Global convergence of an augmented Lagrangian method for nonlinear programming via Riemannian optimization

Roberto Andreani * Kelvin R. Couto†§ Orizon P. Ferreira ‡ Gabriel Haeser §
Leandro F. Prudente ‡

September 9, 2024

Abstract

Considering a standard nonlinear programming problem, one may view a subset of the equality constraints as an embedded Riemannian manifold. In this paper we investigate the differences between the Euclidean and the Riemannian approach for this problem. It is well known that the linear independence constraint qualification for both approaches are equivalent. However, when considering recently introduced constant rank constraint qualifications, the Riemannian approach provides a weaker condition as the rank of the gradients must remain constant only inside the manifold, while the Euclidean approach requires constant rank properties inside a full-dimensional neighborhood of the ambient space. Therefore by employing a Riemannian augmented Lagrangian method to a standard nonlinear programming problem we are able to obtain standard global convergence to a Karush/Kuhn-Tucker point under a new weaker constant rank condition that considers only lower dimensional neighborhoods. In this way we illustrate how the Riemannian perspective can provide new and stronger results to classical problems traditionally addressed through Euclidean theory. We also investigate the two alternative augmented Lagrangian algorithms in a comprehensive computational study, where we show some classes of problems where the Riemannian approach is much more robust in attaining better quality solutions.

Keywords: Safeguarded augmented Lagrangian method, constrained nonlinear programming, constraint qualifications, embedded submanifold.

AMS subject classification: 49J52, 49M15, 65H10, 90C30.

1 Introduction

This paper advances the comprehension of findings concerning constraint qualifications and convergence properties inherent in an augmented Lagrangian method designed for Riemannian manifolds, as initially outlined in [3]. It aims to demonstrate how the theoretical framework based on Riemannian concepts can introduce innovative perspectives and viable alternative solutions to problems traditionally addressed through Euclidean theory. Additionally, this study highlights the capacity of modern Riemannian geometry concepts to enrich conventional Euclidean theory, thereby refining theoretical paradigms within Euclidean space. To achieve this objective, we introduce novel constraint qualifications and explore the applicability of Riemannian augmented Lagrangian methods to a specific category of constrained nonlinear programming problems characterized by both equality and inequality constraints, with the equality constraints further categorized into two

*Department of Applied Mathematics, University of Campinas, Campinas-SP, Brazil. Email: andreani@ime.unicamp.br
†Federal Institute of Goiás, Goiânia-GO, Brazil. Email: kelvin.couto@ifg.edu.br
‡Institute of Mathematics and Statistics, Federal University of Goiás. Goiânia-GO, Brazil. Emails: orizon@ufg.br, lfprudente@ufg.br
§Department of Applied Mathematics, University of São Paulo, São Paulo-SP, Brazil. Email: ghaeser@ime.usp.br

distinct types. The constrained optimization problem under consideration is formally defined as follows:

$$\underset{q\in\mathbb{R}^n}{\text{Minimize}}\, f(q), \qquad \text{subject to} \quad h(q)=0,\ H(q)=0,\ G(q)\leq 0, \tag{1}$$

where the functions $f\colon \mathbb{R}^n \to \mathbb{R}$, $h := (h_1,\dots,h_t)\colon \mathbb{R}^n \to \mathbb{R}^t$, $H := (H_1,\dots,H_s)\colon \mathbb{R}^n \to \mathbb{R}^s$ and $G := (G_1,\dots,G_m)\colon \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable. A standard approach to solving problem (1) is through the *augmented Lagrangian algorithm*, which involves the iterative unconstrained minimization of the *standard Powell-Hestenes-Rockafellar augmented Lagrangian function* given by

$$L_\rho(q,\eta,\lambda,\mu) := f(q) + \frac{\rho}{2}\left(\left\|h(q)+\frac{\eta}{\rho}\right\|_2^2 + \left\|H(q)+\frac{\lambda}{\rho}\right\|_2^2 + \left\|\left[G(q)+\frac{\mu}{\rho}\right]_+\right\|_2^2\right), \tag{2}$$

where $\rho > 0$ is a fixed penalty parameter, and safeguarded Lagrange multipliers $\eta := (\eta_1,\dots,\eta_t) \in \mathbb{R}^t$, $\lambda := (\lambda_1,\dots,\lambda_s) \in \mathbb{R}^s$ and $\mu := (\mu_1,\dots,\mu_m) \in \mathbb{R}^m_+$ are estimated in each (outer) iteration. Here $[u]_+$ stands for the projection of $u \in \mathbb{R}^m$ onto the non-negative orthant $\mathbb{R}^m_+$. An alternative approach to addressing constrained optimization problems in the format (1), previously utilized in [2, 16], involves considering the so-called *lower-level constraints*:

$$\mathbb{M} := \{q \in \mathbb{R}^n \mid h(q) = 0\}. \tag{3}$$

Then, a *constrained augmented Lagrangian method* is employed to solve the problem, which involves iteratively minimizing the *partial Powell-Hestenes-Rockafellar augmented Lagrangian function*

$$\mathbb{L}_\rho(q,\lambda,\mu) := f(q) + \frac{\rho}{2}\left(\left\|H(q)+\frac{\lambda}{\rho}\right\|_2^2 + \left\|\left[G(q)+\frac{\mu}{\rho}\right]_+\right\|_2^2\right), \tag{4}$$

subject to the lower-level set $\mathbb{M}$. The idea behind this division arises from the strategic advantage that augmented Lagrangian methods offer in solving nonlinear programming problems. By partitioning the equality constraints, a level of flexibility is introduced, allowing for the prioritization of constraints based on their relevance to the current problem or the ease with which they can be managed. Consequently, this approach within the augmented Lagrangian framework enables the penalization of a specific set of constraints, potentially the most demanding ones, while ensuring a non-penalized status for the lower-level constraints that we aim to prioritize. As a result, subproblems are formulated as minimizing $\mathbb{L}_\rho(\cdot,\eta,\mu)$ subject to $\mathbb{M}$. For instance, if the goal is to maintain feasibility for a set $\mathbb{M}$, these subproblems can be addressed using methods that keep the (inner) iterates in $\mathbb{M}$. In this manner, the sequence generated by these constrained augmented Lagrangian methods remains feasible with respect to $\mathbb{M}$. Additionally, a notable aspect of this approach applies to scenarios where the objective function and/or constraints are defined solely at points belonging to $\mathbb{M}$, rendering the equality constraint $h(q) = 0$ ineligible for penalization.

Understanding optimality conditions and constraint qualifications is crucial in the study of nonlinear programing problems. The *Karush/Kuhn-Tucker (KKT) conditions* play a pivotal role in identifying optimal solutions, while *constraint qualifications* ensure that these solutions satisfy the KKT conditions. Over time, modern nonlinear programming theory has witnessed the evolution of KKT conditions and the emergence of new constraint qualifications. This evolution has significantly broadened the theoretical framework of nonlinear optimization, allowing the application of augmented Lagrangian methods across a wide range of problem classes. Notably, constraint qualifications such as the *constant rank constraint qualification* (CRCQ) [27], *constant positive linear dependence condition* (CPLD) [34], *relaxed-CRCQ* (RCRCQ) [33], *relaxed-CPLD* (RCPLD) [7], *constant rank of the subspace component* (CRSC) [8] and *quasinormality constraint qualification* (QN) [25], have been introduced to enhance the understanding and application of optimization techniques. Moreover, the introduction of sequential optimality conditions, such as the *approximate Karush-Kuhn-Tucker* (AKKT) [5] and *positive-AKKT* (PAKKT) [4], has provided additional flexibility by relaxing the KKT conditions. These developments represent important advancements in the field and are essential for advancing the state-of-the-art in nonlinear programming research.

For example, within the framework of safeguarded augmented Lagrangian methods, their strength lies in their ability to generate PAKKT sequences for constrained nonlinear programming problems. Under any of the aforementioned constraint qualifications, this ensures that all limit points of such sequences adhere to the KKT conditions, a topic extensively explored in the literature (see, for instance, [4, 6, 8]). We use the adjective *strict* to distinguish the constraint qualifications with the aforementioned sequential property, see [9].

To address nonlinear optimization problems in the format (1), we introduce new strict constraint qualifications termed *lower strict constraint qualifications* (Lower-SCQs) to take into account the lower-level approach of considering augmented Lagrangian subproblems constrained to the lower-level set $\mathbb{M}$. These constraint qualifications serve as less restrictive counterparts to CRCQ, CPLD, RCRCQ, RCPLD, CRSC, and QN. In this new scenario, it is no longer guaranteed that the limit points of the sequence generated by classic augmented Lagrangian methods satisfy the KKT conditions. Therefore, by considering the equality constraints (3) as a Riemannian manifold, we employ tools from Riemannian Geometry to establish a connection between the Lower-SCQs and their Riemannian counterparts recently introduced in [3], referred to as *Riemannian strict constraint qualifications* (Riemannian-SCQs). Furthermore, by introducing the concept of *lower approximate Karush-Kuhn-Tucker* (Lower-AKKT) and *lower positive approximate-KKT* (Lower-PAKKT) for problem (1), which serve as counterparts to AKKT and PAKKT, respectively, we show that the *Riemannian adaptation of the classic safeguarded augmented Lagrangian algorithm*, an intrinsic algorithm presented in [35], is able to produce Lower-PAKKT sequences that are feasible for $\mathbb{M}$. Moreover, under any Lower-SCQ we show that all limit points of this sequence satisfy the KKT conditions for problem (1). Additionally, as we establish a link between these Lower-SCQs and the Riemannian-SCQs, we highlight the robustness of the theory within Riemannian manifolds. This robustness offers valuable support for the convergence analysis of algorithms in nonlinear programming, especially when compared to those formulated in Euclidean spaces. This underscores that there are various subtle aspects concerning constraint qualifications in Riemannian manifold settings that would be overlooked if the problem were solely addressed with the existing Euclidean theory. In this sense, as mentioned earlier, this paper serves as a complement to aid in understanding the range of applications of the theory presented in [3].

The paper is structured as follows: Subsection 1.1 introduces terminology, notations, and basic results on Euclidean space and calculus on embedded submanifolds. Section 2 revisits concepts and results in nonlinear optimization in Euclidean spaces and Riemannian manifolds. Section 3 presents new strict constraint qualifications (SCQs) for problem (1), including Lower-SCQs such as Lower-CRCQ, Lower-CPLD, Lower-RCRCQ, Lower-RCPLD, Lower-CRSC, and Lower-QN. It also introduces the new sequential optimality conditions Lower-AKKT and Lower-PAKKT. Section 4 establishes connections between Lower-SCQs and Riemannian-SCQs, demonstrating that under any Lower-SCQ, limit points of the constrained augmented Lagrangian algorithm satisfy the KKT conditions for problem (1). Section 5 presents numerical experiments, and Section 6 offers concluding remarks.

1.1 Notations, terminology and basics results

The set of all $m \times n$ matrices with real entries is denoted by $\mathbb{R}^{m\times n}$. For $M \in \mathbb{R}^{m\times n}$, the matrix $M^\top \in \mathbb{R}^{n\times m}$ is the *transpose* of M. Let $\mathbb{R}^m \equiv \mathbb{R}^{m\times 1}$ be the m-dimensional Euclidean space with the norm denoted by $\|\cdot\|_2$. We denote the infinity norm in $\mathbb{R}^m$ by $\|\cdot\|_\infty$. The *open* and *closed balls* of radius $r > 0$ in $\mathbb{R}^m$, centered at p, are respectively defined by $B_r(p) := \{q \in \mathbb{R}^m \mid \|p-q\|_2 < r\}$ and $B_r[p] := \{q \in \mathbb{R}^m \mid \|p-q\|_2 \le r\}$. For all $p, q \in \mathbb{R}^m$, $\min\{p,q\} \in \mathbb{R}^m$ is the component-wise minimum of p and q. We denote by $[q]_+$ the Euclidean projection of q onto the non-negative orthant $\mathbb{R}^m_+$. The subspace spanned by a set $\mathcal{C} \subset R^m$ is denoted by $\mathrm{Span}(\mathcal{C})$. For a given subspace $V \subset R^m$, its *orthogonal subspace* is defined by $V^\perp := \left\{z \in \mathbb{R}^m \mid v^\top z = 0,\ \forall v \in V\right\}$ and the *Euclidean projection operator* onto $V^\perp$ is denoted by $\mathrm{Proj}_{V^\perp}$.

Definition 1. *Let $V = \{v_1, \ldots, v_s\}$ and $W = \{w_1, \ldots, w_m\}$ be two finite multisets on $\mathbb{R}^n$. The* *pair (V, W) is said to be positive-linearly dependent if there exist $\alpha = (\alpha_1, \dots, \alpha_s) \in \mathbb{R}^s$ and $\beta = (\beta_1, \dots, \beta_m) \in \mathbb{R}^m_+$ such that $(\alpha, \beta) \neq 0$ and $\sum_{i=1}^s \alpha_i v_i + \sum_{j=1}^m \beta_j w_j = 0$. Otherwise, (V, W) is said to be positive-linearly independent. When clear, we refer to $V \cup W$ instead of (V, W).*

We now introduce two lemmas that are essential in Section 4 to establish a connection between the Lower-SCQs and their Riemannian counterparts. The proofs are straightforward from standard linear algebra arguments.

Lemma 1. *Let $\mathcal{C}_1 := \{v_i \in \mathbb{R}^n \mid i = 1, \dots, t\}$, $\mathcal{C}_2 := \{w_i \in \mathbb{R}^n \mid i = 1, \dots, s\}$, and $\mathcal{C}_3 := \{u_i \in \mathbb{R}^n \mid i = 1, \dots, m\}$. Let $V := \mathrm{Span}(\mathcal{C}_1)$ and $V^\perp$ be its orthogonal subspace. Define $\mathrm{Proj}_{V^\perp} \mathcal{C}_2 := \{\mathrm{Proj}_{V^\perp} w_i \mid i = 1, \dots, s\}$ and $\mathrm{Proj}_{V^\perp} \mathcal{C}_3 := \{\mathrm{Proj}_{V^\perp} u_i \mid i = 1, \dots, m\}$. Assume that $\mathcal{C}_1$ is linearly independent. Then, the following statements are equivalent:*

(i) The set $\mathcal{C} := (\mathcal{C}_1 \cup \mathcal{C}_2) \cup \mathcal{C}_3$ is linearly independent (respectively, positive-linearly independent);

(ii) The set $\mathcal{P} := \mathrm{Proj}_{V^\perp} \mathcal{C}_2 \cup \mathrm{Proj}_{V^\perp} \mathcal{C}_3$ is linearly independent (respectively, positive-linearly independent).

Lemma 2. *Let $\mathcal{C}_1 := \{v_i \in \mathbb{R}^n \mid i = 1, \dots, t\}$, $\mathcal{C}_2 := \{w_i \in \mathbb{R}^n \mid i = 1, \dots, s\}$, $\mathcal{K} \subset \{1, \dots, s\}$ and $\mathcal{C}_\mathcal{K} := \{w_i \in \mathbb{R}^n \mid i \in \mathcal{K}\}$. Let $V := Span(\mathcal{C}_1)$ and $V^\perp$ be its orthogonal subspace. Define $\mathrm{Proj}_{V^\perp} \mathcal{C}_\mathcal{K} := \{\mathrm{Proj}_{V^\perp} w_i \mid i \in \mathcal{K}\}$ and $\mathrm{Proj}_{V^\perp} \mathcal{C}_2 := \{\mathrm{Proj}_{V^\perp} w_i \mid i = 1, \dots, s\}$. Assume that $\mathcal{C}_1$ is linearly independent. Then, the following statements are equivalent:*

(i) The set $\mathcal{C}_1 \cup \mathcal{C}_\mathcal{K}$ is a basis of $\mathrm{Span}(\mathcal{C}_1 \cup \mathcal{C}_2)$;

(ii) $\mathrm{Proj}_{V^\perp} \mathcal{C}_\mathcal{K}$ is a basis of $\mathrm{Span}(\mathrm{Proj}_{V^\perp} \mathcal{C}_2)$.

Since $h = (h_1, \dots, h_t) \colon \mathbb{R}^n \to \mathbb{R}^t$ is continuously differentiable on $\mathbb{R}^n$, by *assuming that the set $\{h_i'(q) \mid i = 1, \dots, t\}$ is linearly independent for all $q \in \mathbb{R}^n$, we conclude that the set (3) is an embedded submanifold of $\mathbb{R}^n$ of dimension $n - t$.* The *open* and *closed balls* of radius $r > 0$ in $\mathbb{M}$, centered at p, are respectively defined by $\mathbb{B}_r(p) := \{q \in \mathbb{M} \mid d(p, q) < r\}$ and $\mathbb{B}_r[p] := \{q \in \mathbb{M} \mid d(p, q) \leq r\}$, where $d(\cdot, \cdot)$ *is the Riemannian distance* associated with the induced metric from $\mathbb{R}^n$. The *tangent plane at* $q \in \mathbb{M}$ is given by

$$T_q\mathbb{M} := \{v \in \mathbb{R}^n \mid h'(q)v = 0\} = \Big\{v \in \mathbb{R}^n \mid h_i'(q)^\top v = 0,\ i = 1, \dots t\Big\}. \tag{5}$$

To simplify the notation we also denote the metric in $T_q\mathbb{M}$ by $\| \cdot \|$. It follows from (5) that

$$T_q\mathbb{M} := \mathrm{Ker}\ h'(q), \qquad T_q\mathbb{M}^\perp = \mathrm{Im}\ h'(q)^T, \qquad \mathbb{R}^n = T_q\mathbb{M} \oplus T_q\mathbb{M}^\perp. \tag{6}$$

Therefore, (5) and the second equality in (6) imply that

$$T_q\mathbb{M}^\perp = \Big\{h'(q)^\top \eta = \sum_{i=1}^t \eta_i h_i'(q) \mid \eta = (\eta_1, \dots, \eta_t) \in \mathbb{R}^t\Big\}. \tag{7}$$

For a given $q \in \mathbb{M}$, it is well known that the *projection operator* $\mathrm{Proj}_q : \mathbb{R}^n \to T_q\mathbb{M}$ is given by

$$\mathrm{Proj}_q v = \Big(I - h'(q)^\top \Big(h'(q)h'(q)^\top\Big)^{-1} h'(q)\Big)v, \tag{8}$$

see, for example, [31, p. 377]. Hence, by using (8), the *intrinsic gradient* of a differentiable function $\varphi : \mathbb{M} \to \mathbb{R}$ is given by

$$\mathrm{grad}\, \varphi(q) = \mathrm{Proj}_q \varphi'(q). \tag{9}$$

We will need the following lemma, which is easily proved:

Lemma 3. *Let $X_1, \dots, X_\ell$ be continuous vector fields on a Riemannian manifold $\mathbb{M}$. Let $p \in \mathbb{M}$ and assume that $\{X_1(p), \dots, X_\ell(p)\}$ are linearly independent on $T_p\mathbb{M}$. Then, there exists $\epsilon > 0$ such that $\{X_1(q), \dots, X_\ell(q)\}$ are also linearly independent on $T_q\mathbb{M}$, for all $q \in \mathbb{B}_\epsilon(p)$.*

We conclude this section by noting that the subspace in $T_q\mathbb{M}$ spanned by a set $\mathbb{C} \subset T_q\mathbb{M}$ will also be denoted by $\mathrm{Span}(\mathbb{C})$.

