 144 instances can be found in `https://github.com/lfprudente/Sl1EPA.git`. Next we present an analysis using performance profiles. The first observation is that, with the exception of a single variation on a single instance, all twelve variations found in all 144 instances a point that satisfies the nonnegativity constraints exactly, i.e., they stopped at a $v^k \geq 0$. (In the one exception, the variant stopped at a v^k that satisfied the constraints with the required precision $\varepsilon_{\text{feas}}$). For a given instance, let $f_1, \ldots, f_{12}$ be the function values of the final iterate v^k of each of the twelve variations of Algorithm 1. Let $f_{\min} = \min_{\{i=1,\ldots,12\}} \{f_i\}$. We say that the variant i found a solution equivalent to the best one if

$$f_i \leq f_{\min} + f_{\text{tol}} \max\{1, |f_{\min}|\},$$

where $f_{\text{tol}} > 0$ is a pre-specified precision. Table 1 shows the number of instances in which each variant of Algorithm 1 found a best-equivalent solution, as a function of $f_{\text{tol}} \in \{0.1, 10^{-2}, \ldots, 10^{-5}\}$. In column SC, the table also shows the number of instances in which each variant managed to find a point that satisfies the stopping criterion (35,36,37). In the remaining instances, the variants stopped because the subproblem solver failed to find a point that satisfies the stopping criterion (23) in two consecutive iterations. Figures in the table show that at low accuracies, the solutions found by the different variants are all considered equivalent, but at higher accuracies, some variants stand out, or in other words, some variants failed to find solutions equivalent to the best solution. For the variants using $\phi_{3,\tau}$ and $\phi_{4,\tau}$ with multiple penalty parameters and the variant using $\phi_{4,\tau}$ with a single penalty parameter, it seems clear that the failure for high accuracies is related to the difficulty in solving the subproblems. It is also worth noting that for the variants using $\phi_{4,\tau}$, $\phi_{5,\tau}$ and $\phi_{6,\tau}$, which approximate the modulus from below, it is practically the same whether a single penalty parameter or a different penalty parameter is used for each constraint. For the three other variants, which approximate the modulus from above, it is always better to consider a single penalty parameter. Therefore, there seems to be no practical advantage in using multiple penalty parameters. A similar result was observed in [2] in the context of augmented Lagrangian methods.

In relation to the quality of the obtained solution, considering $f_{\text{tol}} = 10^{-5}$, the variants using $\phi_{1,\tau}$ and $\phi_{5,\tau}$ stand out. It is worth noting that the variant using $\phi_{1,\tau}$ with a single penalty parameter corresponds, basically, to the Q^{lqr}-algorithm considered in [33]. If, on the other hand, we consider that $f_{\text{tol}} = 0.1$ is sufficient to consider the values of the objective function equivalent, then all methods find equivalent solutions. In that case, it is worth asking which one is more efficient.

Variant			f_{tol}				
Penalty parameter	Index i of $\phi_{i,\tau}$	SC	0.1	10^{-2}	10^{-3}	10^{-4}	10^{-5}
single	1	127	144	144	143	136	124
	2	130	144	143	141	134	107
	3	120	144	144	143	104	46
	4	80	144	141	125	59	23
	5	110	144	143	141	132	115
	6	104	144	144	143	120	75
multiple	1	110	144	143	141	132	115
	2	104	144	144	143	120	75
	3	72	144	144	123	76	36
	4	77	144	141	124	58	22
	5	110	144	143	141	132	115
	6	104	144	144	143	120	75

Table 1: Comparison of the function values found by the different variants of Algorithm 1, as a function of the f_{tol} tolerance used to determine whether a function value is equivalent to the best or not.

Figure 2 shows the performance profiles of the twelve variants. In the figure, for $i = 1, \dots, 12$,

$$\Gamma_i(\kappa) = \frac{\left|\{j \in \{1, \dots, q\} \mid t_{ij} \leq \kappa \min_{\{\ell=1,\dots,12\}}\{t_{\ell j}\}\}\right|}{q} \text{ for } \kappa \geq 1,$$

where $q = 144$ is the number of instances considered, t_{ij} is the CPU time of the variant i applied to instance j, and i from 1 to 6 corresponds to the variants that use $\phi_{\tau,i}$ and a single penalty parameter (solid lines in the graphic), and i from 7 to 12 corresponds to the variants that use $\phi_{\tau,7-i}$ and multiple penalty parameters (dashed lines in the graphic). The figure emphasizes that in cases where the absolute value is approximated from below, it makes little difference whether a single or multiple penalty parameters are considered. But for the variants that approximate it from above, using a single penalty parameter is consistently more efficient. In terms of efficiency, the variants using $\phi_{2,\tau}$ and $\phi_{3,\tau}$ with a single penalty parameter stand out. What the experiments show is that having a theory that encompasses a variety of smoothing functions allows different alternatives to be tested when addressing a particular problem.

5.2 Packing circles within ellipses

The circle packing problem considered in this section seeks to maximize the radius r of N identical circles that can be arranged without overlapping within a fixed-size two-dimensional container [35]. This problem has diverse applications, as highlighted in [8, 35]. If we consider the container to be an ellipse with semi-axes $a \geq b > 0$, then, employing continuous variables $(r; u, v, s) \in \mathbb{R} \times (\mathbb{R}^n)^3$,

