CONSTRAINT QUALIFICATIONS AND STRONG GLOBAL CONVERGENCE PROPERTIES OF AN AUGMENTED LAGRANGIAN METHOD ON RIEMANNIAN MANIFOLDS*

ROBERTO ANDREANI†, KELVIN R. COUTO‡, ORIZON P. FERREIRA§, AND GABRIEL HAESER¶

Abstract. In the past several years, augmented Lagrangian methods have been successfully applied to several classes of nonconvex optimization problems, inspiring new developments in both theory and practice. In this paper we bring most of these recent developments from nonlinear programming to the context of optimization on Riemannian manifolds, including equality and inequality constraints. Many research have been conducted on optimization problems on manifolds, however only recently the treatment of the constrained case has been considered. In this paper we propose to bridge this gap with respect to the most recent developments in nonlinear programming. In particular, we formulate several well-known constraint qualifications from the Euclidean context which are sufficient for guaranteeing global convergence of augmented Lagrangian methods, without requiring boundedness of the set of Lagrange multipliers. Convergence of the dual sequence can also be assured under a weak constraint qualification. The theory presented is based on so-called sequential optimality conditions, which is a powerful tool used in this context. The paper can also be read with the Euclidean context in mind, serving as a review of the most relevant constraint qualifications and global convergence theory of state-of-the-art augmented Lagrangian methods for nonlinear programming.

Key words. constraint qualifications, global convergence, augmented Lagrangian methods, Riemannian manifolds

MSC codes. 49J52, 49M15, 65H10, 90C30

DOI. 10.1137/23M1582382

1. Introduction. The problem of minimizing an objective function defined on a Riemannian manifold has received a lot of attention over the last twenty-five years. Several unconstrained algorithms on Euclidean spaces have been successfully adapted to this more general setting. These adaptations come from the fact that the Riemannian machinery, from a theoretical and practical point of view, allows treating several constrained optimization problems as unconstrained Riemannian problems. It is worth noting that the works on this subject involve more than merely a theoretical experiment in generalizing Euclidean space concepts to Riemannian manifolds, which is challenging in many different aspects. Unlike Euclidean spaces, Riemannian manifolds are nonlinear objects, making it challenging to develop a solid optimization theory in this setting. The most important thing to keep in mind is that these studies

---

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

are important mainly because many problems are most effectively addressed from a point of view of Riemannian geometry. In fact, many optimization problems have an underlying Riemannian geometric structure that can be efficiently exploited with the goal of designing more effective methods to solve them; some references on this subject include [1, 24, 30].

Although unconstrained Riemannian optimization is already somewhat well established, only a few works have appeared dealing with constrained Riemannian optimization (CRO) problems, that is, Riemannian optimization problems where equality and inequality constraints restrict the variables to a subset of the manifold itself. For instance, [60] extended to the Riemannian context the Karush/Kuhn–Tucker (KKT) conditions and second-order optimality conditions under a strong assumption, while in [19] a very interesting intrinsic approach was presented for defining suitable KKT conditions. In [59] the approximate-KKT (AKKT) sequential optimality condition was proposed to support the global convergence theory of an augmented Lagrangian method recently introduced in [45]. In [39] an exact penalty method for special problems on Stiefel manifolds was presented, and some constraint qualifications and the first- and second-order optimality conditions to support the method are discussed. A manifold inexact augmented Lagrangian framework to solve a family of nonsmooth optimization problems on Riemannian submanifolds embedded in Euclidean space is proposed in [28]. In [51], a Riemannian sequential quadratic optimization algorithm is proposed, which uses a line-search technique with an ℓ_1-penalty function as an extension of the standard sequential quadratic optimization algorithm for constrained nonlinear optimization problems in Euclidean spaces. In [42], a Riemannian interior point algorithm is introduced.

It is worth mentioning that the theoretical tools needed to support constrained optimization methods on the Riemannian setting are still under development. In fact, only recently in [19] has a full theory of constraint qualifications and optimality conditions been developed, where a definition of weak constraint qualifications (CQs) such as Guignard's CQ and Abadie's CQ were given. Despite guaranteeing the existence of Lagrange multipliers, more robust applications of these conditions are not known so far, even in the Euclidean setting. This is not the case of stronger conditions such as the linear independence CQ (LICQ) and Mangasarian–Fromovitz CQ (MFCQ), which gives, respectively, uniqueness and compactness of the Lagrange multiplier set, together with boundedness of a typical sequence of approximate Lagrange multipliers generated by several primal-dual algorithms, guaranteeing global convergence to a stationary point. These results were discussed in the Riemannian setting in [59].

In this paper our goal is to introduce several intrinsic weaker CQs in the Riemannian context, such as the constant rank CQ (CRCQ) [37], the constant positive linear dependence (CPLD) [52], and their relaxed variants (RCRCQ [48] and RCPLD [12]). RCPLD is the weakest of these four conditions introduced; however, all of them have their own set of applications, which we mention later. With the exception of CRCQ and RCRCQ, which are independent of MFCQ, all CQs presented are strictly weaker than MFCQ. Thus, despite the fact that no such condition guarantees boundedness of the set of Lagrange multipliers at a solution, they are still able to guarantee global convergence of primal-dual algorithms to a stationary point. In particular, we show that all limit points of a safeguarded augmented Lagrangian algorithm will satisfy the KKT conditions under all proposed conditions. Finally, we present two other conditions, the constant rank of the subspace component CQ (CRSC) [13] and the quasinormality CQ (QN) [35], which we also show to be enough for proving the global convergence result we mentioned. Although we do not pursue these results in the Riemannian setting, CRSC is expected to be strictly weaker than RCPLD, while these

conditions (CRSC and QN) are the weakest ones known in the Euclidean setting such that an Error Bound condition is satisfied. That is, locally, the distance to the feasible set can be measured by means of the norm of the constraint violation [13, 49], which should be an interesting result to be extended to the Riemannian setting. Throughout the text we review several results known in the Euclidean setting in order to serve as a guide for future extensions to Riemannian manifolds. We chose to present in the Riemannian setting an interesting characteristic of QN which is related to the global convergence of the augmented Lagrangian method; namely, under QN, the sequence of approximate Lagrange multipliers generated by the algorithm is bounded, guaranteeing primal-dual convergence even when the set of Lagrange multipliers is itself unbounded, a result that first appeared in [26] in the Euclidean setting. In order to do this, we will need to define a stronger sequential optimality condition known as positive-AKKT (PAKKT [3]). Finally, the machinery of sequential optimality conditions we introduce is relevant due to the fact that it is easy to extend the global convergence results we present to other algorithms. We want also to draw attention to the fact that all of the findings obtained in this study are also valid in Euclidean spaces; thus, this study may also be seen as a review of the recent developments in CQs and their connections with global convergence of algorithms in the Euclidean setting. This paper is organized as follows. Section 1.2 presents some definitions and preliminary results that are important throughout our study. In section 2, we state the CRO problem and also recall the KKT and AKKT conditions, together with the definitions of LICQ and MFCQ for CRO problems. Section 3 is devoted to introducing the new CQs for the CRO problem, namely, (R)CRCQ, (R)CPLD, and CRSC, where we present several examples and the proof that these conditions are indeed CQs associated with the global convergence of the augmented Lagrangian method. In section 4 we introduce the PAKKT condition and the QN, where we show that the Lagrange multiplier sequences generated by the augmented Lagrangian method are bounded under quasinormality. Section 5 presents some interesting examples of problems modeled by our framework. The last section contains some concluding remarks.

1.1. Summary of contributions. We state below a concise summary of the main contributions of this paper:

(i) We introduce several intrinsic CQs in the Riemannian context, including CRCQ and CPLD, along with their relaxed counterparts, RCRCQ and RCPLD, the latter being the weakest of these conditions. Two CQs unrelated with RCPLD are also defined: CRSC and QN.

(ii) We prove that all feasible limit points of the safeguarded augmented Lagrangian algorithm satisfy the KKT conditions under all proposed intrinsic CQs.

(iii) We survey the most relevant properties known about these conditions in the Euclidean setting, and we prove that under QN the sequence of approximate Lagrange multipliers generated by the safeguarded augmented Lagrangian algorithm is bounded, guaranteeing primal-dual convergence even when the set of Lagrange multipliers is unbounded.

(iv) We present several examples in the Riemannian setting to illustrate the different CQs introduced, and we also present some examples of interesting problems naturally modeled as CRO problems.

1.2. Notations, terminology, and basics results. In this section, we recall some notations and basic concepts of Riemannian manifolds used throughout the paper. They can be found in many books on Riemannian Geometry; see, for example, [43, 53, 54].