2 Preliminaries

This section defines essential notations and concepts in Euclidean and Riemannian geometry, reviews the basics for addressing the Euclidean problem (1), and uses a submanifold concept to rewrite it as an intrinsic nonlinear optimization problem, yielding new results.

2.1 Nonlinear optimization problems on Euclidean space

The *feasible set* $\Omega \subset \mathbb{R}^n$ of problem (1) and the *set of indices of active inequality constraints* at $p \in \Omega$, denoted by $\mathcal{A}(p)$, are defined respectively as follows:

$$\Omega := \{q \in \mathbb{R}^n \mid h(q) = 0, H(q) = 0, G(q) \leq 0\}, \qquad \mathcal{A}(p) := \{i \in \{1, \dots, m\} \mid G_i(p) = 0\}. \tag{10}$$

It is easy to see that Ω is closed. We say that the *Karush/Kuhn-Tucker* (KKT) conditions are satisfied at $p \in \Omega$ when there exist Lagrange multipliers $(\eta, \lambda, \mu) \in \mathbb{R}^t \times \mathbb{R}^s \times \mathbb{R}^m_+$ such that the following two conditions hold:

(i) $L'(p, \eta, \lambda, \mu) = 0$, *(ii)* $\mu_i = 0$, for all $i \notin \mathcal{A}(p)$,

where $L(\cdot, \eta, \lambda, \mu) \colon \mathbb{R}^n \to \mathbb{R}$ is the *Lagrangian function* associated with problem (1), defined by

$$L(q, \eta, \lambda, \mu) := f(q) + \sum_{i=1}^{t} \eta_i h_i(q) + \sum_{i=1}^{s} \lambda_i H_i(q) + \sum_{i=1}^{m} \mu_i G_i(q),$$

and $L'(q, \eta, \lambda, \mu)$ is its *gradient*. For $p \in \Omega$, the *linearized cone* $\mathcal{L}(p)$ associated with Ω at p is defined by

$$\mathcal{L}(p) := \big\{v \in \mathbb{R}^n \mid h_i'(p)^\top v = 0,\ i = 1, \dots, t;\ H_i'(p)^\top v = 0,\ i = 1, \dots, s;\ G_j'(p)^\top v \leq 0,\ j \in \mathcal{A}(p)\big\},$$

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

$$\mathcal{L}(p)^\circ := \Big\{v \in \mathbb{R}^n \mid v = \sum_{i=1}^{t} \eta_i h_i'(p) + \sum_{i=1}^{s} \lambda_i H_i'(p) + \sum_{j=1}^{m} \mu_j G_j'(p),\ \mu_j \geq 0,\ \eta_i, \lambda_i \in \mathbb{R}\Big\}. \tag{11}$$

In the following, for the sake of conciseness, we introduce some notations. Define

$$\mathcal{T} := \{1, \dots, t\}, \qquad \mathcal{S} := \{1, \dots, s\}, \tag{12}$$

and consider $\bar{\mathcal{T}} \subseteq \mathcal{T}$, $\mathcal{I} \subseteq \mathcal{S}$, and $\mathcal{J} \subseteq \mathcal{A}(p)$. For a given $q \in \Omega$, we define the following sets of vectors:

$$[h'_{\bar{\mathcal{T}}}, H'_{\mathcal{I}}, G'_{\mathcal{J}}](q) := \big(\{h_i'(q) \mid i \in \bar{\mathcal{T}}\} \cup \{H_i'(q) \mid i \in \mathcal{I}\}\big) \cup \{G_i'(q) \mid i \in \mathcal{J}\}. \tag{13}$$

If one of the sets $\bar{\mathcal{T}}$, $\mathcal{I}$, or $\mathcal{J}$ is empty, then the corresponding set will not appear in (13). For instance, for $\bar{\mathcal{T}} = \varnothing$, the set in (13) will be denoted by $[H'_{\mathcal{I}}, G'_{\mathcal{J}}](q) := \{H_i'(q) \mid i \in \mathcal{I}\} \cup \{G_i'(q) \mid i \in \mathcal{J}\}$. In addition, for sake of simplicity, we set $h' := h'_{\mathcal{T}}$, $H' := H'_{\mathcal{S}}$, and $G' := G'_{\mathcal{A}(p)}$. Two constraint qualifications that will be used later are stated below.

Definition 2. *A point $p \in \Omega$ is said to satisfy the linear independence constraint qualification (LICQ) if the set $[h', H', G'_{\mathcal{A}(p)}](p)$ is linearly independent. It satisfies the Mangasarian-Fromovitz constraint qualification (MFCQ) if the set $[h', H', G'_{\mathcal{A}(p)}](p)$ is positive-linearly independent.*

We end this section by recalling the *(Euclidean) safeguarded augmented Lagrangian algorithm* for solving problem (1), which uses the standard Powell-Hestenes-Rockafellar augmented Lagrangian function given in (2), see [2, 6, 17, 28].

Algorithm 1: Euclidean safeguarded augmented Lagrangian algorithm

Step 0. Let $p^0 \in \mathbb{R}^n$, $\tau \in [0,1)$, $\gamma > 1$, $\eta_{\min} < \eta_{\max}$, $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max} > 0$, and $\rho_1 > 0$ be given. Also, take $\bar{\eta}^1 \in [\eta_{\min}, \eta_{\max}]^t$, $\bar{\lambda}^1 \in [\lambda_{\min}, \lambda_{\max}]^s$ and $\bar{\mu}^1 \in [0, \lambda_{\max}]^m$ initial Lagrange multipliers estimates, and $(\epsilon_k)_{k\in\mathbb{N}} \subset \mathbb{R}_+$ a sequence of tolerance parameters such that $\lim_{k\to\infty} \epsilon_k = 0$. Set $k \leftarrow 1$.

Step 1. *(Solve the subproblem)* Compute (if possible) $p^k \in \mathbb{R}^n$ such that

$$\left\| L'_{\rho_k}(p^k, \bar{\eta}^k, \bar{\lambda}^k, \bar{\mu}^k) \right\|_\infty \leq \epsilon_k.$$

If it is not possible, then stop the execution of the algorithm and declare failure.

Step 2. *(Estimate new multipliers)* Compute

$$\eta^k = \bar{\eta}^k + \rho_k h(p^k), \quad \lambda^k = \bar{\lambda}^k + \rho_k H(p^k), \quad \mu^k = \left[\bar{\mu}^k + \rho_k G(p^k)\right]_+.$$

Step 3. *(Update the penalty parameter)* Define $\nu^k := \frac{\mu^k - \bar{\mu}^k}{\rho_k}$. If $k = 1$ or

$$\max\left\{\left\|(h(p^k), H(p^k))\right\|_\infty, \left\|\nu^k\right\|_\infty\right\} \leq \tau \max\left\{\left\|(h(p^{k-1}), H(p^{k-1}))\right\|_\infty, \left\|\nu^{k-1}\right\|_\infty\right\},$$

set $\rho_{k+1} = \rho_k$. Otherwise, define $\rho_{k+1} = \gamma\rho_k$.

Step 4. *(Update safeguarded multipliers)* Compute $\bar{\eta}^{k+1} \in [\eta_{\min}, \eta_{\max}]^m$, $\bar{\lambda}^{k+1} \in [\lambda_{\min}, \lambda_{\max}]^m$, and $\bar{\mu}^{k+1} \in [0, \mu_{\max}]^p$.

Step 5. *(Begin a new iteration)* Set $k \leftarrow k+1$ and go to **Step 1**.

Algorithm 1 is widely recognized for its ability to generate AKKT sequences (see [5, 14]). Under strict constraint qualifications such as CRSC, or even weaker conditions, all limit points of such a sequence satisfy the KKT conditions (see, for example, [8, 10]). In the subsequent section, we introduce new strict constraint qualifications for problem (1). In this new scenario, it is no longer guaranteed that the limit points of the sequence generated by Algorithm 1 satisfy the KKT conditions. Therefore, we will employ tools from Riemannian Geometry to establish a connection between these new strict constraint qualifications and the Riemannian strict constraint qualifications introduced in [3]. Consequently, the Riemannian version of Algorithm 1, an intrinsic algorithm introduced in [35], generates AKKT sequences for problem (1). Under these new strict constraint qualifications, we will show that all its limit points satisfy the KKT conditions.

2.2 Nonlinear optimization problems on embedded submanifolds

In this subsection, we revisit some intrinsic strict constraint qualifications introduced in general Riemannian manifolds, focusing particularly on cases where the manifold is an embedded submanifold of Euclidean space. Hereafter, we assume that:

(H1) The set $[h'](p)$ is linearly independent, for all $p \in \mathbb{R}^n$.

In this way, the specific submanifold under consideration is as follows:

$$\mathbb{M} := \{q \in \mathbb{R}^n \mid h(q) = 0\}, \tag{14}$$

where $h = (h_1, \ldots, h_t)\colon \mathbb{R}^n \to \mathbb{R}^t$ is continuously differentiable on $\mathbb{R}^n$. We denote by $\langle \cdot, \cdot \rangle$ the metric in $\mathbb{M}$ induced from the Euclidean metric in $\mathbb{R}^n$, and by $\|\cdot\|$ the associated norm.

We use (14) to rewrite problem (1) in a more convenient form as an intrinsic nonlinear optimization problem, stated equivalently as follows:

$$\underset{q\in\mathbb{M}}{\text{Minimize}}\, f(q), \qquad \text{subject to} \qquad H(q) = 0,\ G(q) \leq 0, \tag{15}$$

where $f\colon \mathbb{M} \to \mathbb{R}$, $H = (H_1, \ldots, H_s)\colon \mathbb{M} \to \mathbb{R}^s$, and $G = (G_1, \ldots, G_m)\colon \mathbb{M} \to \mathbb{R}^m$ are continuously differentiable on $\mathbb{M}$. Let us denote the *intrinsic feasible set* of problem (15) by $\Omega_\mathbb{M} \subset \mathbb{M}$ and the *set of indices of active inequality constraints* at $p \in \Omega_\mathbb{M}$ by $\mathcal{A}_\mathbb{M}(p)$, i.e.,

$$\Omega_\mathbb{M} := \{q \in \mathbb{M} \mid H(q) = 0,\ G(q) \leq 0\}, \qquad \mathcal{A}_\mathbb{M}(p) := \{i \in \{1, \ldots, m\} \mid G_i(p) = 0\}. \tag{16}$$

Remark 1. *Problems* (1) *and* (15) *are topologically identical and, in particular, have the same solutions. Additionally, from* (10) *and* (16)*, we have* $\Omega = \Omega_{\mathbb{M}}$ *and* $\mathcal{A}(p) = \mathcal{A}_{\mathbb{M}}(p)$*. However, we emphasize that the functions* f*,* H*, and* G *in problem* (1) *are conceptually different from those in problem* (15)*, as they are now defined as functions on the Riemannian manifold. Consequently, the gradients of the functions* f*,* H_i*, and* G_i *are computed using formula* (9)*, specifically* $\operatorname{grad} f(q) = \operatorname{Proj}_q f'(q)$*,* $\operatorname{grad} H_i(q) = \operatorname{Proj}_q H_i'(q)$*, and* $\operatorname{grad} G_i(q) = \operatorname{Proj}_q G_i'(q)$*.*

The intrinsic Karush/Kuhn-Tucker (KKT) conditions are deemed satisfied at $p \in \Omega_{\mathbb{M}}$ if there exist corresponding Lagrange multipliers $(\lambda, \mu) \in \mathbb{R}^s \times \mathbb{R}^m_+$ that fulfill the following two conditions:

(i) $\operatorname{grad} \mathbb{L}(p, \lambda, \mu) = 0$, *(ii)* $\mu_j = 0$, for all $j \notin \mathcal{A}_{\mathbb{M}}(p)$,

where $\mathbb{L}(\cdot, \lambda, \mu) : \mathcal{M} \to \mathbb{R}$ is the *Lagrangian function* associated with problem (15) defined by

$$\mathbb{L}(q, \lambda, \mu) := f(q) + \sum_{i=1}^{s} \lambda_i H_i(q) + \sum_{j=1}^{m} \mu_j G_j(q),$$

and its *intrinsic gradient*, denoted by $\operatorname{grad} \mathbb{L}(q, \lambda, \mu) \in T_q\mathbb{M}$, is given by

$$\operatorname{grad} \mathbb{L}(q, \lambda, \mu) := \operatorname{grad} f(q) + \sum_{i=1}^{s} \lambda_i \operatorname{grad} H_i(q) + \sum_{j=1}^{m} \mu_j \operatorname{grad} G_j(q).$$

Similarly to Section 2, we introduce some notations for conciseness. Let $\mathcal{S}$ as in (12), and consider $\mathcal{I} \subseteq \mathcal{S}$ and $\mathcal{J} \subseteq \mathcal{A}_{\mathbb{M}}(p)$. For a given $q \in \Omega_{\mathbb{M}}$, define the following sets of vectors

$$[\operatorname{grad} H_{\mathcal{I}}, \operatorname{grad} G_{\mathcal{J}}](q) := \{\operatorname{grad} H_i(q) \mid i \in \mathcal{I}\} \cup \{\operatorname{grad} G_i(q) \mid i \in \mathcal{J}\}. \tag{17}$$

If one of the sets $\mathcal{I}$ or $\mathcal{J}$ is empty, then the corresponding set will not appear in (17). For instance, for $\mathcal{I} = \varnothing$, the set in (13) will be denoted by $[\operatorname{grad} G_{\mathcal{J}}](q) := \{\operatorname{grad} G_i(q) \mid i \in \mathcal{J}\}$. In addition, for the sake of simplicity, we set $\operatorname{grad} H := \operatorname{grad} H_{\mathcal{S}}$. Next, we recall two intrinsic constraint qualifications for problem (15), which were introduced in [36] and [13], respectively.

Definition 3. *A point* $p \in \Omega_{\mathbb{M}}$ *is said to satisfy the linear independence constraint qualification (LICQ) if* $[\operatorname{grad} H, \operatorname{grad} G_{\mathcal{A}_{\mathbb{M}}(p)}](p)$ *is linearly independent. It satisfies the Mangasarian-Fromovitz constraint qualification (MFCQ) if* $[\operatorname{grad} H, \operatorname{grad} G_{\mathcal{A}_{\mathbb{M}}(p)}](p)$ *is positive-linearly independent.*

For the sake of convenience, we recall the following strict constraint qualifications which were originally introduced and studied in a general Riemannian manifold in [3].