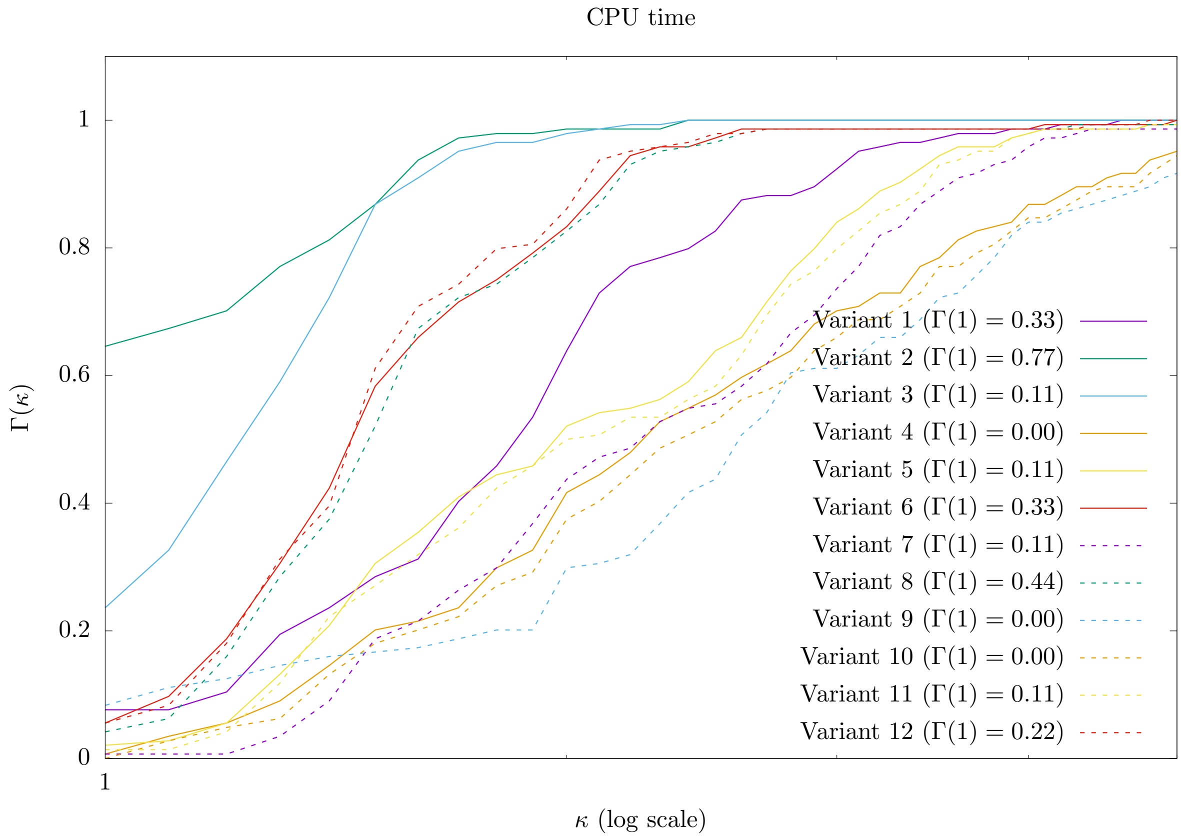


Figure 2: Performance profiles comparing the CPU time spent by the twelve variants of Algorithm 1 in the 144 instances of the non-negative PCA in which, with $f_{\text{tol}} = 0.1$, it was considered that all variants found equivalent solutions.

the problem can be formulated [8] as follows

$$
\begin{array}{lll}
\underset{(r;u,v,s)\in\mathbb{R}\times(\mathbb{R}^n)^3}{\text{Maximize}} & r & \\
\text{subject to} & u_i^2 + v_i^2 = 1, & i = 1,\dots,N \\
& x_i = a[1 + (s_i - 1)(b^2/a^2)]u_i, & i = 1,\dots,N \\
& y_i = b s_i v_i, & i = 1,\dots,N \\
& b^2(s_i - 1)^2[(b^2/a^2)u_i^2 + v_i^2] \geq r^2, & i = 1,\dots,N \\
& 0 \leq s_i \leq 1, & i = 1,\dots,N \\
& (x_i - x_j)^2 + (y_i - y_j)^2 \geq (2r)^2, & i = 1,\dots,N,\ j = i+1,\dots,N \\
& r \geq 0. &
\end{array}
$$

The constraint $u_i^2 + v_i^2 = 1$ ensures that (u_i, v_i) lies on the unit circle $\mathbb{S}^1 \subset \mathbb{R}^2$. Therefore, the problem can be rewritten in the format of (1) as

$$\begin{array}{lll}
\underset{(r;u,v;s)\in\mathbb{R}\times(\mathbb{S}^1)^N\times\mathbb{R}^n}{\text{Minimize}} & -r & \\
\text{subject to} & x_i = a[1 + (s_i - 1)(b^2/a^2)]u_i, & i = 1, \dots, N \\
& y_i = b s_i v_i, & i = 1, \dots, N \\
& b^2(s_i - 1)^2[(b^2/a^2)u_i^2 + v_i^2] \geq r^2, & i = 1, \dots, N \\
& 0 \leq s_i \leq 1, & i = 1, \dots, N \\
& (x_i - x_j)^2 + (y_i - y_j)^2 \geq (2r)^2, & i = 1, \dots, N,\ j = i + 1, \dots, N \\
& r \geq 0. &
\end{array}$$

In both models, x_i and y_i correspond to the Cartesian coordinates (x_i, y_i) of the center of the ith circle. In fact, they do not represent model variables and were included only to simplify the presentation. In practice, the expression on the right in the equations defining x_i and y_i is used to substitute x_i and y_i in the second to last inequality, which represents the non-overlap between the circles.

In the experiments with the packing problem, we also considered $\varepsilon_{\text{feas}} = \varepsilon_{\text{opt}} = \varepsilon_{\text{compl}} = 10^{-4}$. We solved instances with $N \in \{5, 6, \dots, 10\} \cup \{20, 30, \dots, 100\}$ and an ellipse with semi-axes $a = 2$ and $b = 1$ using the six variants of Algorithm 1 with a single penalty parameter. Tables 2 and 3 show the results. Figure 3 illustrates some of the solutions found. In the tables, r is the radius found (i.e., minus the minimized objective function), $\|c\|$, $\|s\|$, and $\|\operatorname{grad}\mathbb{L}\|$ are the left-hand sides of (35), (36), and (37), respectively. Column SC stands for "stopping criterion". In this column, 0 means that (35,36,37) holds, while 4 means that Algorithm 1 stopped because the subproblem solver did not find a point satisfying the stopping criterion (23) in two consecutive iterations. Column k corresponds to the number of "outer" iterations performed by Algorithm 1, while $\sum_{i=1}^k k_i$ corresponds to the sum of the "inner" iterations, i.e., the sum of the iterations required by the inner solver to tackle all subproblems. The last three columns labeled with #fcnt, #gcnt, and Time correspond to different measures of the computational cost, namely, the total number of evaluations of the objective function, the total number of evaluations of its gradient, and the elapsed CPU time in seconds. In the tables, the best solutions are emphasized in bold.