Let $\mathcal{M}$ be an n-dimensional smooth Riemannian manifold. Denote the *tangent space* at a point p by $T_p\mathcal{M}$, the *tangent bundle* by $T\mathcal{M} := \bigcup_{p \in M} T_p\mathcal{M}$, and a *vector field* by a mapping $X : \mathcal{M} \to T\mathcal{M}$ such that $X(p) \in T_p\mathcal{M}$. Assume also that $\mathcal{M}$ has a *Riemannian metric* denoted by $\langle \cdot, \cdot \rangle$ and the corresponding *norm* denoted by $\|\cdot\|$. For $f : U \to \mathbb{R}$, a differentiable function with derivative $df(\cdot)$, where U is an open subset of the manifold $\mathcal{M}$, the Riemannian metric induces the mapping $f \mapsto \operatorname{grad} f$ which associates its *gradient vector field* via the rule $\langle \operatorname{grad} f(p), X(p) \rangle := df(p)X(p)$ for all $p \in U$. The *length* of a piecewise smooth curve $\gamma : [a, b] \to \mathcal{M}$ joining p to q in $\mathcal{M}$, i.e., $\gamma(a) = p$ and $\gamma(b) = q$, is denoted by $\ell(\gamma)$. The *Riemannian distance* between p and q is defined as $d(p, q) = \inf_{\gamma \in \Gamma_{p,q}} \ell(\gamma)$, where $\Gamma_{p,q}$ is the set of all piecewise smooth curves in $\mathcal{M}$ joining points p and q. This distance induces the original topology on $\mathcal{M}$, namely, $(\mathcal{M}, d)$ is a complete metric space, and the bounded and closed subsets are compact. The *open* and *closed balls* of radius $r > 0$, centered at p, are respectively defined by $B_r(p) := \{q \in \mathcal{M} : d(p, q) < r\}$ and $B_r[p] := \{q \in \mathcal{M} : d(p, q) \leq r\}$. Let γ be a curve joining the points p and q in $\mathcal{M}$, and let ∇ be the Levi-Civita connection associated to $(\mathcal{M}, \langle \cdot, \cdot \rangle)$. A vector field Y along a smooth curve γ in $\mathcal{M}$ is said to be *parallel* when $\nabla_{\gamma'} Y = 0$. If γ' itself is parallel, we say that γ is a *geodesic*. If γ is a curve joining points p and q in $\mathcal{M}$, then, for each $t \in [a, b]$, ∇ induces a linear isometry, relative to $\langle \, , \, \rangle$, $P_{\gamma(a)\gamma(t)} : T_{\gamma(a)}\mathcal{M} \to T_{\gamma(t)}\mathcal{M}$, the so-called *parallel transport* along γ from $\gamma(a)$ to $\gamma(t)$. A Riemannian manifold is *complete* if its geodesics $\gamma(t)$ are defined for any value of $t \in \mathbb{R}$. From now on, $\mathcal{M}$ *denotes an n-dimensional smooth and complete Riemannian manifold.* Owing to the completeness of the Riemannian manifold $\mathcal{M}$, the *exponential map* at p, $\exp_p : T_p\mathcal{M} \to \mathcal{M}$, can be given by $\exp_p v = \gamma(1)$, where γ is the geodesic defined by its position p and velocity v at p and $\gamma(t) = \exp_p(tv)$ for any value of t. For $p \in \mathcal{M}$, the *injectivity radius* of $\mathcal{M}$ at p is defined by $r_p := \sup\{r > 0 : \exp_p|_{B_r(0_p)}$ is a diffeomorphism$\}$, where $B_r(0_p) := \{v \in T_p\mathcal{M} : \|v\| < r\}$, and 0_p denotes the origin of $T_p\mathcal{M}$. Hence, for $0 < \delta < r_p$ and $\exp_p(B_\delta(0_p)) = B_\delta(p)$, the map $\exp_p^{-1} : B_\delta(p) \to B_\delta(0_p)$ is a diffeomorphism. Moreover, for all $p, q \in B_\delta(p)$, there exists a unique geodesic segment γ joining p to q, which is given by $\gamma_{pq}(t) = \exp_p(t \exp_p^{-1} q)$ for all $t \in [0, 1]$. Furthermore, $d(q, p) = \|\exp_p^{-1} q\|$ and the map $B_\delta(p) \ni q \mapsto \frac{1}{2}d(q, p)^2$ is C^∞, and its gradient is given by $\operatorname{grad} \frac{1}{2}d(q, p)^2 = -\exp_p^{-1} q$; see, for example, [53, Proposition 4.8, p. 108].

Next we state some elementary facts on (positive-)linear dependence/independence of gradient vector fields, whose proofs are straightforward. Let $h = (h_1, \ldots, h_s) : \mathcal{M} \to \mathbb{R}^s$ and $g = (g_1, \ldots, g_m) : \mathcal{M} \to \mathbb{R}^m$ be continuously differentiable functions on $\mathcal{M}$. Let us denote

$$(1.1) \qquad A(q, \mathcal{I}, \mathcal{J}) := \{\operatorname{grad} h_i(q) : i \in \mathcal{I}\} \cup \{\operatorname{grad} g_j(q) : j \in \mathcal{J}\}, \qquad q \in \mathcal{M},$$

where $\mathcal{I} \subset \{1, \ldots, s\}$, $\mathcal{J} \subset \{1, \ldots, m\}$, while $\{\operatorname{grad} h_i(q) : i \in \mathcal{I}\} \cup \{\operatorname{grad} g_j(q) : j \in \mathcal{J}\}$ is a multiset; that is, repetition of the same element is allowed.

DEFINITION 1.1. *Let $V = \{v_1, \ldots, v_s\}$ and $W = \{w_1, \ldots, w_m\}$ be two finite multisets on $T_p\mathcal{M}$. The pair (V, W) is said to be positive-linearly dependent if there exist $\alpha = (\alpha_1, \ldots, \alpha_s) \in \mathbb{R}^s$ and $\beta = (\beta_1, \ldots, \beta_m) \in \mathbb{R}_+^m$ 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 from the context, we refer to $V \cup W$ instead of (V, W).*

LEMMA 1.2. *Let $p \in \mathcal{M}$, and assume that $A(p, \mathcal{I}, \mathcal{J})$ is (positive-)linearly independent. Then, there exists $\epsilon > 0$ such that $A(q, \mathcal{I}, \mathcal{J})$ is also (positive-)linearly independent for all $q \in B_\epsilon(p)$.*

LEMMA 1.3. *The following two conditions are equivalent:*
  (i) *There exists $\epsilon > 0$ such that, for all $\mathcal{I} \subset \{1,\ldots,s\}$ and $\mathcal{J} \subset \{1,\ldots,m\}$, whenever the set $A(p,\mathcal{I},\mathcal{J})$ is linearly dependent, $A(q,\mathcal{I},\mathcal{J})$ is also linearly dependent for all $q \in B_\epsilon(p)$.*
  (ii) *There exists $\epsilon > 0$ such that, for all $\mathcal{I} \subset \{1,\ldots,s\}$ and $\mathcal{J} \subset \{1,\ldots,m\}$, the rank of $A(q,\mathcal{I},\mathcal{J})$ is constant for any $q \in B_\epsilon(p)$.*

LEMMA 1.4 (Carathéodory's Lemma [12]). *Let $u_1,\ldots,u_s,v_1,\ldots,v_m$ be vectors in a finite-dimensional vector space V such that $\{u_1,\ldots,u_s\}$ is linearly independent. Suppose $x \in V$ is such that there are real scalars $\alpha_1,\ldots,\alpha_s,\beta_1,\ldots,\beta_m$, with $\beta_j \neq 0$ for $j = 1,\ldots,m$ and $x = \sum_{i=1}^s \alpha_i u_i + \sum_{j=1}^m \beta_j v_j$. Then, there exist a subset $\mathcal{J} \subset \{1,\ldots,m\}$ and real scalars $\bar{\alpha}_i, i = 1,\ldots,s$ and $\bar{\beta}_j \neq 0, j \in \mathcal{J}$ such that $x = \sum_{i=1}^m \bar{\alpha}_i u_i + \sum_{j\in\mathcal{J}} \bar{\beta}_j v_j$, $\bar{\beta}_j \beta_j > 0$ for all $j \in \mathcal{J}$, and $\{u_i : i \in \{1,\ldots,s\}\} \cup \{v_j : j \in \mathcal{J}\}$ is linearly independent.*

We end this section by stating some standard notations in Euclidean spaces. The set of all $m \times n$ matrices with real entries is denoted by $\mathbb{R}^{m\times n}$ and $\mathbb{R}^m \equiv \mathbb{R}^{m\times 1}$. For $M \in \mathbb{R}^{m\times n}$, the matrix $M^\top \in \mathbb{R}^{n\times m}$ is the *transpose* of M. For all $x,y \in \mathbb{R}^m$, $\min\{x,y\} \in \mathbb{R}^m$ is the component-wise minimum of x and y. We denote by $[y]_+$ the Euclidean projection of y onto the nonnegative orthant $\mathbb{R}^m_+$, while $\|y\|_2$ and $\|y\|_\infty$ denote its Euclidean and infinity norms, respectively.

2. Preliminaries. In this paper we are interested in the following CRO problem:

$$(2.1) \qquad \begin{aligned} &\underset{q\in\mathcal{M}}{\text{Minimize }} f(q), \\ &\text{subject to } h(q) = 0, \ g(q) \leq 0, \end{aligned}$$

where $\mathcal{M}$ is an n-dimensional smooth and complete Riemannian manifold, the functions $f : \mathcal{M} \to \mathbb{R}$, $h = (h_1,\ldots,h_s) : \mathcal{M} \to \mathbb{R}^s$, and $g = (g_1,\ldots,g_m) : \mathcal{M} \to \mathbb{R}^m$ are continuously differentiable on $\mathcal{M}$. The feasible set $\Omega \subset \mathcal{M}$ of problem (2.1) is defined by

$$(2.2) \qquad \Omega := \{q \in \mathcal{M} : h(q) = 0, \ g(q) \leq 0\},$$

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

$$(2.3) \qquad \mathcal{A}(q) := \{j \in \{1,\ldots,m\} : g_j(q) = 0\}.$$

We say that the KKT conditions are satisfied at $p \in \Omega$ when there exists so-called Lagrange multipliers $(\lambda,\mu) \in \mathbb{R}^s \times \mathbb{R}^m_+$ such that the following two conditions hold:
  (i) $\operatorname{grad} L(p,\lambda,\mu) = 0$,
  (ii) $\mu_j = 0$ for all $j \notin \mathcal{A}(p)$,
where $L(\cdot,\lambda,\mu) : \mathcal{M} \to \mathbb{R}$ is the *Lagrangian function* defined by

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

For $p \in \Omega$, the *linearized/linearization cone* $\mathcal{L}(p)$ is defined as

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

and its polar is given by

$$(2.4)$$

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

A CQ is a condition that refers to the analytic description of the feasible set and that guarantees that every local minimizer is also a KKT point. In [19] it was shown that when p is a local minimizer of (2.1) that satisfies Guignard's CQ, that is, $\mathcal{L}(p)^\circ = \mathcal{T}(p)^\circ$, where $\mathcal{T}(p)$ is the Bouligand tangent cone of Ω at p, then the KKT conditions are satisfied at p.

In [55] the convex inequality constrained problem is studied, under a Slater CQ, on a complete Riemannian manifold. In this case, the objective and inequality constraints are convex along geodesics and the feasible set is described by a finite collection of inequality constraints. In this context KKT conditions are formulated. In [60] it was shown that when p is a local solution of (2.1) and LICQ holds at p, that is, the set $\{\operatorname{grad} h_i(p):\ i = 1,\ldots,s\} \cup \{\operatorname{grad} g_j(p):\ j \in \mathcal{A}(p)\}$ is linearly independent, then the KKT conditions are satisfied at p.

Without CQs, an approximate verison of the KKT conditions are known to be satisfied at local minimizers.

THEOREM 2.1. *Let* $p \in \Omega$ *be a local minimizer of* (2.1). *Then* p *is an AKKT point; that is, there exist sequences* $(p^k)_{k\in\mathbb{N}} \subset \mathcal{M}$, $(\lambda^k)_{k\in\mathbb{N}} \subset \mathbb{R}^s$ *and* $(\mu^k)_{k\in\mathbb{N}} \subset \mathbb{R}_+^m$ *such that*

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

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

(iii) $\mu_j^k = 0$ *for all* $j \notin \mathcal{A}(p)$ *and sufficiently large* k.