Definition 4. *A point* $p \in \Omega_{\mathbb{M}}$ *is said to satisfy:*

(i) the constant rank constraint qualification (CRCQ) if for any $\mathcal{I} \subset \mathcal{S}$ *and* $\mathcal{J} \subset \mathcal{A}_{\mathbb{M}}(p)$*, whenever the set* $[\operatorname{grad} H_{\mathcal{I}}, \operatorname{grad} G_{\mathcal{J}}](p)$ *is linearly dependent, there exists* $\epsilon > 0$ *such that* $[\operatorname{grad} H_{\mathcal{I}}, \operatorname{grad} G_{\mathcal{J}}](q)$ *is linearly dependent for all* $q \in \mathbb{B}_\epsilon(p)$*.*

(ii) the constant positive linear dependence condition (CPLD) if for any $\mathcal{I} \subset \mathcal{S}$ *and* $\mathcal{J} \subset \mathcal{A}_{\mathbb{M}}(p)$*, whenever the set* $[\operatorname{grad} H_{\mathcal{I}}, \operatorname{grad} G_{\mathcal{J}}](p)$ *is positive-linearly dependent, there exists* $\epsilon > 0$ *such that* $[\operatorname{grad} H_{\mathcal{I}}, \operatorname{grad} G_{\mathcal{J}}](q)$ *is linearly dependent for all* $q \in \mathbb{B}_\epsilon(p)$*.*

(iii) the Relaxed-CRCQ (RCRCQ) if there exists $\epsilon > 0$ *such that the following two conditions hold:*

   (a) the rank of $[\operatorname{grad} H](q)$ *is constant for all* $q \in \mathbb{B}_\epsilon(p)$*;*

   (b) Let $\mathcal{K} \subset \mathcal{S}$*, such that* $[\operatorname{grad} H_{\mathcal{K}}](p)$ *is a basis for* $\operatorname{Span}([\operatorname{grad} H](p))$*. For all* $\mathcal{J} \subset \mathcal{A}_{\mathbb{M}}(p)$*, if* $[\operatorname{grad} H_{\mathcal{K}}, \operatorname{grad} G_{\mathcal{J}}](p)$ *is linearly dependent, then* $[\operatorname{grad} H_{\mathcal{K}}, \operatorname{grad} G_{\mathcal{J}}](q)$ *is linearly dependent for all* $q \in \mathbb{B}_\epsilon(p)$*.*

(iv) *the Relaxed-CPLD (RCPLD) if there exists* $\epsilon > 0$ *such that the following two conditions hold:*

(a) *the rank of* $[\operatorname{grad} H](q)$ *is constant for all* $q \in \mathbb{B}_\epsilon(p)$*;*

(b) *Let* $\mathcal{K} \subset \mathcal{S}$*, such that* $[\operatorname{grad} H_{\mathcal{K}}](p)$ *is a basis for* $\operatorname{Span}([\operatorname{grad} H](p))$*. For all* $\mathcal{J} \subset \mathcal{A}_{\mathbb{M}}(p)$*, if* $[\operatorname{grad} H_{\mathcal{K}}, \operatorname{grad} G_{\mathcal{J}}](p)$ *is positive-linearly dependent, then* $[\operatorname{grad} H_{\mathcal{K}}, \operatorname{grad} G_{\mathcal{J}}](q)$ *is linearly dependent for all* $q \in \mathbb{B}_\epsilon(p)$*.*

Below we recall an intrinsic version of the sequential optimality conditions, which are satisfied at a local minimizer of problem (15) in the absence of constraint qualifications. Specifically, the AKKT conditions introduced in the general context of Riemannian manifolds in [35] and the positive-AKKT (PAKKT) conditions introduced in [3]. Let $p \in \Omega_{\mathbb{M}}$, $(p^k)_{k\in\mathbb{N}} \subset \mathbb{M}$, $(\lambda^k)_{k\in\mathbb{N}} \subset \mathbb{R}^s$, $(\mu^k)_{k\in\mathbb{N}} \subset \mathbb{R}^m_+$, and consider the following statements:

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

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

(iii) $\mu_i^k = 0$, for all $i \notin \mathcal{A}_{\mathbb{M}}(p)$ and sufficiently large k;

(iv) if $\gamma_k := \left\|(1, \lambda^k, \mu^k)\right\|_\infty \to +\infty$ it holds:

$$\lim_{k\to\infty} \frac{|\lambda_i^k|}{\gamma_k} > 0 \Longrightarrow \lambda_i^k H_i(p^k) > 0,\ \forall k \in \mathbb{N},$$

and

$$\lim_{k\to\infty} \frac{\mu_i^k}{\gamma_k} > 0 \Longrightarrow \mu_i^k G_i(p^k) > 0,\ \forall k \in \mathbb{N}.$$

Definition 5. *Assume that* $p \in \Omega_{\mathbb{M}}$*. If there exist sequences* $(p^k)_{k\in\mathbb{N}} \subset \mathbb{M}$*,* $(\lambda^k)_{k\in\mathbb{N}} \subset \mathbb{R}^s$ *and* $(\mu^k)_{k\in\mathbb{N}} \subset \mathbb{R}^m_+$ *such that*

(1) *conditions (i), (ii) and (iii) hold, then* $p \in \Omega_{\mathbb{M}}$ *is called an approximate-KKT (AKKT) point for problem* (15)*;*

(2) *conditions (i), (ii), (iii) and (iv) hold, then* $p \in \Omega_{\mathbb{M}}$ *is called a positive approximate-KKT (PAKKT) point for problem* (15)*.*

PAKKT is a necessary optimality condition, as shown in [3]. This is sufficient to conclude that AKKT is also a necessary optimality condition.

Theorem 4. *Let* $p \in \Omega_{\mathbb{M}}$ *be a local minimizer for problem* (15)*. Then* p *is a PAKKT point.*

Next, we recall two other constraint qualifications introduced for general Riemannian manifolds in [3]. To facilitate our discussion, let us first establish the following definition: For $p \in \Omega_{\mathbb{M}}$, we denote by $\mathcal{L}_{\mathbb{M}}(p)$ the *linearized cone associated with* $\Omega_{\mathbb{M}}$ *at* p which is defined as

$$\mathcal{L}_{\mathbb{M}}(p) = \big\{v \in T_p\mathbb{M} \mid \langle \operatorname{grad} H_i(p), v\rangle = 0,\ i \in \mathcal{S};\ \langle \operatorname{grad} G_j(p), v\rangle \le 0,\ j \in \mathcal{A}_{\mathbb{M}}(p)\big\},$$

and its polar is given by

$$\mathcal{L}_{\mathbb{M}}(p)^\circ = \Big\{v \in T_p\mathbb{M} \mid v = \sum_{i=1}^{s} \lambda_i \operatorname{grad} H_i(p) + \sum_{j=1}^{m} \mu_j \operatorname{grad} G_j(p),\ \mu_j \ge 0,\ \lambda_i \in \mathbb{R}\Big\}.$$

Definition 6. *A point* $p \in \Omega_{\mathbb{M}}$ *is said to satisfy the constant rank of the subspace component (CRSC) if there exists* $\epsilon > 0$ *such that for* $\mathcal{J}^-_{\mathbb{M}}(p) = \{j \in \mathcal{A}_{\mathbb{M}}(p) \mid -\operatorname{grad} G_j(p) \in \mathcal{L}_{\mathbb{M}}(p)^\circ\}$*, the rank of the set* $[\operatorname{grad} H, \operatorname{grad} G_{\mathcal{J}^-(p)}](q)$ *remains constant for all* $q \in \mathbb{B}_\epsilon(p)$*.*

Definition 7. *A point* $p \in \Omega_{\mathbb{M}}$ *satisfies the quasinormality constraint qualification (QN) if there do not exist* $\lambda \in \mathbb{R}^s$ *and* $\mu \in \mathbb{R}^m_+$ *such that* *(i)* $\sum_{i=1}^{s} \lambda_i \operatorname{grad} H_i(p) + \sum_{j \in \mathcal{A}_{\mathbb{M}}(p)} \mu_j \operatorname{grad} G_j(p) = 0$;

(ii) $\mu_j = 0$ *for all* $j \notin \mathcal{A}_{\mathbb{M}}(p)$ *and* $(\lambda, \mu) \neq 0$;

(iii) for all $\epsilon > 0$, *there exists* $q \in \mathbb{B}_\epsilon(p)$ *such that* $\lambda_i H_i(q) > 0$ *for all* $i \in \mathcal{S}$ *with* $\lambda_i \neq 0$ *and* $\mu_j G_j(q) > 0$ *for all* $j \in \mathcal{A}_{\mathbb{M}}(p)$ *with* $\mu_j > 0$.

In the following, we present an augmented Lagrangian algorithm to address problem (1) by refraining from penalizing the constraint set $\{q \in \mathbb{R}^n \mid h(q) = 0\}$ and instead focusing on penalizing the constraint set $\{q \in \mathbb{R}^n \mid H(q) = 0,\ G(q) \leq 0\}$. For this purpose, we recall the intrinsic safeguarded augmented Lagrangian algorithm introduced in [35]. This algorithm was initially designed to solve problem (15), which represents the Riemannian version of problem (1). The formulation of the algorithm involves the partial Powell-Hestenes-Rockafellar augmented Lagrangian function defined by (4). The *intrinsic safeguarded augmented Lagrangian algorithm* is stated as follows.

Algorithm 2: Intrinsic safeguarded augmented Lagrangian algorithm

Step 0. Let $p^0 \in \mathcal{M}$, $\tau \in [0\,,\,1)$, $\gamma > 1$, $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max} > 0$, and $\rho_1 > 0$ be given. Also, take $\bar{\lambda}^1 \in [\lambda_{\min}, \lambda_{\max}]^s$ and $\bar{\mu}^1 \in [0, \mu_{\max}]^m$ initial Lagrange multipliers estimates, and $(\epsilon_k)_{k\in\mathbb{N}} \subset \mathbb{R}_+$ a sequence of tolerance parameters such that $\lim_{k\to\infty} \epsilon_k = 0$. Set $k \leftarrow 1$.

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

$$\left\| \operatorname{grad} \mathbb{L}_{\rho_k}(p^k, \bar{\lambda}^k, \bar{\mu}^k) \right\| \leq \epsilon_k. \tag{18}$$

If it is not possible, then stop the execution of the algorithm and declare failure.

Step 2. *(Estimate new multipliers)* Compute

$$\lambda^k = \bar{\lambda}^k + \rho_k H(p^k), \qquad \mu^k = \left[\bar{\mu}^k + \rho_k G(p^k)\right]_+ .$$

Step 3. *(Update the penalty parameter)* Define $\nu^k = \frac{\mu^k - \bar{\mu}^k}{\rho_k}$. If $k = 1$ or

$$\max\left\{ \|H(p^k)\|_\infty\,,\, \|\nu^k\|_\infty \right\} \leq \tau \max\left\{ \|H(p^{k-1})\|_\infty\,,\, \|\nu^{k-1}\|_\infty \right\},$$

set $\rho_{k+1} = \rho_k$. Otherwise, define $\rho_{k+1} = \gamma \rho_k$.

Step 4. *(Update safeguarded multipliers)* Compute $\bar{\lambda}^{k+1} \in [\lambda_{\min}\,,\,\lambda_{\max}]^m$ and $\bar{\mu}^{k+1} \in [0\,,\,\mu_{\max}]^p$.

Step 5. *(Begin a new iteration)* Set $k \leftarrow k+1$ and go to **Step 1**.

The capability of Algorithm 2 to generate AKKT sequences for problem (15) was demonstrated in [35], while in [3] it was proven that it also produces PAKKT sequences. For reader's convenience and future reference, we revisit the main convergence results of Algorithm 2 established in [3].

Theorem 5. *Suppose that* $p \in \Omega_{\mathbb{M}}$ *satisfies RCPLD or CRSC. If* p *is an AKKT point, then* p *is a KKT point for problem* (15).

Theorem 6. *Let* $p \in \Omega_{\mathbb{M}}$ *be a PAKKT point with associated primal sequence* $(p^k)_{k\in\mathbb{N}}$ *and dual sequence* $(\lambda^k, \mu^k)_{k\in\mathbb{N}}$. *Assume that* p *satisfies QN. Then* $(\lambda^k, \mu^k)_{k\in\mathbb{N}}$ *is a bounded sequence. In particular,* p *satisfies the KKT conditions for problem* (15) *and any limit point of* $(\lambda^k, \mu^k)_{k\in\mathbb{N}}$ *is a Lagrange multiplier associated with* p.

Theorem 7. *Assume that Algorithm 2 generates an infinite sequence* $(p^k)_{k\in\mathbb{N}}$ *with a feasible limit point* p, *say,* $\lim_{k\in K} p^k = p$. *Then,* p *is a PAKKT point with correspondent primal sequence* $(p^k)_{k\in K}$ *and dual sequence* $(\lambda^k, \mu^k)_{k\in K}$ *as generated by Algorithm 2. In particular, if RCPLD or CRSC hold,* p *is a KKT point for problem* (15). *Alternatively, if QN holds,* p *is a KKT point for problem* (15) *and* $(\lambda^k, \mu^k)_{k\in K}$ *is bounded with any of its limit points being a Lagrange multiplier associated with* p.

3 Lower strict constraint qualifications

Let us now recall the Euclidean nonlinear programming problem (1). Inspired by the Riemannian approach, we will propose new weak constraint qualifications for problem (1). We will show that the new conditions are able to provide standard global convergence results to a constrained augmented Lagrangian method where a subset of linearly independent equality constraints are kept within the subproblems. These are termed lower-level constraints, which inspire the name of the conditions. Let us start by the extension of the sequential optimality conditions AKKT and PAKKT, which will be generated by the constrained algorithm we propose.

That is, in the absence of constraint qualifications, the following definition introduces sequential optimality conditions, which will be shown to be fulfilled by a local minimizer of problem (1). Consider the nonlinear programming problem (1) under assumption (H1). Let $p \in \Omega$, $(p^k)_{k\in\mathbb{N}} \subset \{q \in \mathbb{R}^n \mid h(q) = 0\}$, $(\eta^k)_{k\in\mathbb{N}} \subset \mathbb{R}^t$, $(\lambda^k)_{k\in\mathbb{N}} \subset \mathbb{R}^s$, $(\mu^k)_{k\in\mathbb{N}} \subset \mathbb{R}^m_+$, and consider the following statements:

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

(ii) $\lim_{k\to\infty} L'(p^k, \eta^k, \lambda^k, \mu^k) = 0$;

(iii) $\mu_i^k = 0$, for all $i \notin \mathcal{A}(p)$ and sufficiently large k;

(iv) if $\gamma_k := \left\|(1, \lambda^k, \mu^k)\right\|_\infty \to +\infty$ it holds:

$$\lim_{k\to\infty} \frac{|\lambda_i^k|}{\gamma_k} > 0 \Longrightarrow \lambda_i^k H_i(p^k) > 0,\ \forall k \in \mathbb{N},$$

and

$$\lim_{k\to\infty} \frac{\mu_i^k}{\gamma_k} > 0 \Longrightarrow \mu_i^k G_i(p^k) > 0,\ \forall k \in \mathbb{N}.$$

Definition 8. *Assume that $p \in \Omega$. If there exist sequences $(p^k)_{k\in\mathbb{N}} \subset \{q \in \mathbb{R}^n \mid h(q) = 0\}$, $(\eta^k)_{k\in\mathbb{N}} \subset \mathbb{R}^t$, $(\lambda^k)_{k\in\mathbb{N}} \subset \mathbb{R}^s$ and $(\mu^k)_{k\in\mathbb{N}} \subset \mathbb{R}^m_+$ such that*

(1) conditions (i), (ii), and (iii) hold, then $p \in \Omega$ is called a lower approximate-KKT (Lower-AKKT) point for problem (1);

(2) conditions (i), (ii), (iii), and (iv) hold, then $p \in \Omega$ is called a lower positive approximate-KKT (Lower-PAKKT) point for problem (1).

The difference with respect to the standard definition is that the sequence $\{p^k\}$ must be feasible with respect to the equality constraints that satisfy assumption (H1) while the sign control *(iv)* is not required for these constraints. The companion Lower constraint qualifications are defined as follows:

Definition 9. *Consider the nonlinear programming problem (1) under assumption (H1). Let Ω and $\mathcal{A}(p)$, where $p \in \Omega$, be given by (10). The point $p \in \Omega$ is said to satisfy:*

(i) the lower constant rank constraint qualification (Lower-CRCQ), if for any $\mathcal{I} \subset \mathcal{S}$ and $\mathcal{J} \subset \mathcal{A}(p)$, whenever the set $[h', H'_{\mathcal{I}}, G'_{\mathcal{J}}](p)$ is linearly dependent, there exists $\epsilon > 0$ such that $[h', H'_{\mathcal{I}}, G'_{\mathcal{J}}](q)$ is linearly dependent for all $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$.

(ii) the lower constant positive linear dependence (Lower-CPLD), if for any $\mathcal{I} \subset \mathcal{S}$ and $\mathcal{J} \subset \mathcal{A}(p)$, whenever the set $[h', H'_{\mathcal{I}}, G'_{\mathcal{J}}](p)$ is positive-linearly dependent, there exists $\epsilon > 0$ such that $[h', H'_{\mathcal{I}}, G'_{\mathcal{J}}](q)$ is linearly dependent for all $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$.

(iii) the Lower Relaxed-CRCQ (Lower-RCRCQ) if there exists $\epsilon > 0$ such that the following two conditions hold:

(a) the rank of $[h', H'](q)$ *is constant, for all* $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$*;*

(b) Let $\mathcal{K} \subset \mathcal{S}$ *be such that* $[h', H'_{\mathcal{K}}](p)$ *is a basis for* $\mathrm{Span}([h', H'](p))$*. For all* $\mathcal{J} \subset \mathcal{A}(p)$*, if* $[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](p)$ *is linearly dependent, then* $[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](q)$ *is linearly dependent for all* $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$*.*

(iv) the Lower Relaxed-CPLD (Lower-RCPLD), if there exists $\epsilon > 0$ *such that the following two conditions hold:*

(a) the rank of $[h', H'](q)$ *is constant for all* $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$*;*

(b) Let $\mathcal{K} \subset \mathcal{S}$ *be such that* $[h', H'_{\mathcal{K}}](p)$ *is a basis for* $\mathrm{Span}([h', H'](p))$*. For all* $\mathcal{J} \subset \mathcal{A}(p)$*, if* $[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](p)$ *is positive-linearly dependent, then the set* $[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](q)$ *is linearly dependent for all* $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$*.*

As previously mentioned, the motivation behind the development of Definition 9 arises from the application of penalty methods or augmented Lagrangian methods for solving nonlinear programming problems where a subset of equality constraints are kept within the subproblems. This approach provides the flexibility to preselect constraints – referred to as lower-level constraints – that align with specific interests or are simpler to handle, while penalizing only the more challenging constraints. Additionally, it ensures that the sequence generated by the chosen minimization method remains feasible for these lower-level constraints. This guarantees that when the stopping criterion for this method is satisfied at some point, the feasibility of those constraints is maintained.

The difference between the new strict constraint qualifications introduced in Definition 9 and the standard strict constraint qualifications lies in the requirement that the condition be satisfied at a smaller number of points. Specifically, the new conditions must be satisfied in a neighborhood of the point restricted to a previously chosen set of constraints, namely, $B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$. In contrast, the standard strict constraint qualifications require the point to satisfy the defining condition in a full neighborhood, i.e., $B_\epsilon(p)$, which is naturally more challenging to fulfill. Additionally, the definitions of (R)CRCQ and (R)CPLD are simplified by taking into account assumption (H1). That is, CRCQ and CPLD require taking into consideration $h'_{\bar{\mathcal{I}}}$, where $\bar{\mathcal{I}} \subset \mathcal{T}$, while Lower-CRCQ and Lower-CPLD require only h'. Also, in item (b) of RCRCQ and RCPLD, $h'_{\bar{\mathcal{K}}}$ is required where $\bar{\mathcal{I}} \subset \mathcal{S}$ while only h' is considered in Lower-RCRCQ and Lower-RCPLD. Thus it is clear that these definitions imply the usual ones. An example where the implication is strict will be given considering a definition of Lower-CRSC which we provide next:

Definition 10. *A point* $p \in \Omega$ *is said to satisfy the lower constant rank of the subspace component (Lower-CRSC) if there exists* $\epsilon > 0$ *such that for* $\mathcal{J}^-(p) = \{j \in \mathcal{A}(p) \mid -G'_j(p) \in \mathcal{L}(p)^\circ\}$*, the rank of the set* $[h', H', G'_{\mathcal{J}^-(p)}](q)$ *is constant for all* $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$*.*

Note that the standard CRSC condition requires constant rank of $[h', H', G'_{\mathcal{J}^-(p)}](q)$ for all $q \in B_\epsilon(p)$, while Lower-CRSC requires constant rank of the same set but restricted to $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$. The following example illustrates that Definition 10 is strictly less restrictive than its standard counterpart.