The first observation is that the problem considered in this section is a non-convex problem with many local non-global minimizers. For this reason, the different variants generally find different solutions. If we consider as equivalent the values of r that are equal to 4 significant digits in the tables, then the variants from 1 to 6 find 5, 9, 3, 1, 7, and 4 best solutions, respectively. In terms of robustness, using this criterion, variant 2 would be the best, followed by variant 5. Variant 1 is the only one where the inner solver never failed to solve a subproblem. For variants 1 to 6, out of a total of 15 instances, the number of failures is 0, 2, 5, 7, 2, and 4. It is important to note that the failures of the inner solver are concentrated on the larger problems, as there are no failures recorded in Table 2. Another observation from the tables is that variants 1, 2, and 3, which approximate the absolute value from above, have fewer failures in solving the subproblems than their respective companions 5, 6, and 4, which approximate the absolute value from below. In summary, variant 2 stands out for having found the largest number of best solutions, and variant 1 stands out for apparently having subproblems that are easier to solve. The latter would be an advantage in a multistart strategy that tries to find good quality local solutions, i.e. better approximations of global minimizers.

N	$\phi_{i,\tau}$	r	$\|c\|$	$\|s\|$	$\|\operatorname{grad}\mathbb{L}\|$	SC	k	$\sum_{i=1}^{k} k_i$	#fcnt	#gcnt	Time
5	1	**5.236e-01**	0e+00	4e-05	9e-05	0	12	254	372	266	4.1
	2	**5.236e-01**	0e+00	2e-05	7e-06	0	6	181	259	187	3.3
	3	5.235e-01	0e+00	5e-05	6e-05	0	5	185	357	190	3.8
	4	4.829e-01	0e+00	1e-04	9e-05	0	5	255	491	260	6.5
	5	**5.236e-01**	0e+00	3e-05	5e-05	0	12	260	376	272	3.8
	6	5.235e-01	0e+00	3e-05	8e-05	0	6	213	341	219	4.1
6	1	**4.917e-01**	0e+00	5e-05	4e-05	0	10	223	307	233	3.4
	2	4.326e-01	0e+00	3e-05	8e-05	0	6	217	360	223	4.7
	3	**4.917e-01**	0e+00	5e-05	6e-05	0	5	196	308	201	4.2
	4	4.400e-01	0e+00	1e-04	8e-05	0	5	385	626	390	9.9
	5	**4.917e-01**	0e+00	7e-05	1e-04	0	11	265	362	276	4.2
	6	4.326e-01	0e+00	4e-05	9e-05	0	6	226	356	232	6.2
7	1	4.326e-01	0e+00	8e-05	8e-05	0	11	326	441	337	8.7
	2	**4.505e-01**	0e+00	5e-05	6e-05	0	6	433	609	439	13.3
	3	**4.505e-01**	0e+00	5e-05	7e-05	0	5	665	1030	670	22.1
	4	4.325e-01	0e+00	1e-04	3e-05	0	5	393	666	398	12.0
	5	4.326e-01	0e+00	3e-05	8e-05	0	13	408	548	421	9.0
	6	**4.505e-01**	0e+00	6e-05	7e-05	0	6	512	768	518	15.8
8	1	4.285e-01	0e+00	8e-05	9e-05	0	11	372	474	383	9.2
	2	4.285e-01	0e+00	3e-05	6e-05	0	6	316	444	322	9.3
	3	4.285e-01	0e+00	5e-05	4e-05	0	5	326	476	331	10.0
	4	**4.293e-01**	0e+00	1e-04	8e-05	0	5	959	1721	964	34.4
	5	4.285e-01	0e+00	4e-05	3e-05	0	13	575	729	588	15.6
	6	4.285e-01	0e+00	4e-05	1e-04	0	6	369	519	375	11.2
9	1	**3.949e-01**	0e+00	7e-05	8e-05	0	12	606	770	618	18.7
	2	**3.949e-01**	0e+00	3e-05	9e-05	0	6	362	555	368	11.7
	3	3.903e-01	0e+00	5e-05	5e-05	0	5	358	492	363	11.9
	4	3.948e-01	0e+00	1e-04	8e-05	0	5	1506	2538	1511	58.7
	5	**3.949e-01**	0e+00	4e-05	8e-05	0	14	647	833	661	18.9
	6	**3.949e-01**	0e+00	4e-05	9e-05	0	6	429	644	435	14.3
10	1	**3.793e-01**	0e+00	3e-05	8e-05	0	12	627	781	639	20.4
	2	**3.793e-01**	0e+00	3e-06	5e-05	0	7	816	1313	821	32.1
	3	**3.793e-01**	0e+00	5e-05	4e-05	0	5	924	1427	929	36.9
	4	3.792e-01	0e+00	1e-04	4e-05	0	5	2721	4935	2726	116.4
	5	**3.793e-01**	0e+00	6e-05	7e-05	0	13	692	858	705	22.3
	6	**3.793e-01**	0e+00	5e-05	5e-05	0	6	673	996	679	27.4
20	1	2.745e-01	0e+00	6e-05	8e-05	0	12	1144	1344	1156	69.4
	2	**2.751e-01**	0e+00	3e-05	9e-05	0	6	725	848	731	45.0
	3	2.750e-01	0e+00	5e-05	9e-05	0	5	993	1322	998	64.0
	4	2.704e-01	0e+00	1e-04	8e-05	0	5	1864	2669	1869	128.3
	5	2.748e-01	0e+00	8e-05	9e-05	0	14	1343	1585	1357	88.6
	6	2.744e-01	0e+00	5e-05	8e-05	0	6	798	986	804	52.9
30	1	2.260e-01	0e+00	2e-05	9e-05	0	13	1907	2167	1920	177.7
	2	**2.270e-01**	0e+00	5e-05	1e-04	0	6	1138	1424	1144	106.2
	3	2.246e-01	0e+00	5e-05	8e-05	0	5	1504	1853	1509	140.9
	4	2.231e-01	0e+00	1e-04	9e-05	0	5	1749	2132	1754	163.9
	5	2.251e-01	0e+00	8e-05	6e-05	0	12	2016	2329	2028	184.8
	6	2.260e-01	0e+00	4e-05	9e-05	0	6	1257	1495	1263	122.1

Table 2: Numerical results of applying the six variants of Algorithm 1 with a single penalty parameter to the problem of maximizing the radius of N identical circles packed into a given elliptic container. In this table, $N \in \{5, 6, \ldots, 10, 20, 30\}$.