This result appeared in [59], as an extension of the well-known nonlinear programming version of this theorem [5]. Any sequence $(p^k)_{k\in\mathbb{N}} \subset \mathcal{M}$ that satisfies (i), (ii), and (iii) above is called a primal AKKT sequence for p, while the correspondent sequence $(\lambda^k, \mu^k)_{k\in\mathbb{N}}$ is its dual sequence. In the Euclidean setting, this notion has shown to be crucial in developing new CQs and expanding global convergence results of several algorithms in different contexts; for instance, nonlinear programming [12, 16], Nash equilibrium problems [26], quasiequilibrium problems [25], multiobjective problems [32], second-order cone programming [7], semidefinite programming [15, 6, 9], Banach spaces [23], equilibrium constraints [14], and cardinality constraints [40], among several other applications and extensions. In [59] the following safeguarded augmented Lagrangian algorithm was defined, and it was proved that its iterates are precisely AKKT sequences. In order to define it, we denote by $\mathcal{L}_\rho(\cdot, \lambda, \mu) : \mathcal{M} \to \mathbb{R}$ the standard Powell–Hestenes–Rockafellar augmented Lagrangian function, defined by

$$\mathcal{L}_\rho(q, \lambda, \mu) := f(q) + \frac{\rho}{2}\left(\left\|h(q) + \frac{\lambda}{\rho}\right\|^2 + \left\|\left[g(q) + \frac{\mu}{\rho}\right]_+\right\|^2\right).$$

In the algorithm, $(\lambda^k, \mu^k)_{k\in\mathbb{N}}$ is the dual sequence associated with $(p^k)_{k\in\mathbb{N}}$, which may be unbounded, while the safeguarded dual sequence $(\bar{\lambda}^k, \bar{\mu}^k)_{k\in\mathbb{N}}$ is bounded and used for defining the subproblems. A standard choice is considering $(\bar{\lambda}^{k+1}, \bar{\mu}^{k+1})$ as the projection of (λ^k, μ^k) onto the corresponding box. It was shown in [59] that any limit point of a sequence $(p^k)_{k\in\mathbb{N}}$ generated by Algorithm 2.1 is such that it is stationary for the problem of minimizing an infeasibility measure, namely,

$$\underset{q\in\mathcal{M}}{\text{Minimize}}\ \frac{1}{2}\|h(q)\|_2^2 + \frac{1}{2}\|g(q)_+\|_2^2.$$

When the limit point is feasible, they showed that it is an AKKT point. The correspondent AKKT sequence is precisely the sequence $(p^k)_{k\in\mathbb{N}}$ of primal iterates, what can be attested by the dual sequences $(\lambda^k, \mu^k)_{k\in\mathbb{N}}$ generated by the algorithm.

Algorithm 2.1 Safeguarded augmented Lagrangian algorithm.

Step 0. Take $p^0 \in \mathcal{M}$, $\tau \in [0, 1)$, $\gamma > 1$, $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max} > 0$, and $\rho_1 > 0$. Take also $\bar{\lambda}^1 \in [\lambda_{\min}, \lambda_{\max}]^s$ 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 such that

$$(2.5) \qquad \left\| \operatorname{grad} \mathcal{L}_{\rho_k}(p^k, \bar{\lambda}^k, \bar{\mu}^k) \right\| \leq \epsilon_k.$$

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

Step 2. (Estimate new multipliers) Compute

$$(2.6) \qquad \lambda^k = \bar{\lambda}^k + \rho_k h(p^k), \qquad \mu^k = \left[\bar{\mu}^k + \rho_k g(p^k) \right]_+ .$$

Step 3. (Update the penalty parameter) Define

$$(2.7) \qquad v^k = \frac{\mu^k - \bar{\mu}^k}{\rho_k}.$$

If $k = 1$ or

$$(2.8) \qquad \max \left\{ \left\| h(p^k) \right\|_2 , \left\| v^k \right\|_2 \right\} \leq \tau \max \left\{ \left\| h(p^{k-1}) \right\|_2 , \left\| v^{k-1} \right\|_2 \right\},$$

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

Step 4. (Update multipliers estimates) 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.

Thus, in order to establish a standard global convergence result to Algorithm 2.1, namely, by showing that its feasible limit points satisfy the KKT conditions, it is sufficient to consider any condition that guarantees that all AKKT points are in fact KKT points. Due to Theorem 2.1, the said condition will necessarily be a CQ. CQs with this additional propery are sometimes called *strict* CQs, and only the following ones have been stated in the Riemannian context.

DEFINITION 2.2. *Let Ω be given by (2.2), let $p \in \Omega$, and let $\mathcal{A}(p)$ be given by (2.3). The point p is said to satisfy*
   (i) *the LICQ if*

$$\{ \operatorname{grad} h_i(p) \colon i = 1, \ldots, s \} \cup \{ \operatorname{grad} g_j(p) \colon j \in \mathcal{A}(p) \}$$

   is linearly independent;
   (ii) *the MFCQ if*

$$\{ \operatorname{grad} h_i(p) \colon i = 1, \ldots, s \} \cup \{ \operatorname{grad} g_j(p) \colon j \in \mathcal{A}(p) \}$$

   is positive-linearly independent.

The definition of LICQ was presented in [60] while MFCQ was introduced in [19], where it was shown that LICQ implies MFCQ. In the next section we will introduce several new weaker CQs, and we will prove that they can still be used for proving global convergence to a KKT point of algorithms that generate AKKT sequences such as Algorithm 2.1.

3. New strict CQs. We will say that a property $\mathcal{P}$ of the constraints defining the feasible set Ω of (2.1) at a given point $p \in \Omega$ is a *strict* CQ for the necessary optimality condition $\mathcal{S}$ if, at a point $p \in \Omega$ that satisfies both $\mathcal{P}$ and $\mathcal{S}$, it is necessarily the case that p satisfies the KKT conditions, according to the definition given in [22]. Thus, after we present and discuss our conditions, we shall prove that they are all strict CQs with respect to the sequential optimality condition AKKT from Theorem 2.1. The first conditions we propose are the following.

DEFINITION 3.1. *Let Ω be given by* (2.2), *let $p \in \Omega$, and let $\mathcal{A}(p)$ be given by* (2.3). *The point p is said to satisfy*
   (i) *the CRCQ if there exists $\epsilon > 0$ such that, for all $\mathcal{I} \subset \{1, \ldots, s\}$ and $\mathcal{J} \subset \mathcal{A}(p)$, the rank of $\{\operatorname{grad} h_i(q) : i \in \mathcal{I}\} \cup \{\operatorname{grad} g_j(q) : j \in \mathcal{J}\}$ is constant for all $q \in B_\epsilon(p)$;*
   (ii) *the CPLD if, for any $\mathcal{I} \subset \{1, \ldots, s\}$ and $\mathcal{J} \subset \mathcal{A}(p)$, whenever the set $\{\operatorname{grad} h_i(p) : i \in \mathcal{I}\} \cup \{\operatorname{grad} g_j(p) : j \in \mathcal{J}\}$ is positive-linearly dependent, there exists $\epsilon > 0$ such that $\{\operatorname{grad} h_i(q) : i \in \mathcal{I}\} \cup \{\operatorname{grad} g_j(q) : j \in \mathcal{J}\}$ is linearly dependent for all $q \in B_\epsilon(p)$;*
   (iii) *the RCRCQ if there exists $\epsilon > 0$ such that, for all $\mathcal{J} \subset \mathcal{A}(p)$, the rank of*

$$\{\operatorname{grad} h_i(q) : i = 1, \ldots, s\} \cup \{\operatorname{grad} g_j(q) : j \in \mathcal{J}\}$$

   is constant for all $q \in B_\epsilon(p)$;
   (iv) *the RCPLD if there exists $\epsilon > 0$ such that the following two conditions hold:*
      (a) *the rank of $\{\operatorname{grad} h_i(q) : i = 1, \ldots, s\}$ is constant for all $q \in B_\epsilon(p)$;*
      (b) *let $\mathcal{K} \subset \{1, \ldots, s\}$ be such that $\{\operatorname{grad} h_i(p) : i \in \mathcal{K}\}$ is a basis for the subspace generated by $\{\operatorname{grad} h_i(p) : i = 1, \ldots, s\}$. For all $\mathcal{J} \subset \mathcal{A}(p)$, if $\{\operatorname{grad} h_i(p) : i \in \mathcal{K}\} \cup \{\operatorname{grad} g_j(p) : j \in \mathcal{J}\}$ is positive-linearly dependent, then $\{\operatorname{grad} h_i(q) : i \in \mathcal{K}\} \cup \{\operatorname{grad} g_j(q) : j \in \mathcal{J}\}$ is linearly dependent for all $q \in B_\epsilon(p)$.*

These conditions are natural versions in the Riemannian setting of the existing conditions in the Euclidean setting. The following diagram shows the relationship among the conditions introduced so far, where an arrow represents strict implication, which shall be proved later in this section.

The reason for introducing these CQs in the Riemannian setting is due to their several applications known in the Euclidean case, which we expect to be extended also to the Riemannian setting. Although we shall only prove results concerning the global convergence of the safeguardaded augmented Lagrangian method, let us briefly review some properties of these conditions in the Euclidean setting.

LICQ is equivalent to the uniqueness of the Lagrange multiplier for any objective function that assumes a constrained minimum at the point [56]. However, it is considered to be too stringent. For instance, it fails when the same constraint is repeated twice in the problem formulation. On the other hand, MFCQ and its many equivalent statements is the most prevalent CQ in the nonlinear programming literature, with several applications. In particular, it considers the correct sign of the Lagrange multiplier in its formulation, what can be thought as it being a more adequate statement than LICQ. However, the simple trick of replacing an inequality constraint $h(x) = 0$ by two inequalities $h(x) \leq 0$ and $-h(x) \leq 0$ is enough for ensuring that MFCQ does not hold. This is due to the fact that, under this very natural formulation, the set of Lagrange multipliers (if nonempty) is necessarily unbounded, while MFCQ is equivalent to the boundedness of this set [56]. Notice also that the case of linear constraints is not automatically covered by any of these two assumptions, which generally requires a separate analysis when one is assuming MFCQ or LICQ.