Example 1. *Let* $n \geq 4$ *and* $u, v, w \in \mathbb{R}^n$ *be linearly independent vectors. Define* $h, H_1, G_1, G_2 \colon \mathbb{R}^n \to \mathbb{R}$ *as follows:*

$$h(x) := u^\top x, \qquad H_1(x) := (u^\top x)^2 (v^\top x), \qquad G_1(x) := (u^\top x)^2 - w^\top x, \qquad G_2(x) := w^\top x,$$

and consider an optimization problem with the feasible set $\Omega := \{x \in \mathbb{R}^n \mid h(x) = 0, H_1(x) = 0, G_1(x) \leq 0, G_2(x) \leq 0\}$*. It is easy to see that* $\Omega = \mathrm{Span}(\{u, w\})^\perp$*. The Euclidean gradients of* h*,* H_1*,* G_1 *and* G_2 *are given, respectively, by*

$$h'(x) = u^\top, \quad H_1'(x) = 2(u^\top x)(v^\top x)u^\top + (u^\top x)^2 v^T, \quad G_1'(x) = 2(u^\top x)u^\top - w^\top, \quad G_2'(x) = w^\top. \tag{19}$$

We claim that not all $x \in \Omega$ *satisfy the usual CRSC constraint qualification. Indeed, it follows from* (11) *and* (19) *that* $\mathcal{J}^-(x) = \{1, 2\}$*, for all* $x \in \Omega$*. In addition, the following two statements hold:*

1. *rank*$(\{h'(y), H_1'(y), G_1'(y), G_2'(y)\}) = 2$, *for all* $y \in \mathbb{R}^n$ *such that* $u^\top y = 0$;

2. *rank*$(\{h'(y), H_1'(y), G_1'(y), G_2'(y)\}) = 3$, *for all* $y \in \mathbb{R}^n$ *such that* $u^\top y \neq 0$.

Thus, not all $x \in \Omega$ *satisfy the usual CRSC, as claimed. On the other hand, all* $x \in \Omega$ *satisfy Lower-CRCQ. Indeed, we have rank*$(\{h'(y), H_1'(y), G_1'(y), G_2'(y)\}) = 2$, *for all* $y \in \{y \in \mathbb{R}^n \mid h(y) = 0\}$.

We conclude this section by introducing a new constraint qualification, which we term *lower quasinormality.*

Definition 11. *A point* $p \in \Omega$ *satisfies the lower quasinormality constraint qualification (Lower-QN) if there do not exist* $\eta \in \mathbb{R}^t$, $\lambda \in \mathbb{R}^s$, *and* $\mu \in \mathbb{R}^m_+$ *such that*

(i) $\sum_{i=1}^t \eta_i h_i'(p) + \sum_{i=1}^s \lambda_i H_i'(p) + \sum_{j \in \mathcal{A}(p)} \mu_j G_j'(p) = 0$;

(ii) $\mu_j = 0$ *for all* $j \notin \mathcal{A}(p)$ *and* $(\eta, \lambda, \mu) \neq 0$;

(iii) for all $\epsilon > 0$, *there exists* $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ *such that* $\lambda_i H_i(q) > 0$ *for all* $i \in \mathcal{S}$ *with* $\lambda_i \neq 0$ *and* $\mu_j G_j(q) > 0$ *for all* $j \in \mathcal{A}(p)$ *with* $\mu_j > 0$.

It is important to note that Lower-QN differs from the usual QN definition solely by item (*iii*), where the respective item in the usual QN definition for problem (1) is given as follows:

(iii) for all $\epsilon > 0$, there exists $q \in B_\epsilon(p)$ such that $\eta_i h_i(q) > 0$ for all $i \in \mathcal{T}$ with $\eta_i \neq 0$, $\lambda_i H_i(q) > 0$ for all $i \in \mathcal{S}$ with $\lambda_i \neq 0$, and $\mu_j G_j(q) > 0$ for all $j \in \mathcal{A}(p)$ with $\mu_j > 0$.

The proof that the usual QN implies Lower-QN for problem (1) under (H1) is not immediate as in the case of the other Lower-SCQs we introduced. Let us prove this.

Theorem 8. *Let* $p \in \Omega$ *satisfy QN for problem* (1) *and assume (H1). Then,* p *satisfies Lower-QN.*

Proof. Assume, by contradiction, that $p \in \Omega$ does not satisfy Lower-QN. Then, there exist $\eta \in \mathbb{R}^t$, $\lambda \in \mathbb{R}^s$, and $\mu \in \mathbb{R}^m_+$ satisfying conditions (*i*), (*ii*), and (*iii*) in Definition 11. Condition (*iii*) implies the existence of a sequence $(p^k)_{k\in\mathbb{N}} \subset \{q \in \mathbb{R}^n \mid h(q) = 0\}$ such that $\lim_{k\to+\infty} p^k = p$. Furthermore, $\lambda_i H_i(p^k) > 0$ for all $i \in \mathcal{S}$ with $\lambda_i \neq 0$, and $\mu_j G_j(p^k) > 0$ for all $j \in \mathcal{A}(p)$ with $\mu_j > 0$. Define the set $\bar{\mathcal{T}} := \{i \in \mathcal{T} \mid \eta_i \neq 0\}$. If $\bar{\mathcal{T}} = \varnothing$, then p also does not satisfy the classical definition of QN. Now, assume $\bar{\mathcal{T}} \neq \varnothing$. Consider the submatrix $[h_{\bar{\mathcal{T}}}'](p)$ of $[h'](p)$, where $[h'](p)$ is the Jacobian matrix of h, and the rows of $[h_{\bar{\mathcal{T}}}'](p)$ correspond to $h_i'(p)$ for $i \in \bar{\mathcal{T}}$. Since assumption (H1) implies that $h_1'(p), \ldots, h_t'(p)$ are linearly independent, it follows that there is no vector $\beta_{\bar{\mathcal{T}}}$ of order $|\bar{\mathcal{T}}|$, with entries β_i for $i \in \bar{\mathcal{T}}$, such that $[h_{\bar{\mathcal{T}}}'](p)^T \beta_{\bar{\mathcal{T}}} = 0$ with $\beta_{\bar{\mathcal{T}}} \geq 0$ and $\beta_{\bar{\mathcal{T}}} \neq 0$. By Gordan's alternative theorem (see, for example, [11, p. 51]), it follows that there exists a vector $d \in \mathbb{R}^n$ such that $[h_{\bar{\mathcal{T}}}'](p)d > 0$, or equivalently, $h_i'(p)^T d > 0$ for all $i \in \bar{\mathcal{T}}$. Since we can make this construction replacing any $h_i'(p)$ by $-h_i'(p)$, we will assume that d is such that $\eta_i h_i'(p)^T d > 0$ for all $i \in \bar{\mathcal{T}}$. On the other hand, since $\lambda_i H_i(p^k) > 0$ for all $i \in \mathcal{S}$ with $\lambda_i \neq 0$ and $\mu_j G_j(p^k) > 0$ for all $j \in \mathcal{A}(p)$ with $\mu_j > 0$, define sequences $(q^k)_{k\in\mathbb{N}}$ and $(\varepsilon_k)_{k\in\mathbb{N}}$ such that $q^k := p^k + \varepsilon_k d$ with $\lim_{k\to+\infty} \varepsilon_k = 0^+$ and $\varepsilon_k \in \mathbb{R}_{++}$ such that $\lambda_i H_i(q^k) > 0$ for all $i \in \mathcal{S}$ with $\lambda_i \neq 0$, and $\mu_j G_j(q^k) > 0$ for all $j \in \mathcal{A}(p)$ with $\mu_j > 0$. Given that $\eta_i \neq 0$, $\lim_{k\to+\infty} p^k = p$, and $\lim_{k\to+\infty} \varepsilon_k = 0^+$, we have $\lim_{k\to+\infty} \left((\eta_i h_i(p^k + \varepsilon_k d) - \eta_i h_i(p))/\varepsilon_k\right) = \eta_i h_i'(p)^T d > 0$ for all $i \in \bar{\mathcal{T}}$. Therefore, since $h_i(p) = 0$, there exists $k_i \in \mathbb{N}$ such that $\eta_i h_i(q^k) = \eta_i h_i(p^k + \varepsilon_k d) > 0$ for all $k > k_i$. Let $\bar{k} = \max\{k_i \mid i \in \bar{\mathcal{T}}\}$. Consequently, since $\lim_{k\to+\infty} p^k = p$, for any $\epsilon > 0$, there exists $q^k \in B_\epsilon(p)$ such that $\eta_i h_i(q^k) > 0$ for all $i \in \mathcal{T}$ with $\eta_i \neq 0$, $\lambda_i H_i(q^k) > 0$ for all $i \in \mathcal{S}$ with $\lambda_i \neq 0$, and $\mu_j G_j(q^k) > 0$ for all $j \in \mathcal{A}(p)$ with $\mu_j > 0$ and all $k \geq \bar{k}$. Since $\eta \in \mathbb{R}^t$, $\lambda \in \mathbb{R}^s$, and $\mu \in \mathbb{R}^m_+$ satisfy conditions (*i*) and (*ii*) in Definition 11, this contradicts the classical definition of QN. □

Let us now show that the above implication is strict.

Example 2. *Let $n \geq 3$ be a positive integer and $u, v \in \mathbb{R}^n$ be such that $u \neq 0$, $v \neq 0$, $u^T v = 0$ and $h, H_1, G_1 \colon \mathbb{R}^n \to \mathbb{R}$ be functions defined, respectively, by*

$$h(x) := u^T x - v^T x, \qquad H_1(x) := (u^T x)^2 + (v^T x)^2, \qquad G_1(x) := (u^T x)^2 - (v^T x)^2. \tag{20}$$

Let $f \colon \mathbb{R}^n \to \mathbb{R}$ be a function. Consider the following constrained optimization problem

$$\operatorname*{Minimize}_{x \in \mathbb{R}^n} f(x) \qquad \text{subject to} \quad h(x) = 0,\ H_1(x) = 0,\ G_1(x) \leq 0. \tag{21}$$

The Euclidean gradients of the functions h, H_1 and G_1 are given, respectively, by

$$h'(x) = u^T - v^T, \qquad H_1'(x) = 2(u^T x)u^T + 2(v^T x)v^T, \qquad G_1'(x) = 2(u^T x)u^T - 2(v^T x)v^T. \tag{22}$$

Denote by $\Omega = \{x \in \mathbb{R}^n \mid \ h(x) = 0,\ H_1(x) = 0,\ G_1(x) \leq 0\}$ the feasible set of Problem (21). *Thus, using* (20) *we have $\Omega = \operatorname{Span}\{u, v\}^{\perp}$ and $\mathcal{A}(x) = \{1\}$, for all $x \in \Omega$. Thus, by using* (22), *note that for items* (*i*) *and* (*ii*) *of Definition 11 to be satisfied at $x \in \Omega$ we must take $\eta = 0$ and $\lambda_1 \in \mathbb{R}$ and $\mu_1 \in \mathbb{R}_+$ are arbitrary. In addition, given $\epsilon > 0$ and $x \in \Omega$, we can choose $y \in B_\epsilon(x)$ with $y \notin \Omega$ satisfying $H_1(y) = (u^T y)^2 + (v^T y)^2 > 0$ and $G_1(y) = (u^T y)^2 - (v^T y)^2 > 0$. For instance, for $y := x + \alpha u$ we can choose $\alpha > 0$ such that $y \in B_\epsilon(x)$ and $H_1(y) = \alpha^2 \|u\|^4 > 0$ and $G_1(y) = \alpha^2 \|u\|^4 > 0$. Thus, choosing $\eta_1 = 0$, $\lambda_1 > 0$ and $\mu_1 > 0$, all three items of the usual definition of QN are satisfied. Hence, all $x \in \Omega$ do not satisfy the usual QN. On the other hand, since $G_1(y) = 0$, for all $y \in B_\epsilon(x) \cap \{y \in \mathbb{R}^n :\ h(y) = 0\}$, there is no $\mu_1 > 0$ satisfying item* (*iii*) *of Definition 11 such that $\mu_1 G_1(y) > 0$. Therefore, all $x \in \Omega$ satisfy Lower-QN.*

4 Connecting the extrinsic and intrinsic approaches

This section establishes connections between the extrinsic concepts discussed earlier in Section 3, related to the nonlinear optimization problem presented in the format (1), and the ideas addressed in Subsection 2.2 regarding the nonlinear optimization problem presented intrinsically in (15). Our goal is to establish new global convergence results of an augmented Lagrangian algorithm for the Euclidean problem by means of the equivalent optimization problem on an embedded manifold. In order to do this we will first show that the Euclidean KKT conditions for problem (1) and the Riemmanian KKT conditions for the equivalent problem (15) are in fact equivalent. This appears to be new as we only found a proof of this fact under convexity assumptions, see [36].

Recall that we are under assumption (H1), hence it follows from (8) that the mapping Proj_q is well-defined for all $q \in \mathbb{R}^n$.

Theorem 9. *A point $p \in \Omega$ is a KKT (respectively, Lower-AKKT and Lower-PAKKT) point of problem* (1) *if and only if $p \in \Omega_{\mathbb{M}}$ is a KKT (respectively, AKKT and PAKKT) point of problem* (15).

Proof. First, we establish the equivalence for KKT points. Assume that $p \in \Omega$ is a KKT point for problem (1). Thus, there exist $(\eta, \lambda, \mu) \in \mathbb{R}^t \times \mathbb{R}^s \times \mathbb{R}^m_+$ such that $L'(p, \eta, \lambda, \mu) = 0$ and $\mu_i = 0$ for all $i \notin \mathcal{A}(p)$, i.e.,

$$f'(p) + \sum_{i=1}^{t} \eta_i h_i'(p) + \sum_{i=1}^{s} \lambda_i H_i'(p) + \sum_{i=1}^{m} \mu_i G_i'(p) = 0, \qquad \mu_i = 0,\ \forall\ i \notin \mathcal{A}(p). \tag{23}$$

From (7), it follows that $h'(p)^\top \eta = \sum_{i=1}^{t} \eta_i h_i'(p) \in T_p\mathbb{M}^{\perp}$. Consequently, considering (8), we conclude that $\operatorname{Proj}_p h'(p)^\top \eta = 0$. Thus, by applying Proj_p to (23) and using (9) along with the fact that $\mathcal{A}(p) = \mathcal{A}_{\mathbb{M}}(p)$, we obtain

$$\operatorname{grad} f(p) + \sum_{i=1}^{s} \lambda_i \operatorname{grad} H_i(p) + \sum_{i=1}^{m} \mu_i \operatorname{grad} G_i(p) = 0, \qquad \mu_i = 0,\ \forall\ i \notin \mathcal{A}_{\mathbb{M}}(p).$$

Therefore, $\operatorname{grad} \mathbb{L}(p,\lambda,\mu)=0$ and $\mu_j=0$, for all $j\notin \mathcal{A}_{\mathbb{M}}(p)$. Hence, since we also have $p\in\Omega_{\mathbb{M}}$, it follows that p is a KKT point for problem (15) as well.

Reciprocally, suppose that $p\in\Omega_{\mathbb{M}}$ is a KKT point for problem (15). Thus, there exist $(\lambda,\mu)\in\mathbb{R}^s\times\mathbb{R}^m_+$ such that $\operatorname{grad}\mathbb{L}(p,\lambda,\mu)=0$ and $\mu_j=0$, for all $j\notin\mathcal{A}_{\mathbb{M}}(p)$, i.e.,

$$\operatorname{grad} f(p)+\sum_{i=1}^{s}\lambda_i \operatorname{grad} H_i(p)+\sum_{i=1}^{m}\mu_i \operatorname{grad} G_i(p)=0, \qquad \mu_i=0,\ \forall\, i\notin\mathcal{A}_{\mathbb{M}}(p).$$

Therefore, using the projection formula Proj_p and (9), we have

$$\operatorname{Proj}_p\Big(f'(p)+\sum_{i=1}^{s}\lambda_i H_i'(p)+\sum_{i=1}^{m}\mu_i G_i'(p)\Big)=0, \qquad \mu_i=0,\ \forall\, i\notin\mathcal{A}_{\mathbb{M}}(p).$$

Hence, by (6), we conclude that $f'(p)+\sum_{i=1}^{s}\lambda_i H_i'(p)+\sum_{i=1}^{m}\mu_i G_i'(p)\in T_p\mathbb{M}^{\perp}$. Thus, (7) implies that there exists $\eta\in\mathbb{R}^t$ such that

$$f'(p)+\sum_{i=1}^{s}\lambda_i H_i'(p)+\sum_{i=1}^{m}\mu_i G_i'(p)=-\sum_{i=1}^{t}\eta_i h_i'(p).$$

Consequently, we conclude that there exists $(\eta,\lambda,\mu)\in\mathbb{R}^t\times\mathbb{R}^s\times\mathbb{R}^m_+$ such that

$$f'(p)+\sum_{i=1}^{t}\eta_i h_i'(p)+\sum_{i=1}^{s}\lambda_i H_i'(p)+\sum_{i=1}^{m}\mu_i G_i'(p)=0, \qquad \mu_i=0,\ \forall\, i\notin\mathcal{A}_{\mathbb{M}}(p).$$

Therefore, $L'(p,\eta,\lambda,\mu)=0$ and $\mu_i=0$, for all $i\notin\mathcal{A}_{\mathbb{M}}(p)$. Since $p\in\Omega_{\mathbb{M}}=\Omega$ and $\mathcal{A}_{\mathbb{M}}(p)=\mathcal{A}(p)$, the point p satisfies the KKT conditions for problem (1), concluding the proof for KKT points.