5.3 Classification problem

In this section, we consider the classification scheme described in [12], which can be modeled as an optimization problem on the set of positive definite matrices. Given a training set of labeled

N	$\phi_{i,\tau}$	r	$\|c\|$	$\|s\|$	$\|\operatorname{grad}\mathbb{L}\|$	SC	k	$\sum_{i=1}^{k} k_i$	#fcnt	#gcnt	Time
40	1	1.975e-01	0e+00	4e-05	9e-05	0	13	2936	3353	2949	398.9
	2	**1.977e-01**	0e+00	3e-05	8e-05	0	6	1712	1971	1718	226.8
	3	1.970e-01	0e+00	5e-05	1e-04	0	5	1651	2000	1656	225.8
	4	1.913e-01	0e+00	1e-03	1e-02	4	4	2178	2691	2182	288.5
	5	1.916e-01	0e+00	6e-05	1e-04	0	13	2292	2649	2305	290.8
	6	1.963e-01	0e+00	6e-05	9e-05	0	6	1704	2004	1710	218.8
50	1	1.766e-01	0e+00	9e-05	9e-05	0	12	2899	3324	2911	471.9
	2	**1.782e-01**	0e+00	3e-05	7e-05	0	6	1859	2176	1865	307.5
	3	1.766e-01	0e+00	5e-05	9e-05	0	5	1688	2003	1693	281.5
	4	1.686e-01	0e+00	1e-03	3e-01	4	4	2206	2701	2210	370.8
	5	1.776e-01	0e+00	3e-04	3e-03	4	13	4414	4954	4427	705.0
	6	1.781e-01	0e+00	6e-05	1e-04	0	6	2276	2651	2282	379.1
60	1	1.619e-01	0e+00	7e-05	9e-05	0	12	4998	5620	5010	1029.7
	2	**1.628e-01**	0e+00	3e-05	1e-04	0	6	2635	3125	2641	549.1
	3	1.619e-01	0e+00	5e-04	4e-03	4	4	2190	2584	2194	462.0
	4	1.522e-01	0e+00	1e-03	1e+00	4	4	2239	2965	2243	482.6
	5	1.626e-01	0e+00	6e-05	1e-04	0	14	5635	6553	5649	1161.5
	6	1.616e-01	0e+00	6e-05	1e-04	0	6	3205	3713	3211	660.6
70	1	1.502e-01	0e+00	9e-05	9e-05	0	11	2917	3296	2928	721.2
	2	1.503e-01	0e+00	3e-05	8e-03	4	6	3198	3782	3204	801.3
	3	1.483e-01	0e+00	5e-04	5e-02	4	4	2352	2723	2356	590.9
	4	1.359e-01	0e+00	1e-03	3e-01	4	4	2458	3383	2462	625.2
	5	**1.511e-01**	0e+00	5e-05	9e-05	0	15	6145	7080	6160	1605.8
	6	1.510e-01	0e+00	7e-05	2e-03	4	6	3768	4420	3774	958.2
80	1	1.409e-01	0e+00	5e-05	9e-05	0	13	4336	5040	4349	1270.5
	2	1.403e-01	0e+00	4e-05	1e-04	0	6	1765	2090	1771	532.9
	3	1.398e-01	0e+00	5e-06	5e-02	4	6	2876	3438	2882	893.5
	4	1.396e-01	0e+00	1e-05	1e-01	4	6	2508	3075	2514	734.4
	5	**1.415e-01**	0e+00	5e-05	8e-05	0	14	5624	6591	5638	1654.3
	6	1.414e-01	0e+00	5e-05	2e-03	4	6	2948	3514	2954	916.4
90	1	1.332e-01	0e+00	3e-05	1e-04	0	13	5728	6683	5741	1947.0
	2	1.333e-01	0e+00	4e-06	8e-05	0	7	3442	4104	3449	1189.1
	3	1.322e-01	0e+00	5e-04	5e-04	4	4	2154	2581	2158	763.6
	4	1.269e-01	0e+00	1e-03	1e-01	4	4	2341	2723	2345	831.8
	5	**1.337e-01**	0e+00	6e-05	9e-05	0	15	6803	7841	6818	2329.4
	6	**1.337e-01**	0e+00	1e-09	1e+00	4	6	2538	3055	2542	866.9
100	1	**1.270e-01**	0e+00	5e-05	1e-04	0	13	5997	6978	6010	2433.6
	2	1.265e-01	0e+00	3e-04	5e-04	4	5	2562	3032	2567	1009.4
	3	1.253e-01	0e+00	5e-04	2e-02	4	4	2198	2572	2202	896.4
	4	1.169e-01	0e+00	1e-03	8e-01	4	4	2459	3180	2463	995.8
	5	1.267e-01	0e+00	3e-04	2e-04	4	13	5878	6725	5891	2246.3
	6	1.259e-01	0e+00	6e-04	1e-03	4	5	2856	3337	2861	1152.2

Table 3: Numerical results of applying the six variants of Algorithm 1 with a single penalty parameter to the problem of maximizing the radius of N identical circles packed into a given elliptic container. In this table, $N \in \{40, 50, \dots, 100\}$.

examples

$$D = \left\{(z_i, w_i), i = 1, \dots, \bar{m},\ z_i \in \mathbb{R}^2, \text{ and } w_i \in \{1, -1\}\right\},$$

we want to find a classifier ellipse $E(A, b) = \{y \in \mathbb{R}^2 \,|\, y^T A y + b^T y = 1\}$ such that $z_i^T A z_i + b^T z_i \le 1$ when $w_i = 1$ and $z_i^T A z_i + b^T z_i \ge 1$ when $w_i = -1$. Since such an ellipse may not exist, we define $I = \{i \in \{1, \dots, \bar{m}\} \,|\, w_i = 1\}$ and $O = \{i \in \{1, \dots, \bar{m}\} \,|\, w_i = -1\}$ and try to minimize the function given by

$$f(A, b) = \frac{1}{\bar{m}} \left[\sum_{i \in I} \max\{0, z_i^T A z_i + b^T z_i - 1\}^2 + \sum_{i \in O} \max\{0, 1 - z_i^T A z_i - b^T z_i\}^2 \right]$$