Condition CRCQ, on the other hand, gives more freedom to someone modeling an optimization problem, given that it is not tricked by repetition of a constraint or by splitting an equality constraint in two inequalities. It also subsumes the linear case, disregarding a separate analysis. However, it does not consider the correct sign of Lagrange multipliers, being thus independent of MFCQ. The CPLD condition, on the other hand, corrects this issue, introducing the correct sign considering positive linear dependence instead of standard linear dependence (see the alternative definition of CRCQ as given by Lemma 1.3), being then strictly weaker than both MFCQ and CRCQ together. This condition has been used mainly for showing global convergence of algorithms, firstly for an SQP method, when it was introduced in [52], and it was popularized for being the basis for the global convergence theory of the popular ALGENCAN software [22]. However, other applications have emerged such as in bilevel optimization [58, 47], switching constraints [44], and exact penalty [33], among several others. On the other hand, CRCQ is more robust in terms of applications, since it has been introduced in order to compute the derivative of the value function [37]. It has also found applications in the characterization of tilt stable minimizers [31]. More interestingly, while MFCQ is still able to provide a second-order necessary optimality condition, the condition depends on the whole set of Lagrange multipliers, which limits its practical use. CRCQ, on the other hand, provides a strong second-order necessary optimality condition depending on a single Lagrange multiplier [12], a result which was recently extended to the context of conic optimization [10]. This difference with respect to MFCQ in terms of the second-order results can somehow be explained by the fact that under CRCQ the value of the quadratic form defined by the Hessian of the Lagrangian evaluated at a direction in the critical cone is invariant to the Lagrange multiplier used to define it [31].

The relaxed versions of CRCQ [48] and CPLD [12] can in fact be thought as the "correct" versions of these conditions, as they enjoy the very same properties previously described. In fact, there is no reason to consider all subsets of equality constraints, and this was already present in the very first results proved under CRCQ by Janin in [37].

Let us now prove that the strict implications shown in Figure 1 hold for any Riemannian manifold $\mathcal{M}$ with dimension $n \geq 2$. We do this by providing several examples that help illustrate the different conditions we propose; however, we only describe in details the computations concerning the most relevant examples, the other ones being analogous. We start by showing in the next two examples that MFCQ and CRCQ are independent conditions; that is, in Example 1, CRCQ holds and MFCQ fails, while in Example 2, MFCQ holds while CRCQ fails.

Example 1. Consider problem (2.1) with feasible set $\Omega := \{q \in \mathcal{M} : h(q) \leq 0, -h(q) \leq 0\}$, where $h : \mathcal{M} \to \mathbb{R}$ is continuously differentiable on $\mathcal{M}$ with grad $h(p) \neq 0$ and $p \in \Omega$. Thus MFCQ fails at p while CRCQ holds.
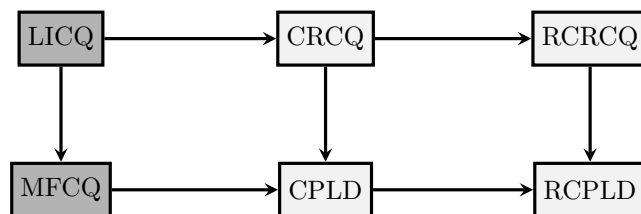

Fig. 1. *Strict CQs for problem* (2.1).

To proceed with the examples, let us define some auxiliary functions. Let $\mathcal{M}$ be an n-dimensional Riemannian manifold and $p \in \mathcal{M}$. Take $0 < \bar{\delta} < r_p$, where r_p is the injectivity radius, such that $B_{\bar{\delta}}(p)$ is a strongly convex neighborhood, which exists by [29, Proposition 4.2]. Let $u, v \in T_p\mathcal{M}$ with $\|u\| = \|v\| = 1$, $\langle v, u \rangle = 0$, and define the geodesics $\gamma_u(t) := \exp_p(tu)$ and $\gamma_v(t) := \exp_p(tv)$. Take also $0 < \delta < \bar{\delta}$, and define $p_1 := \gamma_u(-\delta)$, $p_2 := \gamma_u(\delta)$, and $p_3 = \gamma_v(\delta)$. Note that $p_1, p_2, p_3 \in B_{\bar{\delta}}(p)$ with $p_1 \neq p_2$, $p_1 \neq p_3$, and $p_2 \neq p_3$. Define the following auxiliary functions:

$$(3.1) \qquad \varphi_i(q) = \frac{1}{2}d(q, p_i)^2 - \frac{1}{2}d(p, p_i)^2, \qquad i = 1, 2, 3.$$

Example 2. Define the functions $g_1(q) := \varphi_1(q)$, $g_2(q) := -\varphi_2(q)$, and consider a feasible set $\Omega := \{q \in \mathcal{M} : g(q) \leq 0\}$, where $g := (g_1, g_2)$. One can see that CRCQ is not valid at $p \in \Omega$ while MFCQ holds.

In [52], it was proved that CPLD is strictly weaker than both MFCQ and CRCQ together in the Euclidean setting. The next example shows that the same thing happens for any smooth Riemannian manifold $\mathcal{M}$ with dimension $n \geq 2$.

Example 3. Let $\mathcal{M}$ be an n-dimensional Riemannian manifold with $n \geq 2$. Take $p \in \mathcal{M}$ and $g := (g_1, \ldots, g_4) : \mathcal{M} \to \mathbb{R}^4$ continuously differentiable functions satisfying the following conditions:
   (i) $g(p) = 0$;
   (ii) $\operatorname{grad} g_1(p) = \operatorname{grad} g_2(p) \neq 0$ and $\operatorname{grad} g_3(p) = -\operatorname{grad} g_4(p) \neq 0$;
   (iii) for all $\epsilon > 0$, there exists $q \in B_\epsilon(p)$ such that $\{\operatorname{grad} g_1(q), \operatorname{grad} g_2(q)\}$ is linearly independent with $q \neq p$;
   (iv) $\{\operatorname{grad} g_1(p), \operatorname{grad} g_3(p)\}$ is linearly independent.
   (v) there exists $\epsilon > 0$ such that $\{\operatorname{grad} g_3(q), \operatorname{grad} g_4(q)\}$ is linearly dependent for all $q \in B_\epsilon(p)$.
Consider the feasible set $\Omega := \{q \in \mathcal{M} : g(q) \leq 0\}$. Then, by (i), $p \in \Omega$. It follows from the first equality in condition (ii) that the set $\{\operatorname{grad} g_1(p), \operatorname{grad} g_2(p)\}$ is linearly dependent. Thus, (iii) implies that p does not satisfy CRCQ. Furthermore, the second equality in condition (ii) guarantees that p does not satisfy MFCQ. We will now show that p satisfies CPLD. It is easy to see that the set $\{\operatorname{grad} g_j(p) : j \in \mathcal{J} \subset \mathcal{A}(p)\}$ is positive-linearly dependent if and only if $\{3, 4\} \subset \mathcal{J}$. Therefore, by (v) we concluded that p satisfies CPLD. In the following we build two concrete examples satisfying conditions (i), (ii), (iii), (iv), and (v). The first one considers $\mathcal{M}$ as the 2-sphere while the second one considers an arbitrary 2-dimensional manifold $\mathcal{M}$; however, it is easy to generalize the examples to an arbitrary dimension $n \geq 2$.
   • Consider the sphere $\mathcal{M} := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$ and take $p := (0, 0, 1)$. The functions $g_1(x, y, z) := x$, $g_2(x, y, z) := x + y^2$, $g_3(x, y, z) := x + y$, $g_4(x, y, z) := -x - y$ satisfy conditions (i), (ii), (iii), (iv), and (v). Indeed, for $q = (x, y, z) \in B_\epsilon(p)$ we have $\operatorname{grad} g_1(q) = \Pi_{T_q\mathcal{M}}(1, 0, 0)$, $\operatorname{grad} g_2(q) = \Pi_{T_q\mathcal{M}}(1, 2y, 0)$, $\operatorname{grad} g_3(q) = \Pi_{T_q\mathcal{M}}(1, 1, 0)$, and $\operatorname{grad} g_4(q) = \Pi_{T_q\mathcal{M}}(-1, -1, 0)$, where $\Pi_{T_q\mathcal{M}}$ denotes the orthogonal projection onto $T_q\mathcal{M}$. Clearly, (i) holds. To see that (ii) and (iv) hold it is enough to note that since $T_p\mathcal{M} = \{p\}^\perp$, at $q = p$ the projections coincide with the vectors being projected. We proceed to prove that (iii) holds. Let $q := (x, y, z)$ with $y \neq 0$ and $z \neq 0$; hence $u := (1, 0, 0)$, $v := (1, 2y, 0)$, and q are linearly independent. Take $\alpha, \beta \in \mathbb{R}$ such that $\alpha \operatorname{grad} g_1(q) + \beta \operatorname{grad} g_2(q) = 0$. Since $T_q\mathcal{M} = \{q\}^\perp$, this implies that $\alpha(u - r_u q) + \beta(v - r_v q) = 0$ for some $r_u, r_v \in \mathbb{R}$, which in turn gives $\alpha u + \beta v + (-\alpha r_u - \beta r_v)q = 0$, implying $\alpha = \beta = 0$; hence, (iii). We obtain (v) by noting that $\operatorname{grad} g_3(q) = -\operatorname{grad} g_4(q)$ for all q.