To establish the statements regarding other equivalences, we first note that the projection map Proj_q, defined in (8), is continuous. Consequently, the part of the proof involving the Lagrangian follows arguments similar to those used for establishing the statement related to KKT points. Given that $\mathbb{M}=\{q\in\mathbb{R}^n \mid h(q)=0\}$, $\Omega_{\mathbb{M}}=\Omega$, and $\mathcal{A}_{\mathbb{M}}(p)=\mathcal{A}(p)$, we conclude the proof by observing that the additional conditions required are directly equivalent. □

The next theorem shows that Euclidean LICQ and MFCQ are equivalent to their Riemannian counterparts. The proof for the LICQ case is in [36]. Since both proofs follow directly from Lemma 1, we omit them.

Theorem 10. *A point $p\in\Omega$ satisfies LICQ (respectively, MFCQ) for problem* (1) *if and only if $p\in\Omega_{\mathbb{M}}$ satisfies LICQ (respectively, MFCQ) for problem* (15).

We now begin the discussion where we establish the connection between the extrinsic Definition 9 and the intrinsic Definition 4. Our discussion commences by establishing the connection between the first two items of these definitions.

Theorem 11. *A point $p\in\Omega$ satisfies Lower-CRCQ (respectively, Lower-CPLD) for problem* (1) *if and only if $p\in\Omega_{\mathbb{M}}$ satisfies CRCQ (respectively, CPLD) for problem* (15).

Proof. Suppose that $p\in\Omega$ satisfies Lower-CRCQ (respectively, Lower-CPLD) for problem (1). Assume, by contradiction, that $p\in\Omega_{\mathbb{M}}$ does not satisfies CRCQ (respectively, CPLD) for problem (15). According to Definition 4, there exist $\mathcal{I}\subset\mathcal{S}$ and $\mathcal{J}\subset\mathcal{A}_{\mathbb{M}}(p)=\mathcal{A}(p)$ such that $[\operatorname{grad} H_{\mathcal{I}},\operatorname{grad} G_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent), and for each $k\in\mathbb{N}$, there exists $q_k\in\mathbb{B}_{1/k}(p)$ such that $[\operatorname{grad} H_{\mathcal{I}},\operatorname{grad} G_{\mathcal{J}}](q_k)$ is linearly independent. Taking into account that $[h'](p)$ is linearly independent and $q_k\in\mathbb{B}_{1/k}(p)$ for all $k\in\mathbb{N}$, we can assume without loss of generality that $[h'](q_k)$ is also linearly independent. Consequently, by applying Lemma 1 and considering (5) and (9), we conclude that $[h',H_{\mathcal{I}}',G_{\mathcal{J}}'](q_k)$ is also linearly independent for each

$q_k \in \mathbb{B}_{1/k}(p)$. On the other hand, since $[\operatorname{grad} H_{\mathcal{I}}, \operatorname{grad} G_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent), applying again Lemma 1 and taking into account (5) and (9), we obtain that $[h', H'_{\mathcal{I}}, G'_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent). Thus, since $p \in \Omega$ satisfies Lower-CRCQ (respectively, Lower-CPLD) for problem (1), item (i) (respectively, item (ii)) of Definition 9 implies that there exists $\epsilon > 0$ such that $[h', H'_{\mathcal{I}}, G'_{\mathcal{J}}](q)$ is linearly dependent (respectively, positive-linearly dependent) for all $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$. Hence, as $B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ is an open subset of $\mathbb{M}$ and $\lim_{k \to +\infty} q_k = p$, there exists $\bar{k} \in \mathbb{N}$ such that $q_{\bar{k}} \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\} \cap \mathbb{B}_{1/\bar{k}}(p)$ and $[h', H'_{\mathcal{I}}, G'_{\mathcal{J}}](q_{\bar{k}})$ is linearly dependent (respectively, positive-linearly dependent), which is a contradiction. Therefore, $p \in \Omega_{\mathbb{M}}$ satisfies CRCQ (respectively, CPLD) for problem (15).

Reciprocally, suppose that $p \in \Omega_{\mathbb{M}}$ satisfies CRCQ (respectively, CPLD) for problem (15). Assume, by contradiction, that $p \in \Omega$ does not satisfy Lower-CRCQ (respectively, Lower-CPLD) for problem (1). Thus, there exist $\mathcal{I} \subset \mathcal{S}$ and $\mathcal{J} \subset \mathcal{A}(p) = \mathcal{A}_{\mathbb{M}}(p)$ such that $[h', H'_{\mathcal{I}}, G'_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent) and, for all $k \in \mathbb{N}$, there exists $q_k \in B_{1/k}(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ such that $[h', H'_{\mathcal{I}}, G'_{\mathcal{J}}](q_k)$ is linearly independent. Hence, by using Lemma 1 and considering (5) and (9), the set $[\operatorname{grad} H_{\mathcal{I}}, \operatorname{grad} G_{\mathcal{J}}](q_k)$ is linearly independent. Since $[h', H'_{\mathcal{I}}, G'_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent), by employing Lemma 1 and considering (5) and (9), the set $[\operatorname{grad} H_{\mathcal{I}}, \operatorname{grad} G_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent). Thus, considering that $p \in \Omega_{\mathbb{M}}$ satisfies CRCQ (respectively, CPLD) for problem (15), there exists $\epsilon > 0$ such that $[\operatorname{grad} H_{\mathcal{I}}, \operatorname{grad} G_{\mathcal{J}}](q)$ is linearly dependent for all $q \in \mathbb{B}_\epsilon(p)$. Since $B_{1/k}(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ is an open subset of $\mathbb{M}$ and $\lim_{k \to +\infty} q_k = p$, there exists $\bar{k} \in \mathbb{N}$ such that $q_{\bar{k}} \in \mathbb{B}_\epsilon(p) \cap B_{1/\bar{k}}(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ and $[\operatorname{grad} H_{\mathcal{I}}, \operatorname{grad} G_{\mathcal{J}}](q_{\bar{k}})$ is linearly dependent, which is a contradiction. Therefore, $p \in \Omega$ satisfies Lower-CRCQ (respectively, Lower-CPLD) for problem (1), which concludes the proof. □

The following theorem establishes the connection between Lower-RCRCQ and Lower-RCPLD for problem (1), and RCRCQ and RCPLD for problem (15).

Theorem 12. *A point $p \in \Omega$ satisfies Lower-RCRCQ (respectively, Lower-RCPLD) for problem* (1) *if and only if $p \in \Omega_{\mathbb{M}}$ satisfies RCRCQ (respectively, RCPLD) for problem* (15).

Proof. Suppose that $p \in \Omega$ satisfies Lower-RCRCQ (respectively, Lower-RCPLD) for problem (1). Assume, by contradiction, that $p \in \Omega_{\mathbb{M}}$ does not satisfies RCRCQ (respectively, RCPLD) for problem (15). Thus, taking $\mathcal{K} \subset \mathcal{S}$ such that $[\operatorname{grad} H_{\mathcal{K}}](p)$ is a basis for $\operatorname{Span}([\operatorname{grad} H](p))$, at least one of the following two conditions holds for each $k \in \mathbb{N}$:

- *(a)* there exists $q_k \in \mathbb{B}_{1/k}(p)$ such that $|\mathcal{K}| = \operatorname{rank}([\operatorname{grad} H](q)) \neq \operatorname{rank}([\operatorname{grad} H](q_k))$;
- *(b)* there exist $\mathcal{J} \subset \mathcal{A}(p)$ and $q_k \in \mathbb{B}_{1/k}(p)$ such that $[\operatorname{grad} H_{\mathcal{K}}, \operatorname{grad} G_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent) and $[\operatorname{grad} H_{\mathcal{K}}, \operatorname{grad} G_{\mathcal{J}}](q_k)$ is linearly independent.

First, assume that *(a)* holds for infinitely many $k \in \mathbb{N}$. Using Lemma 3, it follows that there exists a subsequence $(q_{k_j})_{j \in \mathbb{N}}$ of $(q_k)_{k \in \mathbb{N}}$ such that $|\mathcal{K}| < \operatorname{rank}([\operatorname{grad} H](q_{k_j}))$ for all $j \in \mathbb{N}$. Thus, since $\mathcal{S}$ is finite, there exists $\bar{\mathcal{K}} \subset \{1, \ldots, s\}$ satisfying

$$|\mathcal{K}| < |\bar{\mathcal{K}}| := \operatorname{rank}([\operatorname{grad} H_{\bar{\mathcal{K}}}](q_{k_j})), \qquad \forall j \in \mathbb{N}. \tag{24}$$

In particular, the definition of $\bar{\mathcal{K}}$ implies that $[\operatorname{grad} H_{\bar{\mathcal{K}}}](q_{k_j})$ is linearly independent. Since $[h'](q_{k_j})$ is linearly independent, applying Lemma 1 and using (5) and (9), we conclude that $[h', H'_{\bar{\mathcal{K}}}](q_{k_j})$ is also linearly independent for all $j \in \mathbb{N}$. Hence, due to $\bar{\mathcal{K}} \subset \mathcal{S}$, we have

$$t + |\bar{\mathcal{K}}| \leq \operatorname{rank}([h', H'](q_{k_j})), \qquad \forall j \in \mathbb{N}. \tag{25}$$

Since $p \in \Omega$ satisfies Lower-RCRCQ (respectively, Lower-RCPLD) for problem (1), there exists $\epsilon > 0$ such that $\operatorname{rank}([h', H'](q))$ is constant for all $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$. Thus, considering that $[h'](q)$ is linearly independent for all $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$, there exists $\hat{\mathcal{K}} \subset \mathcal{S}$ such that $[H'_{\hat{\mathcal{K}}}](p)$ is linearly independent and

$$t + |\hat{\mathcal{K}}| = \operatorname{rank}([h', H'](q)), \qquad \forall q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}. \tag{26}$$

Since $\lim_{j \to +\infty} q_{kj} = p$, there exist j_0 such that $q_{kj} \in B_\epsilon(p) \cap \{p \in \mathbb{R}^n \mid h(p) = 0\}$ for all $j \geq j_0$. Hence, (25) and (26) imply that

$$|\bar{\mathcal{K}}| \leq |\hat{\mathcal{K}}|. \tag{27}$$

Thus, utilizing Lemma 1 and considering (5) and (9), if follows that $[\operatorname{grad} H_{\hat{\mathcal{K}}}](p)$ is also linearly independent. Hence, considering that $[\operatorname{grad} H_{\mathcal{K}}](p)$ is a basis for $\operatorname{Span}([\operatorname{grad} H](p))$, we conclude that $|\hat{\mathcal{K}}| \leq |\mathcal{K}|$. Thus, by (24), we obtain $|\hat{\mathcal{K}}| \leq |\mathcal{K}| < |\bar{\mathcal{K}}|$, contradicting (27). Therefore, *(a)* must hold only for a finite number of $k \in \mathbb{N}$. Hence, it follows that *(b)* holds for all sufficiently large $k \in \mathbb{N}$. We may assume, without loss of generality, that *(b)* holds for all $k \in \mathbb{N}$. Note that by applying Lemma 2 and using (5) and (9), we obtain that $[h', H'_{\mathcal{K}}](p)$ is a basis for $\operatorname{Span}([h', H'](p))$. Let $\mathcal{J} \subset \mathcal{A}(p)$ be such that $[\operatorname{grad} H_{\mathcal{K}}, \operatorname{grad} G_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent) and $(q_n)_{k \in \mathbb{N}} \subset \mathbb{B}_{1/k}(p)$ be a sequence such that $[\operatorname{grad} H_{\mathcal{K}}, \operatorname{grad} G_{\mathcal{J}}](q_k)$ is linearly independent for all $k \in \mathbb{N}$. Given that $[h'](q_k)$ is linearly independent for all $k \in \mathbb{N}$, by applying Lemma 1 and considering (5) and (9), we have

$$[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](q_k) \tag{28}$$

is also linearly independent for all $k \in \mathbb{N}$. Now, as $[\operatorname{grad} H_{\mathcal{K}}, \operatorname{grad} G_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent), using Lemma 1 and taking into account (5) and (9), we obtain that $[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](p)$ is also linearly dependent (respectively, positive-linearly dependent). Since the point $p \in \Omega$ satisfies Lower-RCRCQ (respectively, Lower-RCPLD) for problem (1), there exists $\epsilon > 0$ such that $[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](q)$ is linearly dependent for all $q \in B_\epsilon(p) \cap \{p \in \mathbb{R}^n \mid h(p) = 0\}$. Thus, due to $\lim_{k \to +\infty} q_k = p$, there exist $q_k \in B_\epsilon(p)$ such that the set in (28) is linearly dependent, which is also a contradiction. Therefore, $p \in \Omega_{\mathbb{M}}$ satisfies RCRCQ (respectively, RCPLD) for problem (15).

Reciprocally, suppose that $p \in \Omega_{\mathbb{M}}$ satisfies RCRCQ (respectively, RCPLD) for problem (15). Assume, by contradiction, that $p \in \Omega$ does not satisfy Lower-RCRCQ (respectively, Lower-RCPLD) for problem (1). Thus, taking $\mathcal{K} \subset \mathcal{S}$ such that $[h', H'_{\mathcal{K}}](p)$ is a basis for $\operatorname{Span}([h', H'](p))$, at least one of the following two conditions holds for each $k \in \mathbb{N}$:

(c) there exists $q_k \in B_{1/k}(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ such that $t + |\mathcal{K}| := \operatorname{rank}([h', H'](p)) \neq \operatorname{rank}([h', H'](q_k))$;

(d) there exist $\mathcal{J} \subset \mathcal{A}(p)$ and $q_k \in B_{1/k}(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ such that $[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent) and $[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](q_k)$ is linearly independent.

Let us assume that condition *(c)* holds for infinitely many $k \in \mathbb{N}$. Then, there exists a subsequence $(q_{k_j})_{j \in \mathbb{N}}$ of $(q_k)_{k \in \mathbb{N}}$ such that

$$t + |\mathcal{K}| < \operatorname{rank}([h', H'](q_{k_j})), \qquad \forall j \in \mathbb{N}.$$

Considering that $\{1, \dots s\}$ is finite, there exists $\bar{\mathcal{K}} \subset \{1, \dots s\}$ satisfying

$$t + |\mathcal{K}| < t + |\bar{\mathcal{K}}| := \operatorname{rank}([h', H'_{\bar{\mathcal{K}}}](q_{k_j})), \qquad \forall j \in \mathbb{N}. \tag{29}$$

Hence, $[h', H'_{\bar{\mathcal{K}}}](q_{k_j})$ is linearly independent for all $j \in \mathbb{N}$. Thus, using Lemma 1 and considering (5) and (9), we conclude that $[\operatorname{grad} H_{\bar{\mathcal{K}}}](q_{k_j})$ is also linearly independent for all $j \in \mathbb{N}$. In particular, we have

$$|\bar{\mathcal{K}}| \leq \operatorname{rank}([\operatorname{grad} H](q_{k_j})), \qquad \forall j \in \mathbb{N}. \tag{30}$$

Since $p \in \Omega_{\mathbb{M}}$ satisfies RCRCQ (respectively, RCPLD) for problem (15), there exists $\epsilon > 0$ such that $\text{rank}([\text{grad}\, H](q))$ is constant for all $q \in \mathbb{B}_\epsilon(p)$. Let $\hat{\mathcal{K}} \subset \{1, \ldots, s\}$ be such that $[\text{grad}\, H_{\hat{\mathcal{K}}}](p)$ is a basis of $\text{Span}([\text{grad}\, H](p))$. Thus,

$$|\hat{\mathcal{K}}| = \text{rank}([\text{grad}\, H](q)), \qquad \forall q \in \mathbb{B}_\epsilon(p). \tag{31}$$

Hence, due to $\lim_{j\to+\infty} q_{k_j} = p$, there exist $q_{k_j} \in \mathbb{B}_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$, which, together with (30) and (31), implies that $|\bar{\mathcal{K}}| \leq |\hat{\mathcal{K}}|$. On the other hand, taking into account Lemma 1, we conclude that $[h', H'_{\hat{\mathcal{K}}}](p)$ is linearly independent. Since $[h', H'_{\mathcal{K}}](p)$ is a basis for $\text{Span}([h', H'](p))$, we have $|\hat{\mathcal{K}}| \leq |\mathcal{K}|$. The latter two inequalities imply that $|\bar{\mathcal{K}}| \leq |\hat{\mathcal{K}}| \leq |\mathcal{K}|$, contradicting (29). Therefore, *(c)* must hold only for a finite number of indexes $k \in \mathbb{N}$. Hence, without loss of generality, we may assume that *(d)* holds for all $k \in \mathbb{N}$. Since $[h', H'_{\mathcal{K}}](p)$ is a basis for $\text{Span}([h', H'](p))$, applying Lemma 2 and using (5) and (9), we obtain that $[\text{grad}\, H_{\mathcal{K}}](p)$ is also a basis for $\text{Span}([\text{grad}\, H](p))$. Let $\mathcal{J} \subset \mathcal{A}(p)$ be such that $[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent), and consider a sequence $(q_k)_{k\in\mathbb{N}} \subset B_{1/k}(p)$ such that $[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](q_k)$ is linearly independent for all $k \in \mathbb{N}$. Since $[h'](q_k)$ is linearly independent, applying Lemma 1 and considering (5) and (9), we have

$$[\text{grad}\, H_{\mathcal{K}}, \text{grad}\, G_{\mathcal{J}}](q_k) \tag{32}$$

is also linearly independent. Now, as $[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent), applying Lemma 1 and considering (5) and (9), we have that $[\text{grad}\, H_{\mathcal{K}}, \text{grad}\, G_{\mathcal{J}}](p)$ is also linearly dependent (respectively, positive-linearly dependent). Taking into account that $p \in \Omega_{\mathbb{M}}$ satisfies RCRCQ (respectively, RCPLD) for problem (15), there exists $\epsilon > 0$ such that $[\text{grad}\, H_{\mathcal{K}}, \text{grad}\, G_{\mathcal{J}}](q)$ is linearly dependent for all $q \in \mathbb{B}_\epsilon(p)$. Given that $\lim_{k\to+\infty} q_k = p$, there exist $q_k \in \mathbb{B}_\epsilon(p)$ such that the set in (32) is linearly dependent, which is a contradiction. ☐

As a consequence of Theorems 10, 11, and 12, along with the relationships established for strict constraint qualifications in [3] for a general Riemannian manifold, the diagram in Figure 1 illustrates the relationship among the lower strict constraint qualifications given in Definition 9.