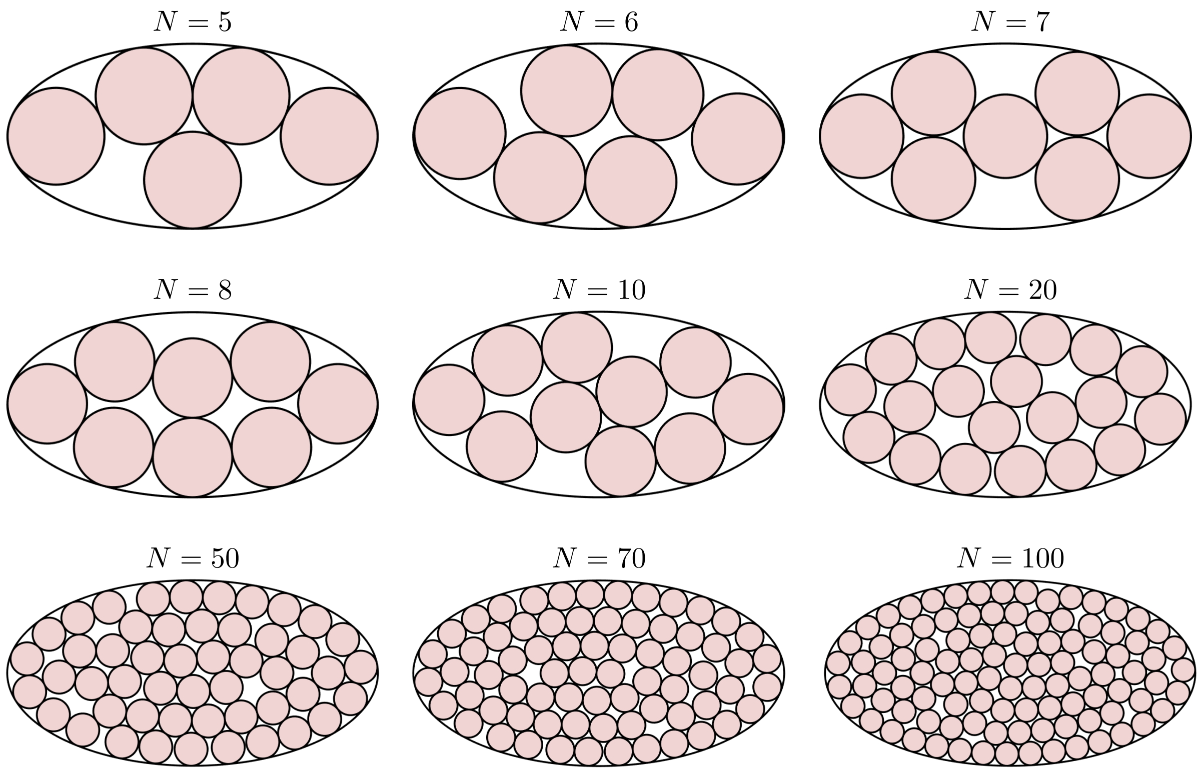


Figure 3: Solutions found for the problem of packing $N \in \{5, 6, 7, 8, 10, 20, 50, 70, 100\}$ identical circles of maximum radius r into an ellipse with semi-axes $(a, b) = (2, 1)$.

provided that A is symmetric and positive definite. To consider a constrained problem, we added a constraint on the center $c = (c_1, c_2)$ of the ellipse to be found. The center corresponds to the solution of the linear system $Ac = -b/2$. Thus,

$$c_1 = \frac{1}{2}\frac{a_{12}\, b_2 - a_{22}\, b_1}{a_{11}\, a_{22} - a_{12}\, a_{21}} \text{ and } c_2 = \frac{1}{2}\frac{a_{21}\, b_1 - a_{11}\, b_2}{a_{11}\, a_{22} - a_{12}\, a_{21}},$$

where a_{ij} with i and j between 1 and 2 are the elements of the matrix A, and b_1, b_2 are the components of the vector b. Constraints are given by $1 \leq c_1 \leq 10$ and $1 \leq c_2 \leq 10$. Therefore, denoting by $\mathrm{Sym}(n)_{++}$ the set of symmetric positive definite real matrices of size n, this problem can be written as

$$\begin{array}{ll} \underset{(A,b)\in \mathrm{Sym}(2)_{++}\times\mathbb{R}^2}{\text{Minimize}} & \frac{1}{\bar{m}}\left[\sum_{i\in I}\max\{0, z_i^T A z_i + b^T z_i - 1\}^2 + \sum_{i\in O}\max\{0, 1 - z_i^T A z_i - b^T z_i\}^2\right] \\ \text{subject to} & c_1 = (a_{12}\, b_2 - a_{22} b_1)/[2(a_{11}\, a_{22} - a_{12}\, a_{21})], \\ & c_2 = (a_{21}\, b_1 - a_{11}\, b_2)/[2(a_{11}\, a_{22} - a_{12}\, a_{21})], \\ & 1 \leq c_1 \leq 10, \\ & 1 \leq c_2 \leq 10. \end{array} \quad (38)$$

Note that c_1 and c_2 are not variables of the problem. They are only used to simplify the representation of the constraints.

Following [12], we considered four instances with $\bar{m} = 10{,}000$ in which the points z_i are randomly generated with uniform distribution in the box $[-10, 10]^2$. In the first instance, the points inside a circle of radius 7 centered on the origin are labeled $w_i = 1$, while the other points are labeled $w_i = -1$. The other three instances correspond to points that get the label 1 if they are inside a square with side 7 centered on the origin, a rectangle with height 7 and width 14 centered on the origin, and a triangle with corners $(-7, 0)$, $(0, -7)$, and $(7, 7)$.