- Let $\mathcal{M}$ be a 2-dimensional complete manifold. Let us show that the functions $g_1(q) := \varphi_1(q)$, $g_2(q) := -\varphi_2(q)$, $g_3(q) := \varphi_3(q)$, and $g_4(q) := -\varphi_3(q)$ satisfy conditions (i), (ii), (iii), (iv), and (v), where these functions are defined in (3.1). Indeed, $g(p) = 0$, which gives (i). Since $\operatorname{grad} g_1(p) = -\exp_p^{-1} p_1 = \delta u \neq 0$, $\operatorname{grad} g_2(p) = \exp_p^{-1} p_2 = \delta u$, $\operatorname{grad} g_3(p) = -\exp_p^{-1} p_3 = -\delta v \neq 0$, and $\operatorname{grad} g_4(p) = \exp_p^{-1} p_3 = \delta v$, then g satisfies (ii). We proceed to show that g_1 and g_2 satisfy (iii). For that, take a point $q \in B_{\bar{\delta}}(p)$ with $q \neq p$ such that $d(q, p_2) = d(q, p_1) < d(p_2, p_1)$. Note that $\operatorname{grad} g_1(q) = -\exp_q^{-1} p_1$ and $\operatorname{grad} g_2(q) = \exp_q^{-1} p_2$. In addition, due to $d(q, p_2) = d(q, p_1)$, we have $\|\operatorname{grad} g_1(q)\| = \|\operatorname{grad} g_2(q)\| = d(q, p_1)$. Assume by contradiction that $\{\operatorname{grad} g_1(q), \operatorname{grad} g_2(q)\}$ is linearly dependent. Since $\mathcal{M}$ is 2-dimensional and $d(q, p_1) < d(p_2, p_1)$, we conclude that $\operatorname{grad} g_1(q) = \operatorname{grad} g_2(q)$. Consider the geodesic

$$\gamma(t) = \exp_q(-t \exp_q^{-1} p_1) = \exp_q(t \exp_q^{-1} p_2),$$

where the second equality holds because we are under the assumption $\operatorname{grad} g_1(q) = \operatorname{grad} g_2(q)$. Hence $\gamma(0) = q$, $\gamma(-1) = p_1$, and $\gamma(1) = p_2$. Considering that there exists a unique geodesic joining p_1 and p_2, we conclude that $\gamma_u = \gamma$. Thus, there exists $\bar{t}$ such that $\gamma(\bar{t}) = p$ and

$$\gamma'(\bar{t}) = P_{q\gamma(\bar{t})}(-\exp_q^{-1} p_1),$$

where $P_{q\gamma(\bar{t})}$ is the parallel transport along γ from $\gamma(0) = q$ to $\gamma(\bar{t})$. We know that the parallel transport is an isometry and $\gamma'(\bar{t}) = \delta \gamma'_u(0) = -\exp_p^{-1} p_1$. Thus, using the last equality we conclude that $d(q, p_1) = d(p, p_1)$. Hence, considering that q, p, and p_1 belong to the same geodesic, we have $q = p$, which is a contradiction. Therefore, $\{\operatorname{grad} g_1(q), \operatorname{grad} g_2(q)\}$ is linearly independent for $q \neq p$. Due to $\operatorname{grad} g_1(p) = -\delta u$, $\operatorname{grad} g_3(p) = \delta v$ and $\langle v, u \rangle = 0$, condition (iv) is satisfied. Finally, due to $g_4 = -g_3$, we have $\operatorname{grad} g_4(q) = -\operatorname{grad} g_3(q)$, and (v) is also satisfied. The situation in consideration is depicted in Figure 2.

As proved in [48], RCRCQ is strictly weaker than CRCQ in the Euclidean context. This fact is also true for any smooth Riemannian manifold $\mathcal{M}$ with dimension $n \geq 2$. In fact, the following example shows that RCRCQ does not imply CPLD.

Example 4. Define the functions $h(q) := -\varphi_3(q)$, $g_1(q) := -\varphi_1(q)$, $g_2(q) := -\varphi_2(q)$ from (3.1), and consider a feasible set $\Omega := \{q \in \mathcal{M} : h(q) = 0, g(q) \leq 0\}$, where $g := (g_1, g_2)$. The point $p \in \Omega$ does not satisfy CPLD, but it satisfies RCRCQ.
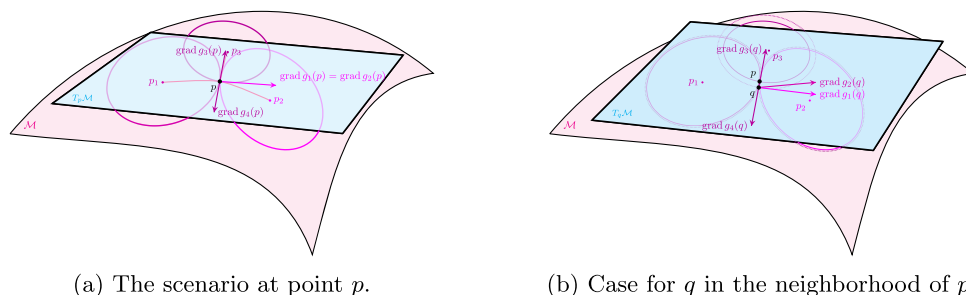

(a) The scenario at point p.   (b) Case for q in the neighborhood of p.

FIG. 2. *Illustrative figure for Example 3 where CPLD holds while MFCQ and CRCQ fail. The rank of the gradients indexed by $\{1, 2\}$ is not constant, but they are positive-linearly independent. In addition, the gradients in $\{3, 4\}$ are positive-linearly dependent, which implies that MFCQ fails, but this remains to be the case in a neighborhood.*

Finally, let us show that RCRCQ implies RCPLD. We do this by providing the following equivalent description of RCRCQ:

PROPOSITION 3.2. *Let* $\mathcal{K} \subset \{1, \ldots, s\}$ *be such that* $\{\operatorname{grad} h_i(p) \colon i \in \mathcal{K}\}$ *is a basis for the subspace generated by* $\{\operatorname{grad} h_i(p) \colon i = 1, \ldots, s\}$. *RCRCQ holds at* $p \in \Omega$ *if and only if there exists* $\epsilon > 0$ *such that the following two conditions hold:*
  (i) *the rank of* $\{\operatorname{grad} h_i(q) \colon i = 1, \ldots, s\}$ *is constant for all* $q \in B_\epsilon(p)$;
  (ii) *for all* $\mathcal{J} \subset \mathcal{A}(p)$, *if* $\{\operatorname{grad} h_i(p) \colon i \in \mathcal{K}\} \cup \{\operatorname{grad} g_j(p) \colon j \in \mathcal{J}\}$ *is linearly dependent, then* $\{\operatorname{grad} h_i(q) \colon i \in \mathcal{K}\} \cup \{\operatorname{grad} g_j(q) \colon j \in \mathcal{J}\}$ *is linearly dependent for all* $q \in B_\epsilon(p)$.

Proof. Assume first that $p \in \Omega$ satisfies RCRCQ. Taking $\mathcal{J} = \emptyset$ in the definition of RCRCQ, we obtain (i). In order to obtain (ii), let $\mathcal{J} \subset \mathcal{A}(p)$ such that $\{\operatorname{grad} h_i(p) \colon i \in \mathcal{K}\} \cup \{\operatorname{grad} g_j(p) \colon j \in \mathcal{J}\}$ is linearly dependent. Since $\{\operatorname{grad} h_i(p) \colon i \in \mathcal{K}\}$ is a basis for the subspace generated by $\{\operatorname{grad} h_i(p) \colon i = 1, \ldots, s\}$ and (i), we have that there exists $\epsilon > 0$ such that $\{\operatorname{grad} h_i(q) \colon i \in \mathcal{K}\}$ is a basis for the subspace generated by $\{\operatorname{grad} h_i(q) \colon i = 1, \ldots, s\}$ for all $q \in B_\epsilon(p)$. Thus, in accordance with RCRCQ, the rank of $\{\operatorname{grad} h_i(q) \colon i \in \mathcal{K}\} \cup \{\operatorname{grad} g_j(q) \colon j \in \mathcal{J}\}$ is constant for all $q \in B_\epsilon(p)$. Consequently, (ii) holds.

To prove the reciprocal assertion, let $\mathcal{J} \subset \mathcal{A}(p)$. It is worth noting that, owing to Lemma 1.2, the rank of the set cannot decrease in a neighborhood. Let us choose $\mathcal{J}_1 \subset \mathcal{J}$ such that $A(p, \mathcal{K}, \mathcal{J}_1)$ is a basis for $A(p, \mathcal{K}, \mathcal{J})$—see the notation introduced in (1.1). The case $\mathcal{J}_1 = \mathcal{J}$ follows trivially. Consider the situation $\mathcal{J}_1 \neq \mathcal{J}$, and let $j \in \mathcal{J}$ with $j \notin \mathcal{J}_1$. As a result of (ii), $A(q, \mathcal{K}, \mathcal{J}_1) \cup \{\operatorname{grad} g_j(q)\}$ must continue to be linearly dependent for q in a neighborhood of p. Therefore, rank $A(q, \mathcal{K}, \mathcal{J}) = |\mathcal{K}| + |\mathcal{J}_1|$ for all $q \in B_\epsilon(p)$ and sufficiently small $\epsilon > 0$. Considering the definition of $\mathcal{K}$ and (i), we must have that $A(q, \{1, \ldots, s\}, \mathcal{J})$ has constant rank for $q \in B_\epsilon(p)$, which completes the proof. $\square$

Clearly, the equivalent definition of RCRCQ given by Proposition 3.2 is independent of the choice of the index set $\mathcal{K}$. It is easy to see that the definition of RCPLD is also independent of this choice. This concludes the analysis of strict implications depicted in Figure 1, where in particular we have that RCPLD is strictly weaker than CPLD and RCRCQ.

At this point, condition RCPLD is the weakest one among the ones we have presented. Thus, we will prove that RCPLD is a strict CQ with respect to the AKKT condition, which will be true for all other conditions that imply it. Before doing this, let us present yet another CQ called the *constant rank of the subspace component* (CRSC) [13]. Noticing that while RCRCQ improves upon CRCQ by noticing that there is no reason to consider every subset of equality constraints, CRSC improves upon RCRCQ by noticing that the same thing is true with respect to the inequality constraints. Namely, it is not the case that every subset of the active inequality constraints must be taken into account; only a particular fixed subset of the constraints maintaining the constant rank property is enough for guaranteeing the existence of Lagrange multipliers. The definition is as follows.

DEFINITION 3.3. *Let* Ω *be given by* (2.2), *let* $p \in \Omega$, $\mathcal{A}(p)$, *and let* $\mathcal{L}(p)^\circ$ *be given by* (2.2) *and* (2.4), *respectively. Define the index set* $\mathcal{J}_-(p) = \{j \in \mathcal{A}(p) \colon -\operatorname{grad} g_j(p) \in \mathcal{L}(p)^\circ\}$. *The point* p *is said to satisfy CRSC if there exists* $\epsilon > 0$ *such that the rank of* $\{\operatorname{grad} h_i(q) \colon i = 1, \ldots, s\} \cup \{\operatorname{grad} g_j(q) \colon j \in \mathcal{J}_-(p)\}$ *is constant for all* $q \in B_\epsilon(p)$.

It is clear that CRSC is weaker than RCRCQ and MFCQ, but its relation with RCPLD is not simple to establish. Somewhat surprisingly, CRSC is strictly weaker

than RCPLD in the Euclidean setting. The proof is somewhat elaborate [13], so we did not pursue it in the Riemannian setting, as it was not needed in our developments. Instead, we simply show that CRSC does not imply RCPLD in any Riemannian manifold $\mathcal{M}$ of dimension $n \geq 2$. Thus, we shall prove the convergence of the augmented Lagrangian method under either of these conditions, even though we expect CRSC to be weaker than RCPLD. At this point, results under CRSC are at least independent of the ones where RCPLD are employed. We proceed with the example where CRSC holds and RCPLD fails.