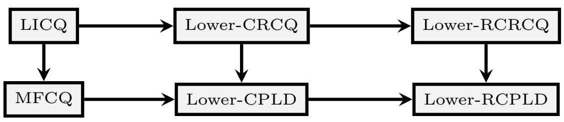

Figure 1: Lower strict constraint qualifications for problem (1).

To establish the relationship between Lower-CRSC and CRSC, we need to show the equality of the sets $\mathcal{J}^-(p)$ and $\mathcal{J}^-_{\mathbb{M}}(p)$. Since the proof is straightforward, we will omit it.

Lemma 13. *Let $\mathcal{J}^-(p)$ and $\mathcal{J}^-_{\mathbb{M}}(p)$ be as in Definitions 10 and 6, respectively. Then, it holds that $\mathcal{J}^-(p) = \mathcal{J}^-_{\mathbb{M}}(p)$.*

In the next theorem we establish the connection between Lower-CRSC and CRSC.

Theorem 14. *A point $p \in \Omega$ satisfies Lower-CRSC for problem* (1) *if and only if $p \in \Omega_{\mathbb{M}}$ satisfies CRSC for problem* (15).

Proof. Suppose that $p \in \Omega$ satisfies Lower-CRSC for problem (1). Assume, by contradiction, that $p \in \Omega_{\mathbb{M}}$ does not satisfies CRSC for problem (15). Thus, for each $k \in \mathbb{N}$, there exists $q_k \in \mathbb{B}_{1/k}(p)$ such that $|\mathcal{K}| := \text{rank}([\text{grad}\, H, \text{grad}\, G_{\mathcal{J}^-_{\mathbb{M}}(p)}](p)) \neq \text{rank}([\text{grad}\, H, \text{grad}\, G_{\mathcal{J}^-_{\mathbb{M}}(p)}](q_k))$. Using Lemma 3, we may assume that there exists a subsequence $(q_{k_j})_{j\in\mathbb{N}}$ of $(q_k)_{k\in\mathbb{N}}$ such that

$$|\mathcal{K}| < \text{rank}([\text{grad}\, H, \text{grad}\, G_{\mathcal{J}^-_{\mathbb{M}}(p)}](q_{k_j})), \qquad \forall j \in \mathbb{N}.$$

Thus, due to $\mathcal{S}$ and $\mathcal{J}^-_{\mathbb{M}}(p)$ being finite sets, there exist $\bar{\mathcal{K}} \subset \mathcal{S}$ and $\bar{\mathcal{K}}^- \subset \mathcal{J}^-_{\mathbb{M}}(p)$ such that $|\bar{\mathcal{K}}| := \text{rank}([\text{grad}\, H_{\bar{\mathcal{K}}}](q_{k_j}))$ and $|\bar{\mathcal{K}}^-| := \text{rank}([\text{grad}\, G_{\bar{\mathcal{K}}^-}](q_{k_j}))$, satisfying

$$|\mathcal{K}| < |\bar{\mathcal{K}}| + |\bar{\mathcal{K}}^-|, \tag{33}$$

and $[\operatorname{grad} H_{\bar{\mathcal{K}}}, \operatorname{grad} G_{\bar{\mathcal{K}}^-}](q_{k_j})$ is linearly independent. Hence, considering that $[h'](q_{k_j})$ is linearly independent, applying Lemma 1 and using (5) and (9), we conclude that $[h', H'_{\bar{\mathcal{K}}}, G'_{\bar{\mathcal{K}}^-}](q_{k_j})$ is also linearly independent for all $j \in \mathbb{N}$. By using Lemma 13, we obtain that $\mathcal{J}^-_{\mathbb{M}}(p) = \mathcal{J}^-(p)$. Therefore, due to $\bar{\mathcal{K}} \subset \mathcal{S}$ and $\bar{\mathcal{K}}^- \subset \mathcal{J}^-_{\mathbb{M}}(p) = \mathcal{J}^-(p)$, we have

$$t + |\bar{\mathcal{K}}| + |\bar{\mathcal{K}}^-| \leq \operatorname{rank}([h', H', G'_{\mathcal{J}^-(p)}](q_{k_j})), \qquad \forall j \in \mathbb{N}. \tag{34}$$

Considering that $p \in \Omega$ satisfies Lower-CRSC for problem (1), there exists $\epsilon > 0$ such that $\operatorname{rank}([h', H', G'_{\mathcal{J}^-(p)}](q))$ is constant for all $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$. Thus, there exist $\hat{\mathcal{K}} \subset \mathcal{S}$ and $\hat{\mathcal{K}}^- \subset \mathcal{J}^-(p)$ such that $[h', H'_{\hat{\mathcal{K}}}, G'_{\hat{\mathcal{K}}^-}](q)$ is linearly independent for all $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$, and

$$t + |\hat{\mathcal{K}}| + |\hat{\mathcal{K}}^-| = \operatorname{rank}([h', H', G'_{\mathcal{J}^-(p)}](q)), \tag{35}$$

for all $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ (perhaps decreasing ϵ if necessary). Given that $\lim_{j\to+\infty} q_{k_j} = p$, there exists $\hat{j}$ such that $q_{k_j} \in B_\epsilon(p) \cap \{p \in \mathbb{R}^n \mid h(p) = 0\}$ for all $j \geq \hat{j}$. Hence, (34) and (35) imply that

$$|\bar{\mathcal{K}}| + |\bar{\mathcal{K}}^-| \leq |\hat{\mathcal{K}}| + |\hat{\mathcal{K}}^-|. \tag{36}$$

On the other hand, since $[h', H'_{\hat{\mathcal{K}}}, G'_{\hat{\mathcal{K}}^-}](q)$ is linearly independent, applying Lemma 1 and taking into account (5) and (9), we obtain that $[\operatorname{grad} H'_{\hat{\mathcal{K}}}, \operatorname{grad} G_{\hat{\mathcal{K}}^-}](q)$ is also linearly independent. Thus, considering that $|\mathcal{K}| := \operatorname{rank}([\operatorname{grad} H, \operatorname{grad} G_{\mathcal{J}^-_{\mathbb{M}}(p)}](p))$, we have $|\hat{\mathcal{K}}| + |\hat{\mathcal{K}}^-| \leq |\mathcal{K}|$. This, together with (33), implies $|\hat{\mathcal{K}}| + |\hat{\mathcal{K}}^-| \leq |\mathcal{K}| < |\bar{\mathcal{K}}| + |\bar{\mathcal{K}}^-|$, leading to a contradiction with (36). Therefore, $p \in \Omega_{\mathbb{M}}$ satisfies CRSC for problem (15).

Reciprocally, suppose that $p \in \Omega_{\mathbb{M}}$ satisfies CRSC for problem (15). Consider $\mathcal{J}^-(p) = \{j \in \mathcal{A}(p) \mid -G'_j(p) \in \mathcal{L}(p)^\circ\}$ and assume, by contradiction, that $p \in \Omega$ does not satisfy CRSC for problem (1). Thus, for each $k \in \mathbb{N}$, there exists $q_k \in B_{1/k}(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ such that $t + |\mathcal{K}| := \operatorname{rank}([h', H', G'_{\mathcal{J}^-(p)}](q)) \neq \operatorname{rank}([h', H', G'_{\mathcal{J}^-(p)}](q_k))$. Therefore, by using Lemma 3, we conclude that there exists a subsequence $(q_{k_j})_{j\in\mathbb{N}}$ of $(q_k)_{k\in\mathbb{N}}$ such that $t + |\mathcal{K}| < \operatorname{rank}([h', H', G'_{\mathcal{J}^-(p)}](q_{k_j}))$ for all $j \in \mathbb{N}$. Considering that $\mathcal{S}$ and $\mathcal{J}^-(p)$ are finite, there exist $\bar{\mathcal{K}} \subset \mathcal{S}$ and $\bar{\mathcal{K}}^- \subset \mathcal{J}^-(p) = \mathcal{J}^-_{\mathbb{M}}(p)$ such that $t + |\mathcal{K}| < t + |\bar{\mathcal{K}}| + |\bar{\mathcal{K}}^-| := \operatorname{rank}([h', H'_{\bar{\mathcal{K}}}, G'_{\bar{\mathcal{K}}^-}](q_{k_j}))$ and $[h', H'_{\bar{\mathcal{K}}}, G'_{\bar{\mathcal{K}}^-}](q_{k_j})$ are linearly independent for all $j \in \mathbb{N}$. Thus, by applying Lemma 1 and considering (5) and (9), we conclude that $[\operatorname{grad} H_{\bar{\mathcal{K}}}, \operatorname{grad} G_{\bar{\mathcal{K}}^-}](q_{k_j})$ are also linearly independent for all $j \in \mathbb{N}$. In particular, we have

$$|\bar{\mathcal{K}}| + |\bar{\mathcal{K}}^-| \leq \operatorname{rank}([\operatorname{grad} H, \operatorname{grad} G_{\mathcal{J}^-_{\mathbb{M}}(p)}](q_{k_j})), \qquad \forall j \in \mathbb{N}. \tag{37}$$

Since $p \in \Omega_{\mathbb{M}}$ satisfies CRSC for problem (15), there exists $\epsilon > 0$ such that $\operatorname{rank}([\operatorname{grad} H, \operatorname{grad} G_{\mathcal{J}^-_{\mathbb{M}}(p)}](q))$ remains constant for all $q \in \mathbb{B}_\epsilon(p)$. Let $\hat{\mathcal{K}} \subset \mathcal{S}$ and $\hat{\mathcal{K}}^- \subset \mathcal{J}^-(p) = \mathcal{J}^-_{\mathbb{M}}(p)$ be such that $[\operatorname{grad} H_{\hat{\mathcal{K}}}, \operatorname{grad} G_{\hat{\mathcal{K}}^-}](p)$ is a basis of $\operatorname{Span}([\operatorname{grad} H, \operatorname{grad} G_{\mathcal{J}^-_{\mathbb{M}}(p)}](p))$. Thus,

$$|\hat{\mathcal{K}}| + |\hat{\mathcal{K}}^-| = \operatorname{rank}([\operatorname{grad} H, \operatorname{grad} G_{\mathcal{J}^-_{\mathbb{M}}(p)}](q)), \qquad \forall q \in \mathbb{B}_\epsilon(p). \tag{38}$$

Hence, given that $\lim_{j\to+\infty} q_{k_j} = p$, there exist $q_{k_j} \in \mathbb{B}_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$, which together with (37) and (38) imply that

$$|\bar{\mathcal{K}}| + |\bar{\mathcal{K}}^-| \leq |\hat{\mathcal{K}}| + |\hat{\mathcal{K}}^-|. \tag{39}$$

On the other hand, by applying Lemma 1 and taking into account (5) and (9), we obtain that $[h', H'_{\hat{\mathcal{K}}}, G'_{\hat{\mathcal{K}}^-}](p)$ is linearly independent. Since $t + |\mathcal{K}| := \operatorname{rank}([h', H', G'_{\mathcal{J}^-(p)}](p))$, we have $|\hat{\mathcal{K}}| + |\hat{\mathcal{K}}^-| \leq |\mathcal{K}|$, which together with (39) implies that $|\bar{\mathcal{K}}| + |\bar{\mathcal{K}}^-| \leq |\hat{\mathcal{K}}| + |\hat{\mathcal{K}}^-| \leq |\mathcal{K}|$. Considering that $|\mathcal{K}| < |\bar{\mathcal{K}}| + |\bar{\mathcal{K}}^-|$, we have a contradiction. □

As a consequence of Theorems 5, 9, 12 and 14, we obtain that Lower-RCPLD and Lower-CRSC are constraint qualifications for problem (1). In particular, all other conditions shown in Figure 1 are also constraint qualifications. More specifically, Theorem 5 translated to problem (1) gives the following:

Corollary 15. *Suppose that $p \in \Omega$ satisfies Lower-RCPLD or Lower-CRSC. If p is a Lower-AKKT point for problem* (1)*, then p is a KKT point for problem* (1)*.*

In [35], the capability of Algorithm 2 to generate AKKT sequences for problem (15) was demonstrated. This finding, supported by Theorem 9, indicates that Algorithm 2 also produces Lower-AKKT sequences for the related problem (1). It is noteworthy that a comprehensive global convergence analysis of Algorithm 2, specifically designed to address problem (15), was presented in [3]. Given that problem (15) is essentially the Riemannian version of problem (1), Algorithm 2 is applicable to solving both problem instances. The next theorem establishes that, subject to any lower strict constraint qualification in Definition 9, any feasible limit point of the sequence generated by Algorithm 2 is a KKT point for problem (1). It is important to highlight that the sequence $(p^k)_{k\in\mathbb{N}}$ generated by Algorithm 2 is feasible for the constraint $\{q \in \mathbb{R}^n \mid h(q) = 0\}$.

Theorem 16. *Let $p \in \Omega$ be a limit point of a sequence $(p^k)_{k\in\mathbb{N}}$ generated by Algorithm 2. Assume that p satisfies Lower-RCPLD or Lower-CRSC. Then, p satisfies the KKT conditions for problem* (1)*.*

Proof. Let $p \in \Omega$ satisfy Lower-RCPLD (respectively Lower-CRSC). By Theorem 12 (respectively, Theorem 14), it follows that $p \in \Omega_{\mathbb{M}}$ and also satisfies RCPLD (respectively, CRSC). According to [35, Theorem 3], p is an AKKT point for problem (15). Using Theorem 9, we conclude that p is a Lower-AKKT point for problem (1). Therefore, by Corollary 15, p is a KKT point for problem (1). □

The next theorem establishes the connection between Lower-QN and QN for problem (15).

Theorem 17. *A point $p \in \Omega$ satisfies Lower-QN for problem* (1) *if and only if $p \in \Omega_{\mathbb{M}}$ satisfies QN for problem* (15)*.*

Proof. First, assume that $p \in \Omega$ satisfies Lower-QN for problem (1). Assume, by contradiction, that $p \in \Omega_{\mathbb{M}}$ does not satisfies QN for problem (15). Since $\Omega = \Omega_{\mathbb{M}}$, we have $p \in \Omega_{\mathbb{M}}$, and item *(i)* of Definition 7 implies that there exist $\lambda \in \mathbb{R}^s$ and $\mu \in \mathbb{R}^m_+$ such that

$$\sum_{i=1}^{s} \lambda_i \operatorname{grad} H_i(p) + \sum_{j\in\mathcal{A}_{\mathbb{M}}(p)} \mu_j \operatorname{grad} G_j(p) = 0.$$

Given that $\mathcal{A}_{\mathbb{M}}(p) = \mathcal{A}(p)$, and using similar arguments as in the proof of Theorem 9, we can conclude that there exist $(\eta, \lambda, \mu) \in \mathbb{R}^t \times \mathbb{R}^s \times \mathbb{R}^m_+$ such that

$$\sum_{i=1}^{t} \eta_i h_i'(p) + \sum_{i=1}^{s} \lambda_i H_i'(p) + \sum_{j\in\mathcal{A}(p)} \mu_i G_i'(p) = 0. \tag{40}$$

Additionally, considering item *(ii)* of Definition 7 and the fact that $\mathcal{A}_{\mathbb{M}}(p) = \mathcal{A}(p)$, we obtain

$$\mu_j = 0, \quad \forall j \notin \mathcal{A}(p), \quad (\eta, \lambda, \mu) \neq 0. \tag{41}$$

Finally, take $\epsilon > 0$. It follows from item *(iii)* of Definition 7 that, for each $k \in \mathbb{N}$, there exists $q_k \in \mathbb{B}_{1/k}(p)$ such that $\lambda_i H_i(q_k) > 0$ for all $i \in \mathcal{S}$ with $\lambda_i \neq 0$, and $\mu_j G_j(q_k) > 0$ for all $j \in \mathcal{A}_{\mathbb{M}}(p)$ with $\mu_j > 0$. Given that $\lim_{k\to+\infty} q_k = p$, there exists $\hat{k} \in \mathbb{N}$ such that $q_k \in B_\epsilon(p) \cap \{p \in \mathbb{R}^n \mid h(p) = 0\}$ for all $k \geq \hat{k}$. Hence, considering that $\mathcal{A}_{\mathbb{M}}(p) = \mathcal{A}(p)$, it follows that there exists $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ such that $\lambda_i H_i(q) > 0$ for all $i \in \mathcal{S}$ with $\lambda_i \neq 0$, and $\mu_j G_j(q) > 0$ for all $j \in \mathcal{A}(p)$ with $\mu_j > 0$. This, together with (40) and (41), contradicts Definition 11. Therefore, $p \in \Omega_{\mathbb{M}}$ satisfies QN for problem (15).