In this problem we considered $\varepsilon_{\text{feas}} = \varepsilon_{\text{opt}} = \varepsilon_{\text{compl}} = 10^{-6}$. We solved the four instances using the six variants of Algorithm 1 with a single penalty parameter. Table 4 shows the results. Figure 4 illustrates some of the solutions found. All variants tackled the problem effectively, producing solutions that met the feasibility criteria and provided high quality objective function values. Although variants 3 and 4 always stopped due to failures of the inner solver (SC equal to 4), they still find objective function values equivalent to the best value found across all tested algorithms in most cases. Variant 5 stopped with SC equal to 4 only in instance 2, where the algorithm almost successfully converged: during earlier iterations, the feasibility and optimality conditions ((35) and (37), respectively) were satisfied within the tolerance 10^{-6}, and the complementarity measure (36) reached the order of 10^{-6}, but the inner solver subsequently failed. Despite this failure, the algorithm successfully identified a solution equivalent to the best solution found. In particular, all instances produced a final solution where the center of the ellipse coincided with $(c_1, c_2) = (1, 1)$ up to three decimal places, while variants 1, 2, 5, and 6 achieved at least five correct decimal places, confirming that the constraints $c_1 \geq 1$ and $c_2 \geq 1$ are active in the solution. This accuracy reflects the symmetry of the problem, where the global solution is expected to have its center at $(1, 1)$.

Inst.	$\phi_{i,\tau}$	f	$\Vert c\Vert$	$\Vert s\Vert$	$\Vert \text{grad}\,\mathbb{L}\Vert$	SC	k	$\sum_{i=1}^{k} k_i$	#fcnt	#gcnt	Time
	1	**3.521e-03**	0e+00	9e-07	8e-07	0	17	415	1191	426	22.7
	2	**3.521e-03**	0e+00	5e-08	6e-07	0	9	286	1001	290	17.7
1	3	**3.521e-03**	0e+00	0e+00	2e-03	4	7	394	2300	392	36.5
	4	**3.521e-03**	0e+00	0e+00	2e-03	4	7	440	2514	438	40.0
	5	**3.521e-03**	0e+00	4e-07	6e-07	0	18	418	1213	430	23.0
	6	**3.521e-03**	0e+00	5e-07	8e-07	0	8	277	938	280	16.8
	1	7.682e-03	0e+00	7e-07	9e-07	0	18	314	1068	327	19.1
	2	**7.681e-03**	0e+00	0e+00	7e-04	4	9	199	806	202	14.1
2	3	7.682e-03	0e+00	0e+00	7e-04	4	5	211	1613	208	24.1
	4	**7.681e-03**	0e+00	0e+00	7e-04	4	7	294	1910	294	29.5
	5	**7.681e-03**	0e+00	1e-07	7e-04	4	20	361	1236	371	22.2
	6	**7.681e-03**	0e+00	6e-07	9e-07	0	8	175	754	180	12.7
	1	**9.499e-03**	0e+00	5e-07	5e-07	0	18	531	1561	541	30.1
	2	**9.499e-03**	0e+00	6e-07	4e-07	0	8	322	973	325	18.9
3	3	**9.499e-03**	0e+00	5e-06	1e-03	4	6	365	2165	365	34.1
	4	**9.499e-03**	0e+00	0e+00	2e-03	4	7	365	2402	361	37.5
	5	**9.499e-03**	0e+00	7e-07	7e-07	0	18	467	1238	478	24.5
	6	**9.499e-03**	0e+00	6e-07	4e-07	0	8	308	979	310	18.2
	1	**1.195e-02**	0e+00	1e-06	9e-07	0	17	392	928	406	18.9
	2	**1.195e-02**	0e+00	5e-07	1e-06	0	8	282	696	287	14.3
4	3	1.196e-02	0e+00	0e+00	4e-03	4	5	268	1555	266	25.0
	4	**1.195e-02**	0e+00	0e+00	4e-03	4	7	340	2005	338	31.7
	5	**1.195e-02**	0e+00	5e-07	8e-07	0	18	435	1136	447	22.2
	6	**1.195e-02**	0e+00	6e-07	1e-06	0	8	190	617	195	11.7

Table 4: Numerical results of applying the six variants of Algorithm 1 with a single penalty parameter to solve the four instances of the classification problem (38).

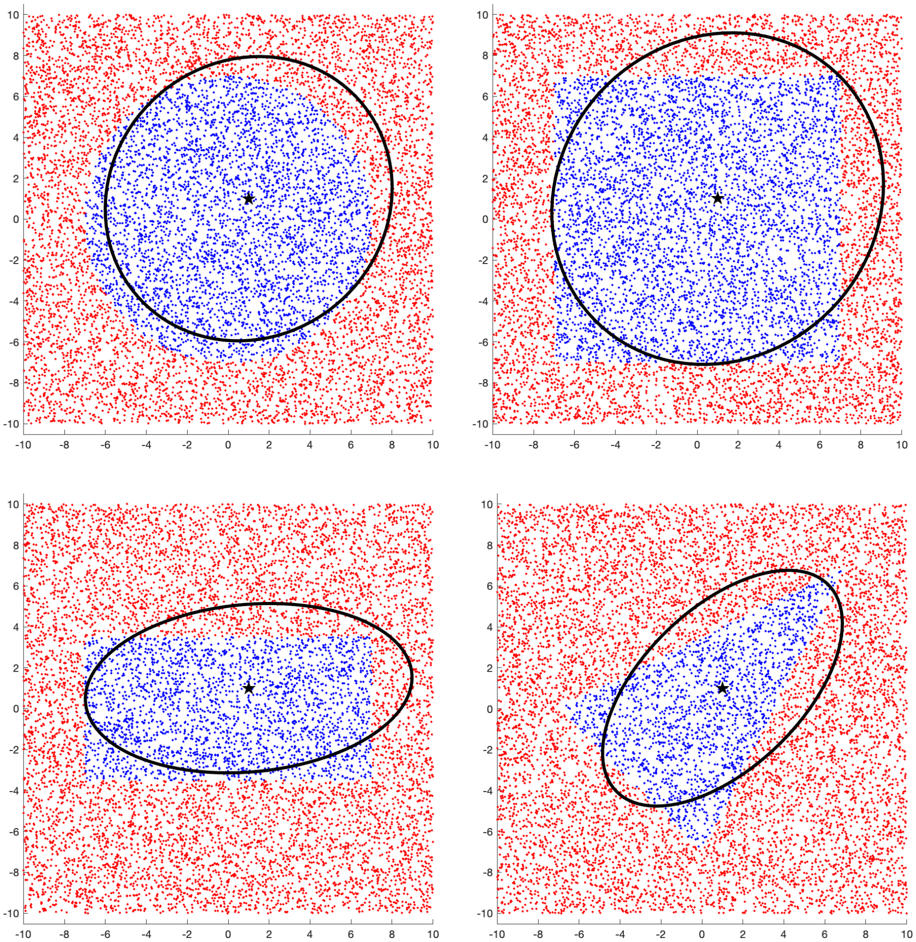


Figure 4: Illustration of the solutions obtained by Algorithm 1 for different smoothing functions in solving the four instances of the classification problem (38). The center of each ellipse is marked with a star.