Example 5. Let $\mathcal{M}$ be an n-dimensional smooth Riemannian manifold with $n \geq 2$. Take $p \in \mathcal{M}$ and $g := (g_1, \ldots, g_4) : \mathcal{M} \to \mathbb{R}^4$ continuously differentiable functions satisfying the following conditions:

(i) $g(p) = 0$;

(ii) $\operatorname{grad} g_1(p) = -\operatorname{grad} g_2(p)$ and $\operatorname{grad} g_3(p) = -\operatorname{grad} g_4(p)$;

(iii) for all $\epsilon > 0$, there exists $q \in B_\epsilon(p)$ such that the set $\{\operatorname{grad} g_1(q), \operatorname{grad} g_2(q)\}$ is linearly independent with $q \neq p$;

(iv) the set $\{\operatorname{grad} g_1(p), \operatorname{grad} g_3(p)\}$ is linearly independent;

(v) there exists $\epsilon > 0$ such that $rank\{\operatorname{grad} g_j(q) : j = 1, \ldots, 4\} = 2$ for all $q \in B_\epsilon(p)$.

Consider a feasible set $\Omega := \{q \in \mathcal{M} : g(q) \leq 0\}$ and $p \in \Omega$. It follows from condition (ii) that the set $\{\operatorname{grad} g_1(p), \operatorname{grad} g_2(p)\}$ is positive-linearly dependent. Hence, using conditions (i) and (iii), we conclude that RCPLD does not hold at p. In order to see that CRSC is valid at p, it is enough to note (v) together with the fact that (i) and (ii) imply that $\mathcal{J}_-(p) = \{1, 2, 3, 4\}$. Next, we will present two examples in which conditions (i), (ii), (iii), (iv), and (v) are satisfied.

- Consider the sphere $\mathcal{M} := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$, and take $p := (0, 0, 1)$, $g_1(x, y, z) := x - y^2$, $g_2(x, y, z) := -x$, $g_3(x, y, z) := y - x^2$, and $g_4(x, y, z) := -y$, where clearly (i) holds. Similarly to Example 3, let $q := (x, y, z)$ with $y \neq 0$ and $z \neq 0$, $u_1 := (1, -2y, 0)$, $u_2 := (-1, 0, 0)$, $u_3 := (-2x, 1, 0)$, and $u_4 := (0, -1, 0)$. Since $u_i \in \{p\}^\perp, i = 1, 2, 3, 4$, it is easy to see (ii) and (iv). In order to prove (iii), notice that $\{q, u_1, u_2\}$ is linearly independent. Hence, since $\operatorname{grad} g_i(q) = \Pi_{\{q\}^\perp}(u_i) = u_i - r_{u_i} q, r_{u_i} \in \mathbb{R}$ for $i = 1, 2, 3, 4$, similarly to Example 3 we have that $\{\operatorname{grad} g_1(q), \operatorname{grad} g_2(q)\}$ is linearly independent. To see that (v) holds, take $\epsilon > 0$ such that $z \neq 0$ for all $q = (x, y, z) \in B_\epsilon(p)$. Hence, $\{u_2, u_4, q\}$ is linearly independent, which implies that $\{\operatorname{grad} g_2(q), \operatorname{grad} g_4(q)\}$ is linearly independent. From the fact that $\mathcal{M}$ is 2-dimensional, (v) holds.

- Consider a 2-dimensional complete manifold $\mathcal{M}$ and define the functions $g_1(q) := \varphi_1(q)$, $g_2(q) := \varphi_2(q)$, $g_3(q) := \varphi_3(q)$, and $g_4(q) := -\varphi_3(q)$ from (3.1). Similarly to the computations in Example 3, one can prove that items (i), (ii), (iii), and (iv) are satisfied. By (iv) and Lemma 1.2, we have that the rank of $\{\operatorname{grad} g_j(q) : j = 1, \ldots, 4\}$ is at least 2 for all $q \in B_\epsilon(p)$ and some $\epsilon > 0$. Therefore, taking into account that $\mathcal{M}$ is 2-dimensional, condition (v) is also satisfied. Figure 3 illustrates this example.

In our view, CRSC is the most interesting one of all previously defined conditions. Although we do not pursue its extensions to the Riemannian setting, we mention a few of its properties known in the Euclidean case. First, it has an elegant mathematical description. Also, the index set $\mathcal{J}_-(p)$ can be viewed as the index set of active inequality constraints that are treated as equality constraints in the polar of the linearized cone $\mathcal{L}^\circ(p)$ (2.4). Actually, this interpretation holds also for the cone $\mathcal{L}(p)$, since $\mathcal{J}_-(p)$ can be equivalently stated as the set of indexes $j \in \mathcal{A}(p)$ such that
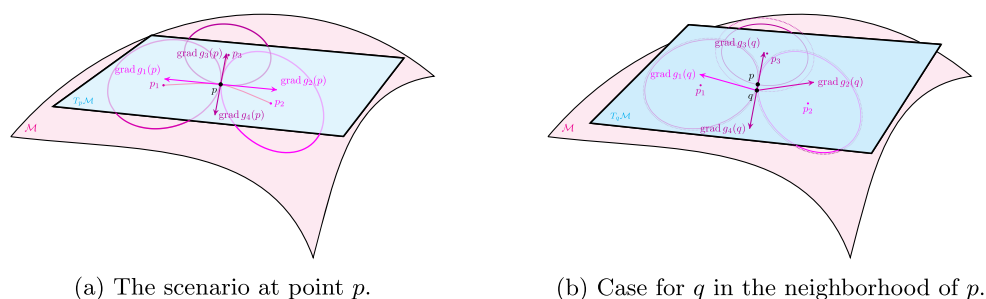
(a) The scenario at point p.   (b) Case for q in the neighborhood of p.

FIG. 3. *Illustrative figure for Example* 5, *where RCPLD fails but CRSC is satisfied. Although the subset of gradients indexed by* $\{1,2\}$ *loses positive linear dependence for q near p, all gradients remain to span a* 2-*dimensional vector space (the whole tangent space) for q near p.*

$\langle \operatorname{grad} g_j(p), v \rangle = 0$ for all $v \in \mathcal{L}(p)$ [41]. However, surprisingly, these interpretations are not by chance, since it was proved [13] that when $p \in \Omega$ satisfies CRSC, the constraints $g_j(q) \le 0, j \in \mathcal{J}_-(p)$ can only be satisfied as equalities for a feasible point q in a small enough neighborhood around p. That is, one can safely replace the inequalities $g_j(q) \le 0, j \in \mathcal{J}_-(p)$ with equalities $g_j(q) = 0, j \in \mathcal{J}_-(p)$ without locally altering the feasible set and in such a way that MFCQ holds. This result is connected to the one in [46] which shows that whenever CRCQ is satisfied, there exists a local reformulation of the problem such that MFCQ holds. In fact, this procedure is well known in linear conic programming as *facial reduction;* that is, when the constraints are such that there is no interior point, there is an efficient procedure to replace the cone with one of its faces in such a way that a relative interior point exists. The extension of CRSC to the conic context and its connections with the facial reduction procedure are described in [8], where they also show that CRSC may also provide the strong second-order necessary optimality condition depending on a single Lagrange multiplier by considering the constant rank property for all subsets of constraints that include $\mathcal{J}_-(p)$ and all equalities. Finally, we mention an additional property that holds in the Euclidean setting for all CQs discussed in this paper; that is, that they all imply that an error bound can be computed. We state this as the following conjecture in the Riemannian setting:

Conjecture: Let $\mathcal{M}$ be a complete Riemannian manifold with dimension $n \ge 2$. Let $p \in \Omega$ be such that CRSC or RCPLD is satisfied. Then, there exists $\epsilon > 0$ and $\alpha > 0$ such that

$$\inf_{w \in \Omega} d(q, w) \le \alpha \max\{0, g_1(q), \dots, g_m(q), |h_1(q)|, \dots, |h_s(q)|\}$$

for all $q \in B_\epsilon(p)$.

Finally, let us prove that all conditions proposed so far are CQs. We do this by showing that they are strict CQs with respect to the necessary optimality condition AKKT from Theorem 2.1, since this gives us the main result of global convergence of Algorithm 2.1. We start by showing that CRSC is a strict CQ, and we note that when the condition was introduced in the Euclidean setting [13], only an indirect proof of this fact was presented. Thus, a clear direct proof was not available in the literature even in the Euclidean setting.

THEOREM 3.4. *Suppose that* $p \in \Omega$ *satisfies CRSC. If* p *is an AKKT point, then* p *is a KKT point.*

Proof. As in Definition 3.3, let $\mathcal{J}_-(p) := \{j \in \mathcal{A}(p) \colon -\operatorname{grad} g_j(p) \in \mathcal{L}(p)^\circ\}$, and we denote $\mathcal{J}_+(p) := \mathcal{A}(p) \backslash \mathcal{J}_-(p)$. Since p is an AKKT point of problem (2.1), there exist sequences $(p^k)_{k \in \mathbb{N}} \subset \mathcal{M}$, $(\lambda^k)_{k \in \mathbb{N}} \subset \mathbb{R}^s$, and $(\mu^k)_{k \in \mathbb{N}} \subset \mathbb{R}_+^m$ with $\mu_j^k = 0$ for all $j \notin \mathcal{A}(p)$ such that $\lim_{k \to \infty} p^k = p$ and

(3.2)

$$\operatorname{grad} f(p^k) + \sum_{i=1}^s \lambda_i^k \operatorname{grad} h_i(p^k) + \sum_{\ell \in \mathcal{J}_-(p)} \mu_\ell^k \operatorname{grad} g_\ell(p^k) + \sum_{j \in \mathcal{J}_+(p)} \mu_j^k \operatorname{grad} g_j(p^k)$$

$$=: \epsilon_k \quad \forall k \in \mathbb{N},$$

where $\lim_{k \to \infty} \epsilon_k = 0$. Let $\mathcal{I} \subset \{1, \dots, s\}$ and $\mathcal{J} \subset \mathcal{J}_-(p)$ be such that $A(p, \mathcal{I}, \mathcal{J})$ is a basis for the subspace generated by $A(p, \{1, \dots, s\}, \mathcal{J}_-(p))$ (here we use the notation introduced in (1.1)). By CRSC, $A(p^k, \mathcal{I}, \mathcal{J})$ is also a basis for the subspace generated by $A(p^k, \{1, \dots, s\}, \mathcal{J}_-(p))$, and thus (3.2) can be rewritten as