Reciprocally, suppose that $p \in \Omega_{\mathbb{M}}$ satisfies QN for problem (15). Assume, by contradiction, that $p \in \Omega$ does not satisfy Lower-QN for problem (1). According to item *(i)* of Definition 11, there exist $\eta \in \mathbb{R}^t$, $\lambda \in \mathbb{R}^s$, and $\mu \in \mathbb{R}^m_+$ such that

$$\sum_{i=1}^{t} \eta_i h_i'(p) + \sum_{i=1}^{s} \lambda_i H_i'(p) + \sum_{j\in\mathcal{A}(p)} \mu_i G_i'(p) = 0. \tag{42}$$

Given that $\mathcal{A}_{\mathbb{M}}(p) = \mathcal{A}(p)$, and using similar arguments as in the proof of Theorem 9, we can conclude that there exist $\lambda \in \mathbb{R}^s$ and $\mu \in \mathbb{R}^m_+$ such that

$$\sum_{i=1}^{s} \lambda_i \operatorname{grad} H_i(p) + \sum_{j \in \mathcal{A}_{\mathbb{M}}(p)} \mu_j \operatorname{grad} G_j(p) = 0. \tag{43}$$

Additionally, item *(ii)* of Definition 11 implies that $\mu_j = 0$ for all $j \notin \mathcal{A}(p) = \mathcal{A}_{\mathbb{M}}(p)$ and $(\eta, \lambda, \mu) \neq 0$. Moreover, since $\{h_i'(q) \mid i = 1, \ldots, t\}$ is linearly independent, (42) implies that if $\lambda = \mu = 0$ then $\eta = 0$, which means $(\eta, \lambda, \mu) = 0$. Given that $(\eta, \lambda, \mu) \neq 0$ we have $(\lambda, \mu) \neq 0$. Thus,

$$\mu_j = 0, \quad \forall j \notin \mathcal{A}_{\mathbb{M}}(p), \quad (\lambda, \mu) \neq 0. \tag{44}$$

Now, take $\epsilon > 0$. It follows from item *(iii)* of Definition 11 that for each $k \in \mathbb{N}$, there exists $q_k \in B_{1/k}(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ such that $\lambda_i H_i(q_k) > 0$ for all $i \in \mathcal{S}$ with $\lambda_i \neq 0$, and $\mu_j G_j(q_k) > 0$ for all $j \in \mathcal{A}(p)$ with $\mu_j > 0$. Since $\mathcal{A}_{\mathbb{M}}(p) = \mathcal{A}(p)$, $B_{1/k}(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ is an open subset of $\mathbb{M}$, and $\lim_{k \to +\infty} q_k = p$, there exists $\bar{k} \in \mathbb{N}$ such that $q_{\bar{k}} \in \mathbb{B}_\epsilon(p) \cap B_{1/\bar{k}}(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$, with $\lambda_i H_i(q_{\bar{k}}) > 0$ for all $i \in i \in \mathcal{S}$ with $\lambda_i \neq 0$, and $\mu_j G_j(q_{\bar{k}}) > 0$ for all $j \in \mathcal{A}_{\mathbb{M}}(p)$ with $\mu_j > 0$. This, together with (43) and (44), contradicts Definition 7. Therefore, $p \in \Omega$ satisfies Lower-QN for problem (1). □

Theorem 18. *Let $p \in \Omega$ be a Lower-PAKKT point with associated primal sequence $(p^k)_{k \in \mathbb{N}}$ and dual sequence $(\eta^k, \lambda^k, \mu^k)_{k \in \mathbb{N}}$. Assume that p satisfies Lower-QN. Then, $(\eta^k, \lambda^k, \mu^k)_{k \in \mathbb{N}}$ is a bounded sequence. In particular, p satisfies the KKT conditions, and any limit point of $(\eta^k, \lambda^k, \mu^k)_{k \in \mathbb{N}}$ is a Lagrange multiplier associated with p.*

Proof. Using Theorems 6, 9 and 17, it follows that $(\lambda^k, \mu^k)_{k \in \mathbb{N}}$ is bounded. Assume, by contradiction, that (η^k) is unbounded. Let $K_1 \subset \mathbb{N}$ be an infinite set and $\eta \in \mathbb{R}^t$ with $\|\eta\|_2 = 1$, such that $\lim_{k \in K_1} \|\eta^k\|_2 = +\infty$ and $\lim_{k \in K_1} (\eta^k / \|\eta^k\|_2) = \eta$. Since $p \in \Omega_{\mathbb{M}}$ is Lower-PAKKT with associated primal sequence $(p^k)_{k \in \mathbb{N}}$ and dual sequence $(\eta^k, \lambda^k, \mu^k)_{k \in \mathbb{N}}$, we have $\lim_{k \to \infty} L'(p^k, \eta^k, \lambda^k, \mu^k) = 0$. Therefore,

$$\lim_{k \to \infty} \sum_{i=1}^{t} \frac{\eta_i^k}{\|\eta^k\|} h_i'(p^k) + \sum_{i=1}^{s} \frac{\lambda_i^k}{\|\eta^k\|} H_i'(p^k) + \sum_{j \in \mathcal{A}(p)} \frac{\mu_j^k}{\|\eta^k\|} G_j'(p^k) = 0.$$

Consequently, we obtain $\sum_{i=1}^{t} \eta_i h_i'(p) = 0$, with $\eta \neq 0$, contradicting assumption (H1). Thus, $(\eta^k, \lambda^k, \mu^k)_{k \in \mathbb{N}}$ is a bounded sequence. □

Next, we show that Algorithm 2 produces Lower-PAKKT sequences, thereby establishing its global convergence under the Lower-QN condition.

Theorem 19. *Assume that Algorithm 2 generates an infinite sequence $(p^k)_{k \in \mathbb{N}}$ with a feasible limit $p \in \Omega$, say, $\lim_{k \in K} p^k = p$. Then, p is a Lower-PAKKT point with the correspondent primal sequence $(p^k)_{k \in K}$ and dual sequence $(\eta^k, \lambda^k, \mu^k)_{k \in K}$, where $(\eta^k)_{k \in K}$ can be determined from $(\lambda^k, \mu^k)_{k \in K}$ which is generated by the algorithm. In particular, p is a KKT point, and any limit point of $(\eta^k, \lambda^k, \mu^k)_{k \in K}$ is a Lagrange multiplier associated with p.*

Proof. By Theorem 7, we obtain that p is a PAKKT point with the correspondent primal sequence $(p^k)_{k \in K}$ and dual sequence $(\lambda^k, \mu^k)_{k \in K}$ as generated by Algorithm 2. Using Theorem 9, we conclude that p is a Lower-PAKKT point. Furthermore, similar to the proof of Theorem 9, we find that $(\eta^k)_{k \in K}$ can be determined from $(\lambda^k, \mu^k)_{k \in K}$. □

In [3], the authors introduced the concept of Scaled-PAKKT point for problem (15). Essentially, it coincides with the definition of PAKKT with the condition $\lim_{k \to \infty} \operatorname{grad} \mathbb{L}(p^k, \lambda^k, \mu^k)/\gamma_k = 0$, where $\gamma_k := \left\|(1, \lambda^k, \mu^k)\right\|_\infty$, replacing item *(ii)* in Definition 5. We conclude this section by noting that, similarly to Lower-PAKKT, we can also define the concept of *Lower-Scaled-PAKKT* and establish its connection with Scaled-PAKKT. It was demonstrated in [3] that Algorithm 2 is capable of generating Scaled-PAKKT sequences by ensuring $\|\text{grad}\,\mathbb{L}_{\rho_k}(p^k, \bar\lambda^k, \bar\mu^k)/\gamma_k\| \leq \epsilon_k$, rather than (18) in Step 1. Moreover, QN ensure that the dual Scaled-PAKKT sequence is bounded. Therefore, Theorems 9 and 17 imply that Lower-QN is sufficient to guarantee that the dual Lower-Scaled-PAKKT sequence is bounded.

5 Numerical experiments

This section presents numerical results to illustrate the practical advantages of exploiting Riemannian techniques in solving certain classes of optimization problems. The experiments were conducted in MATLAB version 9.11.0 (R2021b), on a computer with a 3.7 GHz Intel Core i5 6-Core processor and 8GB 2667MHz DDR4 RAM, running macOS Sonoma 14.4.1. All codes are available at `https://github.com/lfprudente/RiemannianAL`. We compare the performance of the Riemannian and Euclidean safeguarded augmented Lagrangian methods as follows:

- **Riemannian-AL** (Riemannian augmented Lagrangian): Algorithm 2 with Manopt [18] to solve the subproblem in Step 1. We use the RLBFGS solver, a Riemannian limited memory BFGS algorithm [26].
- **Euclidian-AL** (Euclidian augmented Lagrangian): Algorithm 1 with ASA [24] to solve the subproblem in Step 1. In the Euclidian version, boxes are treated as lower-level constraints. ASA is an active set algorithm coded in C for box-constrained optimization that combines a gradient projection algorithm [21] and the conjugate gradient algorithm, as implemented in the code CG_DESCENT [22, 23]. We use the MATLAB interface provided in [12].

Given tolerances $\varepsilon_{\text{opt}} > 0$, $\varepsilon_{\text{feas}} > 0$, and $\varepsilon_{\text{compl}} > 0$ for optimality, feasibility, and complementarity, respectively, the Riemannian-AL (resp. Euclidian-AL) algorithm stops successfully at iteration k with $(p^k, \lambda^k, \mu^k) \in \mathbb{M} \times \mathbb{R}^s \times \mathbb{R}^m_+$ (resp. $(p^k, \eta^k, \lambda^k, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^t \times \mathbb{R}^s \times \mathbb{R}^m_+$) when:

$$\|\,\text{grad}\,\mathbb{L}(p^k, \lambda^k, \mu^k)\| \leq \varepsilon_{\text{opt}} \quad (\text{resp. } \|L'(p^k, \eta^k, \lambda^k, \mu^k)\| \leq \varepsilon_{\text{opt}}),$$

$$\max\{\|H(p^k)\|_\infty, \|G(p^k)_+\|_\infty\} \leq \varepsilon_{\text{feas}} \quad (\text{resp. } \max\{\|h(p^k)\|_\infty, \|H(p^k)\|_\infty, \|G(p^k)_+\|_\infty\} \leq \varepsilon_{\text{feas}}),$$

$$\min\{-G_i(p^k), \mu_i^k\} \leq \varepsilon_{\text{compl}}, \quad \text{for all} \quad i = 1, \ldots, m.$$

These conditions correspond to the approximate fulfillment of the KKT conditions. We also stopped the execution of the algorithms if the penalty parameter became too large ($\rho_k > 10^{20}$) or if the algorithms exceeded the maximum number of 50 outer iterations allowed. These two criteria are related to failures. For both algorithms, we used the following parameters: $\tau = 0.5$, $\gamma = 10$, $\lambda_{\min} = -10^{20}$, $\lambda_{\max} = \mu_{\max} = 10^{20}$, $\bar\lambda^1 = \bar\mu^1 = 0$, $\varepsilon_{\text{opt}} = \varepsilon_{\text{feas}} = \varepsilon_{\text{compl}} = 10^{-5}$, $\varepsilon_k = \max\{\varepsilon_{\text{opt}}, \sqrt{\varepsilon_{\text{opt}}}/10^{k-1}\}$ for all $k \geq 1$, and

$$\rho_1 = \max\left\{10^{-8}, \min\left\{10\frac{\max\{1, f(p^0)\}}{\max\{1, \|h(p^0)\|_2^2 + \|H(p^0)\|_2^2 + \|G(p^0)_+\|_2^2\}}, 10^8\right\}\right\},$$

as suggested in [14]. For each experiment, both solvers used the same starting point. In the following subsections, we describe three applications that can be modeled as problem (15), along with the corresponding results.

5.1 Greediness phenomenon

The *greediness* phenomenon is the tendency of some nonlinear programming methods to find highly infeasible points with very small functional values, typically in the initial iterations, see [20]. The phenomenon can occur in problems for which the objective function assumes significantly lower values at infeasible points than in the feasible region. In an augmented Lagrangian algorithm, unconstrained minimizers may attract the iterates at early stages of the calculations, causing the penalty parameter to grow excessively. This excessive growth leads to ill-conditioning, which harms the overall convergence. Consider the following example [20]:

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad -\sum_{i=1}^{n}(x_i^8 + x_i), \quad \text{subject to} \quad \|x\|^2 = 1, \quad x_2 + \sum_{i=1}^{n} x_i \leq 0. \tag{45}$$

Since $\|x\|^2 = 1$ defines the unit sphere $\mathbb{S}^{n-1} \subset \mathbb{R}^n$, problem (45) can be rewritten as:

$$\underset{x \in \mathbb{S}^{n-1}}{\text{Minimize}} \quad -\sum_{i=1}^{n}(x_i^8 + x_i), \quad \text{subject to} \quad x_2 + \sum_{i=1}^{n} x_i \leq 0. \tag{46}$$

We applied Euclidian-AL and Riemannian-AL algorithms to problems (45) and (46), respectively. We set $n = 50$ and randomly generated 100 starting points on the sphere $\mathbb{S}^{n-1}$. The Euclidian-AL algorithm failed in all instances, typically generating the first iterate with $\|x^1\|_\infty \approx 10^{43}$ and $f(x^1) = -\infty$. From there, two situations were observed: either NaNs were generated, resulting in algorithm failure, or the penalty parameter grew beyond the maximum allowed. In contrast, the Riemannian-AL algorithm successfully solved the problem in all instances. The maximum number of iterations was 6, and the greatest final penalty parameter was less than 20. For illustrative purposes, Figure 2 shows the behavior of the Riemannian-AL algorithm in the case where $n = 2$, $x^0 = (\sqrt{2}/2, \sqrt{2}/2)$, and $\rho_1 \approx 7$. The Riemannian-AL algorithm converged to the global solution $x^* = (2\sqrt{5}/5, -\sqrt{5}/5)$ in 4 iterations. Visually, the second iterate x^2 is virtually identical to the solution x^*. In contrast, as for the larger instances, the Euclidian-AL algorithm failed to solve the problem.

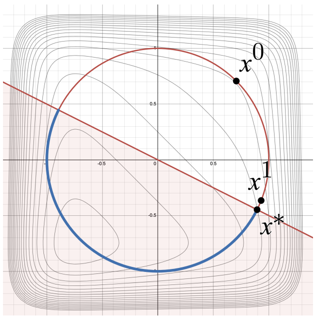

Figure 2: Iterates generated by the Riemannian-AL algorithm for problem (46) with $n = 2$. The feasible set is highlighted in blue and the level sets of the objective function are given in light grey. The algorithm converges in 4 iterations, with the second iterate x^2 closely coinciding with the global solution x^*.

We conclude that when the greediness is associated with the manifold $\mathbb{M}$, the Riemannian algorithm uses the geometric structure to maintain feasibility with respect to $\mathbb{M}$, control the penalty parameter, and improve the convergence of the augmented Lagrangian method.

5.2 Packing circles within ellipses

The circle packing problem involves finding the maximum radius r of N identical circles that can be fitted without overlapping into a two-dimensional fixed-size container [32]. This problem has a wide range of applications, as discussed in [32, 15] and the references therein. In this section, we consider the container to be an ellipse with semi-axes $a \geq b > 0$. Using continuous variables $(r; u, v, s) \in \mathbb{R} \times (\mathbb{R}^n)^3$, this problem can be modeled [15] as follows:

$$\begin{array}{ll}
\underset{(r;u,v,s) \in \mathbb{R} \times (\mathbb{R}^n)^3}{\text{Maximize}} & r \\
\text{subject to} & u_i^2 + v_i^2 = 1, \quad \forall i = 1, \ldots, N, \\
& b^2(s_i - 1)^2[(b^2/a^2)u_i^2 + v_i^2] \geq r^2, \quad \forall i = 1, \ldots, N, \\
& a^2\left[(1 + (s_i - 1)(b^2/a^2))u_i - (1 + (s_j - 1)(b^2/a^2))u_j\right]^2 + b^2(s_i v_i - s_j v_j)^2 \geq 4r^2, \quad \forall i < j, \\
& 0 \leq s_i \leq 1, \quad \forall i = 1, \ldots, N, \\
& r \geq 0.
\end{array} \tag{47}$$

The Cartesian coordinates (x_i, y_i) of the circles' centers can be recovered using:

$$x_i = a[1 + (s_i - 1)(b^2/a^2)]u_i, \quad y_i = bs_iv_i, \quad \forall i = 1, \ldots, N.$$

Since, for all $i = 1, \ldots, N$, the constraint $u_i^2 + v_i^2 = 1$ defines the unit circle $\mathbb{S}^1 \subset \mathbb{R}^2$, problem (47) can be written in the format of (15) by taking $(r; (u, v); s) \in \mathbb{M} := \mathbb{R} \times (\mathbb{S}^1)^n \times \mathbb{R}^n$ and omitting the first group of constraints.

We considered eight instances of the problems with $(a, b) = (2, 1)$ and $N \in \{3, 10, 20, 30, 40, 50, 80, 100\}$. The starting points $(r^0; (u^0, v^0); s^0)$ were randomly generated in $[0, 1] \times (\mathbb{S}^1)^n \times [0, 1]^n$. Table 1 shows the results. In the table, "N" is the number of circles to be packed, "n" is the number of variables and "#c." is the number of constraints. For the Riemannian-AL algorithm, the manifold constraints are excluded from the count, and for the Euclidean-AL algorithm, the box constraints are excluded. "It" is the number of augmented Lagrangian iterations, "Obj." is the final optimal radius, "Feas." is the feasibility measure at the final iterate p^* given by $\max\{\|h(p^*)\|_\infty, \|H(p^*)\|_\infty, \|G(p^*)_+\|_\infty\}$, and "Time" is the CPU time in seconds. The best reported data for each instance is highlighted in bold. Figure 3 illustrates the "solutions" found by the Riemannian-AL algorithm.

		Riemannian-AL					Euclidian-AL				
N	n	#c.	It.	Obj.	Feas.	Time	#c.	It.	r	Feas.	Time
3	10	13	**5**	**6.667e-01**	**5.0e-07**	1.0	9	**5**	**6.667e-01**	3.0e-06	**0.2**
10	31	76	8	**3.793e-01**	**1.9e-06**	9.7	65	**5**	3.638e-01	3.1e-06	**3.6**
20	61	251	**6**	**2.751e-01**	**2.7e-06**	27.4	230	13	2.747e-03	7.5e-06	**5.7**
30	91	526	**6**	**2.254e-01**	1.0e-06	**136.3**	495	7	2.238e-01	**8.8e-07**	184.2
40	121	901	**9**	**1.977e-01**	5.1e-06	315.0	860	14	2.096e-03	**4.3e-06**	**24.9**
50	151	1376	8	1.757e-01	**2.5e-06**	**587.3**	1325	**6**	**1.776e-01**	6.9e-06	815.6
80	241	3401	**8**	**1.401e-01**	9.7e-06	1235.2	3320	13	2.659e-03	**7.0e-06**	**103.9**
100	301	5251	**6**	**1.262e-01**	5.0e-06	1794.4	5150	12	1.905e-03	**3.6e-06**	**347.8**

Table 1: Performance of Riemannian-AL and Euclidean-AL algorithms for packing circles within an ellipse.