6 Conclusion

Our study advances the field of CRO by introducing and thoroughly analyzing a novel penalty-type method. Rooted in Riemannian optimization principles and the smoothing of exact penalty functions, this methodology addresses key challenges in constrained optimization on Riemannian manifolds, offering robust and efficient solutions. Through our investigation, we identified improvements over existing approaches. These advancements include greater flexibility in applying smoothing techniques, a more refined strategy for adjusting penalty multipliers, and an analysis under EMFCQ. Our convergence analysis provides global optimality guarantees under the assumption of convergence. Furthermore, assuming the feasibility of limit points, we demonstrated that these points satisfy the AKKT conditions. Additionally, when AKKT points are combined with either the Relaxed-CPLD or the CRSC conditions—both recently introduced in [3]—we show that the limit points indeed satisfy the KKT conditions. Moreover, we conducted an analysis under the assumptions that the penalty sequences are bounded and that the smoothing parameter sequence associated with the inequality constraints is unbounded, yielding similar positive results. By leveraging these methodologies, our research contributes valuable insights and strategies to the CRO community, paving the way for future advancements. As the field of CRO continues to evolve, we anticipate that penalty-type methods will remain essential for addressing practical constrained problems within the Riemannian framework. Our work lays the groundwork for further research and applications, fostering continued innovation and effectiveness in this domain.

References

[1] P.-A. Absil, C. G. Baker, and K. A. Gallivan. Trust-region methods on Riemannian manifolds. Foundations of Computational Mathematics, 7(3):303–330, 2007.

[2] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. Mathematical Programming, 111(1-2):5–32, 2006.

[3] R. Andreani, K. R. Couto, O. P. Ferreira, and G. Haeser. Constraint qualifications and strong global convergence properties of an augmented Lagrangian method on riemannian manifolds. SIAM Journal on Optimization, 34(2):1799–1825, 2024.

[4] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva. Two new weak constraint qualifications and applications. SIAM Journal on Optimization, 22(3):1109–1135, 2012.

[5] R. Bergmann and R. Herzog. Intrinsic formulation of KKT conditions and constraint qualifications on smooth manifolds. SIAM Journal on Optimization, 29(4):2423–2444, 2019.

[6] D. P. Bertsekas. Nondifferentiable optimization via approximation. In M. L. Balinski and P. Wolfe, editors, Nondifferentiable Optimization, volume 3 of Mathematical Programming Studies, pages 1–25. Springer, Berlin, Heidelberg, 1975.

[7] D. P. Bertsekas. Constrained optimization and Lagrange multiplier methods. Academic Press, New York, 1982.

[8] E. G. Birgin, L. H. Bustamante, H. F. Callisaya, and J. M. Martínez. Packing circles within ellipses. International Transactions in Operational Research, 20(3):365–389, 2013.

[9] E. G. Birgin, R. A. Castillo, and J. M. Martínez. Numerical comparison of augmented Lagrangian algorithms for nonconvex problems. Computational Optimization and Applications, 31(1):31–55, 2005.

[10] E. G. Birgin, G. Haeser, N. Maculan, and L. M. Ramirez. On the global convergence of a general class of augmented Lagrangian methods. Technical Report MCDO160524, University of São Paulo, São Paulo, SP, 2024.

[11] E. G. Birgin and J. M. Martínez. Practical Augmented Lagrangian Methods for Constrained Optimization. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014.

[12] E. G. Birgin, J. M. Martínez, and M. Raydan. Spectral projected gradient methods: Review and perspectives. Journal of Statistical Software, 60(3), 2014.

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

[14] C. Chen and O. L. Mangasarian. Smoothing methods for convex inequalities and linear complementarity problems. Mathematical Programming, 71(1):51–69, 1995.

[15] K. Deng and Z. Peng. A manifold inexact augmented Lagrangian method for nonsmooth optimization on Riemannian submanifolds in Euclidean space. IMA Journal of Numerical Analysis, 43(3):1653–1684, 2023.

[16] G. Di Pillo and L. Grippo. Exact penalty functions in constrained optimization. SIAM Journal on Control and Optimization, 27(6):1333–1360, 1989.

[17] O. P. Ferreira, M. S. Louzeiro, and L. F. Prudente. Gradient method for optimization on Riemannian manifolds with lower bounded curvature. SIAM Journal on Optimization, 29(4):2517–2541, 2019.

[18] A. V. Fiacco and G. P. McCormick. Nonlinear programming: Sequential unconstrained minimization techniques. John Wiley & Sons, Inc., New York-London-Sydney, 1968.

[19] C. C. Gonzaga and R. A. Castillo. A nonlinear programming algorithm based on non-coercive penalty functions. Mathematical Programming, 96(1):87–101, 2003.

[20] F. Goyens, A. Eftekhari, and N. Boumal. Computing second-order points under equality constraints: revisiting Fletcher's augmented Lagrangian. Journal of Optimization Theory and Applications, 201(3):1198–1228, 2024.

[21] C. Grossmann. Smoothing techniques for exact penalty methods. In A panorama of mathematics: pure and applied. Conference mathematics and its applications, Kuwait University, Safat, Kuwait, November 14–17, 2014, pages 249–265. Providence, RI: American Mathematical Society (AMS), 2016.

[22] W. Huang, P.-A. Absil, and K. A. Gallivan. A Riemannian symmetric rank-one trust-region method. Mathematical Programming, 150(2):179–216, 2015.