(3.3)

$$\operatorname{grad} f(p^k) + \sum_{i \in \mathcal{I}} \tilde{\lambda}_i^k \operatorname{grad} h_i(p^k) + \sum_{\ell \in \mathcal{J}} \tilde{\mu}_\ell^k \operatorname{grad} g_\ell(p^k) + \sum_{j \in \mathcal{J}_+(p)} \mu_j^k \operatorname{grad} g_j(p^k) = \epsilon_k \ \forall k \in \mathbb{N},$$

for suitable $\tilde{\lambda}_i^k \in \mathbb{R}, i \in \mathcal{I}$ and $\tilde{\mu}_\ell^k \in \mathbb{R}, \ell \in \mathcal{J}$. If all the sequences $(\tilde{\lambda}_i^k)_{k \in \mathbb{N}}, i \in \mathcal{I}$, $(\tilde{\mu}_\ell^k)_{k \in \mathbb{N}}, \ell \in \mathcal{J}$, and $(\mu_j^k)_{k \in \mathbb{N}}, j \in \mathcal{J}_+(p)$ are bounded, we may take a suitable convergent subsequence $(\tilde{\lambda}_i^k)_{k \in K_1} \to \tilde{\lambda}_i \in \mathbb{R}, i \in \mathcal{I}$, $(\tilde{\mu}_\ell^k)_{k \in K_1} \to \tilde{\mu}_\ell \in \mathbb{R}, \ell \in \mathcal{J}$, and $(\mu_j^k)_{k \in K_1} \to \mu_j \in \mathbb{R}_+, j \in \mathcal{J}_+(p)$ such that

$$\operatorname{grad} f(p) + \sum_{i \in \mathcal{I}} \tilde{\lambda}_i \operatorname{grad} h_i(p) + \sum_{\ell \in \mathcal{J}} \tilde{\mu}_\ell \operatorname{grad} g_\ell(p) + \sum_{j \in \mathcal{J}_+(p)} \mu_j \operatorname{grad} g_j(p) = 0.$$

Let us see that this implies that p is a KKT point. First, note that $\mathcal{J} \subset \mathcal{J}_-(p)$ with $\mathcal{J}_-(p) \cup \mathcal{J}_+(p) = A(p)$. If some $\tilde{\mu}_{\ell_0} < 0$, $\ell_0 \in \mathcal{J}$, by the definition of the set $\mathcal{J}_-(p)$, we have $\tilde{\mu}_{\ell_0} \operatorname{grad} g_{\ell_0} \in \mathcal{L}^\circ(p)$. But from (2.4), one can see that $\mathcal{L}^\circ(p)$ is closed under addition, which implies that $-\operatorname{grad} f(p) \in \mathcal{L}^\circ(p)$; that is, p is a KKT point. Otherwise, if it is not the case that all sequences are bounded, let us take a subsequence $K_2 \subset \mathbb{N}$ such that $\lim_{k \in K_2} M_k = +\infty$, where $M_k = \max\{|\tilde{\lambda}_i^k|, i \in \mathcal{I}; |\tilde{\mu}_\ell^k|, \ell \in \mathcal{J}; \mu_j^k, j \in \mathcal{J}_+(p)\}$. Dividing (3.3) by M_k and taking the limit on a suitable subsequence $K_3 \subset K_2$ such that $\lim_{k \in K_3} \frac{\tilde{\lambda}_i^k}{M_k} = \alpha_i \in \mathbb{R}$, $\lim_{k \in K_3} \frac{\tilde{\mu}_\ell^k}{M_k} = \beta_\ell \in \mathbb{R}$, and $\lim_{k \in K_3} \frac{\mu_j^k}{M_k} = \gamma_j \geq 0$ with not all $\alpha_i, \beta_\ell, \gamma_j$ equal to zero, we arrive at

$$\sum_{i \in \mathcal{I}} \alpha_i \operatorname{grad} h_i(p) + \sum_{\ell \in \mathcal{J}} \beta_\ell \operatorname{grad} g_\ell(p) + \sum_{j \in \mathcal{J}_+(p)} \gamma_j \operatorname{grad} g_j(p) = 0.$$

Let us now show that $\gamma_j = 0$ for all $j \in \mathcal{J}_+(p)$. Suppose that $\gamma_{j_0} > 0$ for some $j_0 \in \mathcal{J}_+(p)$ and let us isolate $-\gamma_{j_0} \operatorname{grad} g_{j_0}(p)$ in the above equality. The remaining sum is a sum of vectors in $\mathcal{L}(p)^\circ$ and hence $-\operatorname{grad} g_{j_0}(p) \in \mathcal{L}(p)^\circ$, which contradicts the definition of $\mathcal{J}_+(p)$. Notice that this is the case since if $\beta_\ell < 0$, we must have $\beta_\ell \operatorname{grad} g_\ell(p) \in \mathcal{L}(p)^\circ$ by the definition of $\mathcal{J}_-(p)$. Therefore $A(p, \mathcal{I}, \mathcal{J})$ must be linearly dependent, which contradicts the definition of the index sets $\mathcal{I}$ and $\mathcal{J}$, concluding the proof. $\square$

Similarly, we show that RCPLD is a strict CQ.

THEOREM 3.5. *Suppose that $p \in \Omega$ satisfies RCPLD. If p is an AKKT point, then p is a KKT point.*

Proof. The proof is similar to the previous one, but without partitioning $A(p)$. That is, consider the previous proof with $\mathcal{J}_-(p)$ replaced by $\emptyset$ and $\mathcal{J}_+(p)$ replaced by $\mathcal{A}(p)$. We arrive similarly to (3.3) to a sequence

$$\operatorname{grad} f(p^k) + \sum_{i \in \mathcal{I}} \tilde{\lambda}_i^k \operatorname{grad} h_i(p^k) + \sum_{j \in \mathcal{A}(p)} \mu_j^k \operatorname{grad} g_j(p^k) = \epsilon_k \quad \forall k \in \mathbb{N},$$

with $\tilde{\lambda}_i^k \in \mathbb{R}, i \in \mathcal{I}, \mu_j^k \geq 0, j \in \mathcal{A}(p), \lim_{k \in \mathbb{N}} p^k = p, \lim_{k \in \mathbb{N}} \epsilon_k = 0$, and $\{\operatorname{grad} h_i(p) : i \in \mathcal{I}\}$ linearly independent.

For every $k \in \mathbb{N}$, we apply Lemma 1.4 to arrive at

$$(3.4) \qquad \operatorname{grad} f(p^k) + \sum_{i \in \mathcal{I}} \bar{\lambda}_i^k \operatorname{grad} h_i(p^k) + \sum_{j \in \mathcal{J}_k} \bar{\mu}_j^k \operatorname{grad} g_j(p^k) = \epsilon_k,$$

for some $\bar{\lambda}_i^k \in \mathbb{R}, i \in \mathcal{I}, \bar{\mu}_j^k \geq 0, j \in \mathcal{J}_k \subset \mathcal{A}(p)$ and all $k \in \mathbb{N}$, where $A(p^k, \mathcal{I}, \mathcal{J}_k)$ is linearly independent. Let us take a subsequence such that $\mathcal{J}_k$ is constant, say, $\mathcal{J}_k \equiv \mathcal{J}$ for all $k \in K_1 \subset \mathbb{N}$. The proof now follows similarly to the previous one considering $M_k := \max\{|\bar{\lambda}_i^k|, i \in \mathcal{I}; \bar{\mu}_j^k, j \in \mathcal{J}\}$. If $(M_k)_{k \in K_1}$ is bounded, one may take the limit in (3.4) for a suitable subsequence to see that p is a KKT point. Otherwise, dividing (3.4) by M_k we see that $A(p, \mathcal{I}, \mathcal{J})$ is positive-linearly dependent, which contradicts the definition of RCPLD. □

We formalize our results in the following.

COROLLARY 3.6. *Let p be a feasible limit point of a sequence $(p^k)_{k \in \mathbb{N}}$ generated by Algorithm 2.1 such that p satisfies RCPLD or CRSC. Then p satisfies the KKT conditions.*

Notice that differently from the result under MFCQ, where the dual AKKT sequence $(\lambda^k, \mu^k)_{k \in \mathbb{N}}$ is necessarily bounded, and thus dual convergence to a Lagrange multiplier is obtained, our result does not include convergence of the dual sequence. In the next section we prove that this can be obtained under a condition weaker than CPLD and independent of RCPLD known as quasinormality [35]. In order to do this, we shall extend a stronger sequential optimality condition to the Riemannian setting known as the PAKKT condition [3].

4. A stronger sequential optimality condition. The QN was introduced in [35], and it was popularized in the book [20] in connection with convergence of the external penalty method. Recently, it has been connected with the notion of so-called enhanced KKT conditions, guaranteeing boundedness of the corresponding set of enhanced Lagrange multipliers [18]. QN is a fairly weak CQ and has been known to be strictly weaker than CPLD [17] while still implying the error bound property [49] in the Euclidean setting. In this section we will extend an important algorithmic property of QN that goes beyond what we have proved for RCPLD and CRSC. That is, besides QN being a strict CQ with respect to the AKKT condition, namely, global convergence of Algorithm 2.1 in the sense that Corollary 3.6 is also valid under QN, we will show that the dual sequence generated by Algorithm 2.1 under QN is in fact bounded. In order to do this, we will show that QN is a strict CQ with respect to a stronger sequential optimality condition known as the PAKKT condition [3].

We start by introducing the PAKKT condition in the Riemannian setting, showing that it is indeed a genuine necessary optimality condition for problem (2.1). Our definition considers a modification of the original one as suggested in [11].