As can be seen, the Riemannian-AL algorithm consistently outperformed the Euclidean-AL algorithm in terms of the final optimal radius. For larger instances, the Euclidean-AL algorithm often encountered issues where the centers of some circles moved to the ellipse boundary, resulting in a significantly smaller radius (r in the order of 10^{-3}). This phenomenon led to poor local minimizers and was never observed with the Riemannian-AL algorithm. For example, when packing 20 circles, the Euclidean-AL algorithm required 15 different starting points, taking a total of 183.1 seconds to find a solution with r in the order of 10^{-1}. Similarly, for packing 40 circles, the Euclidean-AL algorithm needed 7 initial points, taking a total of 1060.7 seconds. Notably, for packing 80 circles, the Euclidean-AL algorithm used 100 starting points and spent over 13 hours without finding a solution with a final radius r in the order of 10^{-1}. These results show that while the Riemannian-AL algorithm efficiently solves all instances without needing multiple starting points, the Euclidean-AL algorithm frequently requires several initial points to find good solutions, resulting in more computational time.

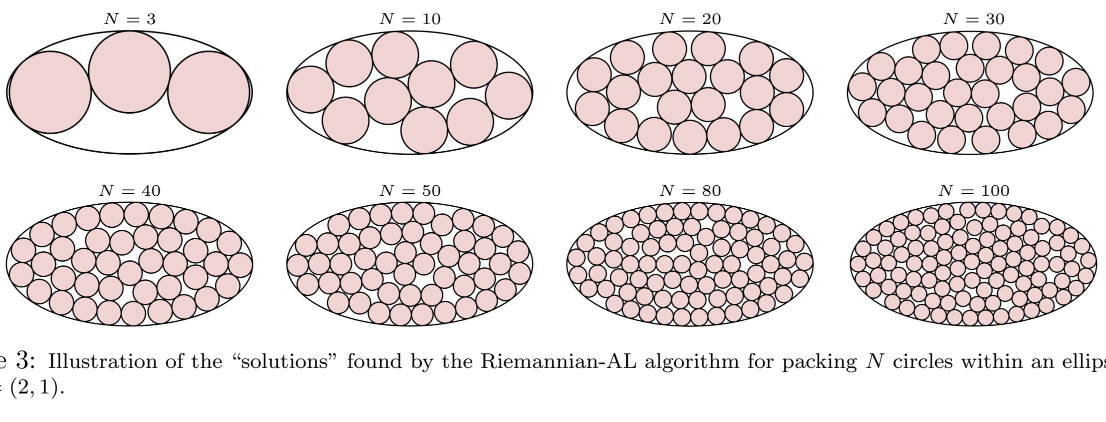

Figure 3: Illustration of the "solutions" found by the Riemannian-AL algorithm for packing N circles within an ellipse with $(a, b) = (2, 1)$.

5.3 k-Means clustering

Given a set of N data points, the k-means clustering problem involves partitioning these points into k clusters, with the goal of minimizing the sum of squared distances between each data point and the centroid of its corresponding cluster. This process helps in grouping similar data points together, uncovering underlying patterns or structures within the data. We refer to [1] for numerous applications across various domains. Let $\mathcal{P} := \{x_1, \ldots, x_N\} \subset \mathbb{R}^\ell$ represent the given data points, and denote the clusters by $\mathcal{C}_1, \ldots, \mathcal{C}_k \subset \mathcal{P}$. The k-means clustering problem can be formulated as follows:

$$\underset{\mathcal{C}_1,\ldots,\mathcal{C}_k}{\text{Minimize}} \quad \sum_{j=1}^{k} \sum_{x_i \in \mathcal{C}_j} \|x_i - \mu_j\|^2, \quad \text{subject to} \quad \mathcal{P} = \bigcup_{j=1}^{k} \mathcal{C}_j, \quad \mathcal{C}_i \cap \mathcal{C}_j = \emptyset,\ \forall i \neq j,$$

where $\mu_j := \frac{1}{|\mathcal{C}_j|} \sum_{x_i \in \mathcal{C}_j} x_i$ and $|\mathcal{C}_j|$ is the cardinality of $\mathcal{C}_j$. According to [19], this problem can be equivalently reformulated as a continuous optimization problem with nonnegative orthogonality constraints, expressed as follows:

$$\underset{Y \in \mathbb{R}^{N\times k}}{\text{Minimize}} \quad -\operatorname{tr}(Y^\top D Y), \quad \text{subject to} \quad Y^\top Y = I, \quad Y \geq 0, \quad YY^\top e = e, \tag{48}$$

where $D := (D_{ij}) \in \mathbb{R}^{N\times N}$ with $D_{ij} = x_i^\top x_j$ for all $i, j = 1, \ldots, N$, I is the N-dimensional identity matrix, the inequality $Y \geq 0$ is component-wise, and $e \in \mathbb{R}^N$ is the vector of ones. The constraint $Y^\top Y = I$ is the Stiefel manifold embedded in the $N \times k$ real matrix space, denoted by $St_{N,k} := \{Y \in \mathbb{R}^{N\times k} \mid Y^\top Y = I\}$. Thus, problem (48) can be written in the format of (15) by taking $Y \in St_{N,k}$ and omitting the constraint $Y^\top Y = I$. A feasible point $Y \in \mathbb{R}^{N\times k}$ has exactly one non-negative entry per row and all non-zero entries of a column are equal. A point x_i is assigned to cluster $\mathcal{C}_j$ if $Y_{ij} \neq 0$. This property ensures that the clustering structure can be directly recovered from the solution matrix Y.

As in [30], we considered some datasets from the UCI Machine Learning Repository [29]. The main characteristics of the considered problems are described in Table 2. The table also shows the number of variables related to problem (48), and the number of constraints of its Riemannian (#c(Riem.)) and Euclidean (#c(Eucl.)) versions. The starting points were generated corresponding to a random cluster classification. For all k-means problems, we used $\varepsilon_{\text{opt}} = 10^{-4}$ in the stopping criterion.

Problem name	Number of datas (N)	Features (ℓ)	Clusters (k)	n	#c(Riem.)	#c(Eucl.)
Breast cancer	569	30	2	1138	1707	572
Cloud	2048	10	2	4096	6144	2051
Ecoli	336	7	8	2688	3024	372
Ionosphere	351	34	2	702	1053	354
Iris	150	4	3	450	600	156
Parkinsons	195	22	2	390	585	198
Pima diabetes	768	8	2	1536	2304	771
Raisin	900	7	2	1800	2700	903
Seeds	210	7	3	630	840	216
SPECTF	267	44	2	534	801	270
Thyroid	215	5	3	645	860	221
Transfusion	748	4	2	751	2244	1496
Wine	178	13	3	534	712	184

Table 2: Main characteristics of the considered k-means problems.

Table 3 shows the results organized similarly to Table 1, with the addition of the "Acc." column, which reports the accuracy of correct classifications. The results indicate that the Riemannian-AL algorithm consistently produces higher clustering accuracy (Acc.) compared to the Euclidean-AL algorithm. Despite generally requiring more computational time, the Riemannian-AL algorithm achieves lower objective values (Obj.) and maintains a high level of feasibility (Feas.). The Euclidean-AL algorithm often faced convergence difficulties and failed to provide results for several datasets. Specifically, for the Cloud, Iris, Pima diabetes, Raisin, and SPECTF problems, the Euclidean-AL algorithm got stuck at infeasible points that are stationary for an infeasibility measure, leading to excessive increases in the penalty parameter until the algorithm failed. The convergence to such infeasible points impairs the overall performance of an augmented Lagrangian algorithm and is often used as a stopping criterion related to failure, see [14].

	Riemannian-AL					Euclidian-AL				
Problem	It.	Obj.	Feas.	Time	Acc.(%)	It.	Obj.	Feas.	Time	Acc. (%)
Breast cancer	**9**	**-2.732e+03**	2.2e-06	59.4	**91.0**	**11**	**-2.732e+03**	**1.7e-06**	**14.8**	**91.0**
Cloud	**13**	**-4.274e+03**	**9.9e-07**	**178.2**	**100.0**	-	-	-	-	-
Ecoli	9	**-9.046e+02**	**2.8e-06**	487.8	**65.8**	**5**	-8.820e+02	3.1e-06	**25.9**	55.1
Ionosphere	**8**	**-1.134e+03**	**3.5e-07**	29.0	**71.2**	9	**-1.134e+03**	3.2e-06	**9.6**	**71.2**
Iris	**9**	**-2.276e+02**	**2.9e-06**	**42.7**	**83.3**	-	-	-	-	-
Parkinsons	6	**-7.969e+06**	**2.3e-07**	44.1	**75.4**	**5**	-7.479e+06	1.1e-06	**1.0**	51.3
Pima diabetes	**9**	**-5.070e+02**	**1.0e-06**	**139.4**	**67.4**	-	-	-	-	-
Raisin	**10**	**-1.449e+03**	**2.8e-06**	**98.7**	**76.8**	-	-	-	-	-
Seeds	11	-5.158e+02	7.4e-06	80.4	91.4	**7**	**-5.172e+02**	**7.0e-06**	**12.2**	**92.4**
SPECTF	**8**	**-1.377e+03**	**4.3e-06**	**28.6**	**66.3**	-	-	-	-	-
Thyroid	**8**	**-3.049e+02**	**1.6e-06**	**56.4**	**87.4**	17	-1.573e+02	9.2e-06	**20.1**	77.2
Transfusion	**7**	**-1.160e+09**	**6.0e-07**	23.6	**73.9**	**7**	**-1.160e+09**	1.1e-06	**13.4**	**73.9**
Wine	9	**-5.151e+02**	**9.0e-07**	49.0	**96.6**	**8**	**-5.151e+02**	9.8e-06	**3.1**	**96.6**

Table 3: Performance of Riemannian-AL and Euclidean-AL algorithms for the k-means clustering problem.

We conclude this section by illustrating the "solution" found by the Riemannian-AL algorithm in a synthetic k-means problem with $\ell = 2$. We generated 500 randomly perturbed data points around the reference points $(3, 3)$, $(-3, -3)$, and $(6, -6)$, resulting in a problem with 1500 variables and 2000 constraints. The Riemannian-AL algorithm took 8 iterations and a total of 103.3 seconds to find a solution with Obj $= -9.919 \times 10^3$ and Feas $= 1.7 \times 10^{-8}$. Figure 4 shows the algorithm's success in correctly clustering the dataset.

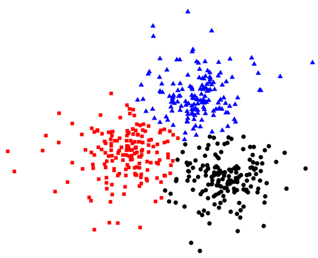

Figure 4: Illustration of the "solution" found by the Riemannian-AL algorithm in a synthetic k-means problem with $\ell = 2$.

6 Conclusions

This paper examines the use of augmented Lagrangian methods for solving constrained nonlinear programming problems involving both equality and inequality constraints, with an emphasis on lower and upper-level constraints. We introduce and analyze lower strict constraint qualifications (Lower-SCQs) within this framework, demonstrating their reduced restrictiveness compared to traditional constraint qualifications. By applying Riemannian Geometry, we connect Lower-SCQs with their Riemannian counterparts, thereby enhancing the theoretical foundation of optimization on Riemannian manifolds. Our study reveals significant theoretical advancements and practical implications, including the introduction of new sequential optimality conditions in the safeguarded augmented Lagrangian algorithm. This algorithm generates Lower-PAKKT sequences that adhere to manifold constraints and ensure all limit points satisfy the KKT conditions under any Lower-SCQ. This finding highlights the robustness of our framework and the advantages of Riemannian optimization methods. The comparison between classic and Riemannian versions of the algorithm reveals that the intrinsic approach often outperforms the extrinsic one in certain cases. This advocates for the use of Riemannian methods in specific optimization problems and suggests new research directions beyond Euclidean spaces. In conclusion, this work advances both the theory and practical applications of nonlinear programming, emphasizing the dynamic nature of optimization research and encouraging further investigation into Riemannian methods across various theoretical and practical settings.

References

[1] M. Ahmed, R. Seraj, and S. M. S. Islam. The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8), 2020.

[2] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. On augmented Lagrangian methods with general lower-level constraints. *SIAM J. Optim.*, 18(4):1286–1309, 2007.

[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 J. Optim.*, 34(2):1799–1825, 2024.

[4] R. Andreani, N. S. Fazzio, M. L. Schuverdt, and L. D. Secchin. A sequential optimality condition related to the quasi-normality constraint qualification and its algorithmic consequences. *SIAM J. Optim.*, 29(1):743–766, 2019.

[5] R. Andreani, G. Haeser, and J. M. Martínez. On sequential optimality conditions for smooth constrained optimization. *Optimization*, 60(5):627–641, 2011.

[6] R. Andreani, G. Haeser, M. L. Schuverdt, L. D. Secchin, and P. J. S. Silva. On scaled stopping criteria for a safeguarded augmented Lagrangian method with theoretical guarantees. *Math. Program. Comput.*, 14(1):121–146, 2022.

[7] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva. A relaxed constant positive linear dependence constraint qualification and applications. *Math. Program.*, 135(1-2, Ser. A):255–273, 2012.

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

[9] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva. Strict constraint qualifications and sequential optimality conditions for constrained optimization. *Math. Oper. Res.*, 43(3):693–717, 2018.

[10] R. Andreani, J. M. Martinez, and M. L. Schuverdt. On the relation between constant positive linear dependence condition and quasinormality constraint qualification. *J. Optim. Theory Appl.*, 125(2):473–485, 2005.

[11] M. S. Bazaraa and C. M. Shetty. *Nonlinear programming.* John Wiley & Sons, New York-Chichester-Brisbane, 1979. Theory and algorithms.

[12] S. Becker. mex interface for bound constrained optimization via ASA. https://www.mathworks.com/matlabcentral/fileexchange/35814-mex-interface-for-bound-constrained-optimization-via-asa, 2024. Retrieved June 10, 2024.

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

[14] E. Birgin and J. Martínez. *Practical Augmented Lagrangian Methods for Constrained Optimization.* SIAM, Philadelphia, 2014.

[15] E. G. Birgin, L. H. Bustamante, H. F. Callisaya, and J. M. Martínez. Packing circles within ellipses. *Int. Trans. Oper. Res.*, 20(3):365–389, 2013.

[16] E. G. Birgin, G. Haeser, and A. Ramos. Augmented Lagrangians with constrained subproblems and convergence to second-order stationary points. *Comput. Optim. Appl.*, 69(1):51–75, 2018.

[17] E. G. Birgin and J. M. Martínez. Complexity and performance of an augmented Lagrangian algorithm. *Optim. Methods Softw.*, 35(5):885–920, 2020.

[18] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre. Manopt, a Matlab toolbox for optimization on manifolds. *J. Mach. Learn. Res.*, 15(42):1455–1459, 2014.

[19] T. Carson, D. G. Mixon, S. Villar, and R. Ward. Manifold optimization for k-means clustering. In *2017 International Conference on Sampling Theory and Applications (SampTA)*, pages

73–77, 2017.

[20] E. V. Castelani, A. L. M. Martinez, J. M. Martínez, and B. F. Svaiter. Addressing the greediness phenomenon in nonlinear programming by means of proximal augmented lagrangians. *Comput. Optim. Appl.*, 46(2):229–245, Jun 2010.

[21] Y.-H. Dai, W. W. Hager, K. Schittkowski, and H. Zhang. The cyclic Barzilai-–Borwein method for unconstrained optimization. *IMA J. Numer. Anal.*, 26(3):604–627, 07 2006.

[22] W. W. Hager and H. Zhang. A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM J. Optim.*, 16(1):170–192, 2005.

[23] W. W. Hager and H. Zhang. Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent. *ACM Trans. Math. Softw.*, 32(1):113–137, mar 2006.

[24] W. W. Hager and H. Zhang. A new active set algorithm for box constrained optimization. *SIAM J. Optim.*, 17(2):526–557, 2006.

[25] M. R. Hestenes. *Optimization theory: The finite dimensional case.* Pure and Applied Mathematics. Wiley-Interscience [John Wiley & Sons], New York-London-Sydney, 1975.

[26] W. Huang, K. A. Gallivan, and P.-A. Absil. A broyden class of quasi-newton methods for riemannian optimization. *SIAM J. Optim.*, 25(3):1660–1685, 2015.

[27] R. Janin. Directional derivative of the marginal function in nonlinear programming. *Math. Programming Stud.*, 21:110–126, 1984.

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

[29] M. Kelly, R. Longjohn, and K. Nottingham. The UCI machine learning repository. https://archive.ics.uci.edu, 2024.

[30] C. Liu and N. Boumal. Simple algorithms for optimization on Riemannian manifolds with constraints. *Appl. Math. Optim.*, 82(3):949–981, 2020.

[31] D. G. Luenberger and Y. Ye. *Linear and Nonlinear Programming.* Springer New York, NY, 3rd edition, 2015.

[32] C. López and J. Beasley. A heuristic for the circle packing problem with a variety of containers. *Eur. J. Oper. Res.*, 214(3):512–525, 2011.

[33] L. Minchenko and S. Stakhovski. On relaxed constant rank regularity condition in mathematical programming. *Optimization*, 60(4):429–440, 2011.

[34] L. Qi and Z. Wei. On the constant positive linear dependence condition and its application to SQP methods. *SIAM J. Optim.*, 10(4):963–981, 2000.

[35] Y. Yamakawa and H. Sato. Sequential optimality conditions for nonlinear optimization on Riemannian manifolds and a globally convergent augmented Lagrangian method. *Comput. Optim. Appl.*, 81(2):397–421, 2022.

[36] W. H. Yang, L.-H. Zhang, and R. Song. Optimality conditions for the nonlinear programming problems on Riemannian manifolds. *Pac. J. Optim.*, 10(2):415–434, 2014.