[23] W. Huang, K. A. Gallivan, and P.-A. Absil. A Broyden class of quasi-Newton methods for Riemannian optimization. SIAM Journal on Optimization, 25(3):1660–1685, 2015.

[24] W. Huang, K. A. Gallivan, and P.-A. Absil. A Broyden class of quasi-Newton methods for Riemannian optimization. SIAM Journal on Optimization, 25(3):1660–1685, 2015.

[25] W. Huang and K. Wei. Riemannian proximal gradient methods. Mathematical Programming, 194(1-2):371–413, 2022.

[26] B. Jiang and Y.-F. Liu. A Riemannian exponential augmented Lagrangian method for computing the projection robust Wasserstein distance. Advances in Neural Information Processing Systems, 36, 2024.

[27] B. Jiang, X. Meng, Z. Wen, and X. Chen. An exact penalty approach for optimization with nonnegative orthogonality constraints. Mathematical Programming, 198(1):855–897, 2023.

[28] C. Kanzow and D. Steck. An example comparing the standard and safeguarded augmented Lagrangian methods. Operations Research Letters, 45(6):598–603, 2017.

[29] A. A. Kaplan. Convex programming algorithms using smoothing of exact penalty functions. Siberian Mathematical Journal, 23:491–500, 1983.

[30] Z. Lai and A. Yoshise. Riemannian interior point methods for constrained optimization on manifolds. Journal of Optimization Theory and Applications, 201(1):433–469, 2024.

[31] S. Lang. Differential and Riemannian manifolds, volume 160 of Graduate Texts in Mathematics. Springer-Verlag, New York, third edition, 1995.

[32] B. Liu. On smoothing exact penalty functions for nonlinear constrained optimization problems. Journal of Applied Mathematics and Computing, 30(1-2):259–270, 2009.

[33] C. Liu and N. Boumal. Simple algorithms for optimization on Riemannian manifolds with constraints. Applied Mathematics and Optimization, 82(3):949–981, 2020.

[34] G. Liuzzi and S. Lucidi. A derivative-free algorithm for inequality constrained nonlinear programming via smoothing of an ℓ_∞ penalty function. SIAM Journal on Optimization, 20(1):1–29, 2009.

[35] C. López and J. Beasley. A heuristic for the circle packing problem with a variety of containers. European Journal of Operational Reserach, 214(3):512–525, 2011.

[36] L. Mallma Ramirez, N. Maculan, A. E. Xavier, and V. L. Xavier. Dislocation hyperbolic augmented Lagrangian algorithm for nonconvex optimization. RAIRO - Operations Research, 57(5):2941–2950, 2023.

[37] L. Mallma Ramirez, N. Maculan, A. E. Xavier, and V. L. Xavier. Dislocation hyperbolic augmented Lagrangian algorithm in convex programming. An International Journal of Optimization and Control: Theories & Applications, 14(2):147–155, 2024.

[38] A. Montanari and E. Richard. Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. IEEE Transactions on Information Theory, 62(3):1458–1484, 2016.

[39] J. Nocedal and S. J. Wright. Numerical Optimization. Springer, New York, NY, 1e edition, 1999.

[40] M. Obara, T. Okuno, and A. Takeda. Sequential quadratic optimization for nonlinear optimization problems on Riemannian manifolds. SIAM Journal on Optimization, 32(2):822–853, 2022.

[41] M. Ç. Pinar and S. A. Zenios. On smoothing exact penalty functions for convex constrained optimization. SIAM Journal on Optimization, 4(3):486–511, 1994.

[42] B. Saheya, C.-H. Yu, and J.-S. Chen. Numerical comparisons based on four smoothing functions for absolute value equation. Journal of Applied Mathematics and Computing, 56(1-2):131–149, 2018.

[43] A. Sahiner, G. Kapusuz, and N. Yilmaz. A new smoothing approach to exact penalty functions for inequality constrained optimization problems. Numerical Algebra, Control and Optimization, 6(2):161–173, 2016.

[44] T. Sakai. Riemannian geometry, volume 149 of Translations of Mathematical Monographs. American Mathematical Society, Providence, RI, 1996.

[45] H. Schichl and A. Neumaier. Transposition theorems and qualification-free optimality conditions. SIAM Journal on Optimization, 17(4):1035–1055, 2006.

[46] W. Si, P.-A. Absil, W. Huang, R. Jiang, and S. Vary. A Riemannian proximal Newton method. SIAM Journal on Optimization, 34(1):654–681, 2024.

[47] L. W. Tu. An introduction to manifolds. Universitext. Springer-Verlag New York, second edition, 2011.

[48] C. Udrişte. Convex functions and optimization methods on Riemannian manifolds, volume 297 of Mathematics and its Applications. Kluwer Academic Publishers Group, Dordrecht, 1994.

[49] G. O. Wesolowsky and R. F. Love. A nonlinear approximation method for solving a generalized rectangular distance Weber problem. Management Science, 18(11):656–663, 1972.

[50] A. E. Xavier. *Penalização hiperbólica: Um novo método para resolução de problemas de otimização.* Master's thesis, COPPE-UFRJ, Rio de Janeiro, RJ, 1982.

[51] A. E. Xavier. Penalização hiperbólica. PhD thesis, COPPE-UFRJ, Rio de Janeiro, RJ, 1992.

[52] A. E. Xavier. Hyperbolic penalty: a new method for nonlinear programming with inequalities. International Transactions in Operational Research, 8(6):659–671, 2001.

[53] Y. Yamakawa and H. Sato. Sequential optimality conditions for nonlinear optimization on Riemannian manifolds and a globally convergent augmented Lagrangian method. Computational Optimization and Applications, 81(2):397–421, 2022.

[54] W. H. Yang, L.-H. Zhang, and R. Song. Optimality conditions for the nonlinear programming problems on Riemannian manifolds. Pacific Journal of Optimization, 10(2):415–434, 2014.

[55] W. I. Zangwill. Non-linear programming via penalty functions. Management Science, 13(5):344–358, 1967.

[56] W. Zhao, R. Wang, and D. Song. Global convergence of a class new smooth penalty algorithm for constrained optimization problem. Journal of Applied Mathematics and Computing, 69(5):3987–3997, 2023.