DEFINITION 4.1. *The PAKKT condition for problem* (2.1) *is satisfied at a point* $p \in \Omega$ *if there exist sequences* $(p^k)_{k \in \mathbb{N}} \subset \mathcal{M}$, $(\lambda^k)_{k \in \mathbb{N}} \subset \mathbb{R}^s$, *and* $(\mu^k)_{k \in \mathbb{N}} \subset \mathbb{R}_+^m$ *such that*

    (i) $\lim_{k \to \infty} p^k = p$;
    (ii) $\lim_{k \to \infty} \left\| \operatorname{grad} L(p^k, \lambda^k, \mu^k) \right\| = 0$;
    (iii) $\mu_j^k = 0$ *for all* $j \notin \mathcal{A}(p)$ *and sufficiently large* k;
    (iv) *If* $\gamma_k := \left\| (1, \lambda^k, \mu^k) \right\|_\infty \to +\infty$ *it holds:*

$$(4.1) \qquad \lim_{k \to \infty} \frac{\left| \lambda_i^k \right|}{\gamma_k} > 0 \quad \Longrightarrow \quad \lambda_i^k h_i(p^k) > 0 \, \forall k \in \mathbb{N};$$

$$(4.2) \qquad \lim_{k \to \infty} \frac{\mu_j^k}{\gamma_k} > 0 \quad \Longrightarrow \quad \mu_j^k g_j(p^k) > 0 \, \forall k \in \mathbb{N}.$$

A point p satisfying Definition 4.1 is called a PAKKT point; the correspondent sequence $(p^k)_{k \in \mathbb{N}}$ is its associated primal sequence, while $(\lambda^k, \mu^k)_{k \in \mathbb{N}}$ is its associated dual sequence. In order to present our results, we will make use of the following lemmas extended to the Riemannian setting in [59].

LEMMA 4.2. *Let* p *be a local minimizer of problem* (2.1) *and* $\alpha > 0$. *Then, for each* $k \in \mathbb{N}$ *and* $\rho_k > 0$, *the problem*

$$\operatorname*{Minimize}_{q \in \mathcal{M}} f(q) + \frac{1}{2} d(q, p)^2 + \frac{\rho_k}{2} \left(\|h(q)\|_2^2 + \|g(q)_+\|_2^2 \right),$$
$$\textit{subject to } d(q, p) \leq \alpha,$$

admits a solution p^k. *Moreover, if* $\lim_{k \to \infty} \rho_k = +\infty$, *then* $\lim_{k \to \infty} p^k = p$.

LEMMA 4.3. *Let* $\phi : \mathcal{M} \to \mathbb{R}$ *be a differentiable function, let* $\alpha > 0$, *and let* $p_0 \in \mathcal{M}$. *Suppose that* $p \in \mathcal{M}$ *is an optimal solution of the following optimization problem:*

$$\operatorname*{Minimize}_{q \in \mathcal{M}} \phi(q),$$
$$\textit{subject to } d(q, p_0) \leq \alpha.$$

If $d(p, p_0) < \alpha$, *then* $\operatorname{grad} \phi(p) = 0$.

We now show that PAKKT is a genuine necessary optimality condition for problem (2.1).

THEOREM 4.4. *Let* $p \in \Omega$ *be a local minimizer of* (2.1). *Then,* p *is a PAKKT point.*

Proof. Let p be a local minimizer of problem (2.1). Thus, there is a sufficiently small parameter $\alpha > 0$ such that the problem

$$\operatorname*{Minimize}_{q \in \mathcal{M}} f(q) + \frac{1}{2} d(q, p)^2,$$
$$\textit{subject to } h(q) = 0, \ g(q) \leq 0, \ d(q, p) \leq \alpha,$$

has p as the unique global minimizer. For each $k \in \mathbb{N}$, take $\rho_k > 0$ such that $\lim_{k \to \infty} \rho_k = +\infty$. Consider the penalized problem

$$(4.3) \qquad \operatorname*{Minimize}_{q \in \mathcal{M}} f(q) + \frac{1}{2} d(q, p)^2 + \frac{\rho_k}{2} \left(\|h(q)\|_2^2 + \|g(q)_+\|_2^2 \right),$$
$$\textit{subject to } d(q, p) \leq \alpha.$$

It follows from Lemma 4.2 that there exists a sequence $(p^k)_{k \in \mathbb{N}}$ such that p^k is a solution of (4.3) and $\lim_{k \to \infty} p^k = p$. Thus, item (i) of Definition 4.1 is satisfied. Moreover, there exists an infinite index set K_1 such that $d(p^k, p) < \alpha$ for all $k \in K_1$. Consequently, using Lemma 4.3, we conclude that

$$\operatorname{grad} f(p^k) - \exp_{p^k}^{-1} p + \sum_{i=1}^{s} \rho_k h_i(p^k) \operatorname{grad} h_i(p^k) + \sum_{j=1}^{m} \rho_k \max\{0, g_j(p^k)\} \operatorname{grad} g_j(p^k) = 0$$

for all $k \in K_1$. Therefore, we have

$$\lim_{k \in K_1} \operatorname{grad} L(p^k, \lambda^k, \mu^k) = \lim_{k \in K_1} \left(\operatorname{grad} f(p^k) + \sum_{i=1}^{s} \lambda_i^k \operatorname{grad} h_i(p^k) + \sum_{j=1}^{m} \mu_j^k \operatorname{grad} g_j(p^k) \right)$$
$$= \lim_{k \in K_1} - \exp_{p^k}^{-1} p = 0,$$

where, for each $k \in K_1$, we denote $\lambda^k := \rho_k h(p^k)$ and $\mu^k := \rho_k [g(p^k)]_+ \geq 0$. Therefore, (ii) and (iii) of Definition 4.1 are satisfied. We will now analyze the validity of (4.1) and (4.2). Let $\gamma_k := \left\| (1, \lambda^k, \mu^k) \right\|_\infty$ for all $k \in K_1$ be such that $\lim_{k \in K_1} M_k = +\infty$, and assume that $\lim_{k \in K_1}(|\lambda_i^k|/\gamma_k) > 0$. Thus, $|\lambda_i^k|/\gamma_k > 0$ for sufficiently large $k \in K_1$, which implies that $h(p^k) \neq 0$. Hence, $\lambda_i^k h_i(p^k) = \rho_k h_i(p^k)^2 > 0$ for all sufficiently large $k \in K_1$. Similarly, if $\lim_{k \in K_1}(\mu_j^k/\gamma_k) > 0$, we have $\mu_j^k > 0$, which implies $g_j(p^k) > 0$ for sufficiently large $k \in K_1$. Therefore, (4.1) and (4.2) are fulfilled. Consequently, p satisfies Definition 4.1, which concludes the proof. □

Let us now introduce QN in the Riemannian context. We will show that under QN the dual sequence $(\lambda^k, \mu^k)_{k \in \mathbb{N}}$ associated with any PAKKT sequence $(p^k)_{k \in \mathbb{N}}$ is bounded. Later, we will show that Algorithm 2.1 generates PAKKT sequences, which will provide the main algorithmic relevance of QN.

DEFINITION 4.5. *Let Ω be given by (2.2), let $p \in \Omega$, and let $\mathcal{A}(p)$ be given by (2.3). The point p satisfies the QN if there are no $\lambda \in \mathbb{R}^s$ and $\mu \in \mathbb{R}_+^m$ such that*
  (i) $\sum_{i=1}^{s} \lambda_i \operatorname{grad} h_i(p) + \sum_{j \in \mathcal{A}(p)} \mu_j \operatorname{grad} g_j(p) = 0$;
  (ii) $\mu_j = 0$ *for all $j \notin \mathcal{A}(p)$ and $(\lambda, \mu) \neq 0$;*
  (iii) *for all $\epsilon > 0$ there exists $q \in B_\epsilon(p)$ such that $\lambda_i h_i(q) > 0$ for all $i \in \{1, \dots, s\}$ with $\lambda_i \neq 0$ and $\mu_j g_j(q) > 0$ for all $j \in \mathcal{A}(p)$ with $\mu_j > 0$.*

In the next example we show that QN holds, but both RCPLD and CRSC fail.

Example 6. Define the functions $h_1(q) := \varphi_1(q)e^{\varphi_2(q)}$ and $h_2(q) := \varphi_1(q)$ as defined in (3.1). Note that

(4.4)
$$\operatorname{grad} h_1(q) := e^{\varphi_2(q)} \operatorname{grad} \varphi_1(q) + \varphi_1(q)e^{\varphi_2(q)} \operatorname{grad} \varphi_2(q), \qquad \operatorname{grad} h_2(q) := \operatorname{grad} \varphi_1(q).$$

The point $p \in \Omega$ satisfies QN. Indeed, we first note that $\varphi_1(p) = 0$ and $\varphi_2(p) = 0$. Moreover, we have $\operatorname{grad} h_1(p) = \operatorname{grad} h_2(p) = \operatorname{grad} \varphi_1(p)$. Consider the linear combination $\lambda_1 \operatorname{grad} h_1(p) + \lambda_2 \operatorname{grad} h_2(p) = 0$ with λ_1 and $\lambda_2 \in \mathbb{R}$. Thus, we have $(\lambda_1 + \lambda_2) \operatorname{grad} \varphi_1(p) = 0$. Since $\operatorname{grad} \varphi_1(p) \neq 0$, we conclude that unless $\lambda_1 = \lambda_2 = 0$, we must have $\lambda_1 \lambda_2 < 0$. In this case, take $\epsilon > 0$ and $q \in B_\epsilon(p)$ such that $q \neq p$. Since $h_1(q)h_2(q) > 0$, we conclude that $\lambda_1 h_1(q)$ and $\lambda_2 h_2(q)$ have opposite signs, which implies that p satisfies QN.

Now, we are going to show that p does not satisfy RCPLD nor CRSC. For that, we first note that rank of $\{\operatorname{grad} h_1(p), \operatorname{grad} h_2(p)\}$ is equal to one. On the other hand,

similarly to the computations in Example 3, one can prove that, for all $\epsilon > 0$, there exists $q \in B_\epsilon(p)$ such that $\{\operatorname{grad}\varphi_1(q), \operatorname{grad}\varphi_2(q)\}$ is linearly independent with $q \neq p$. By the definition of φ_1, notice that $\varphi_1(q) \neq 0$ for all $q \in B_\epsilon(p)$ and sufficiently small $\varepsilon > 0$; it follows from (4.4) that $\{\operatorname{grad}h_1(q), \operatorname{grad}h_2(q)\}$ is also linearly independent. Therefore, p does not satisfy RCPLD or CRSC.

THEOREM 4.6. *Let $p \in \Omega$ 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, and any limit point of $(\lambda^k, \mu^k)_{k\in\mathbb{N}}$ is a Lagrange multiplier associated with p.*

Proof. Let $p \in \Omega$ be a PAKKT point with primal sequence $(p^k)_{k\in\mathbb{N}}$ and dual sequence $(\lambda^k, \mu^k)_{k\in\mathbb{N}}$, and let us assume that the dual sequence is unbounded. Then, we will conclude that the point p does not satisfy the quasinormality condition; i.e., we will prove the existence of $\lambda \in \mathbb{R}^s$ and $\mu \in \mathbb{R}^m_+$ such that items (i), (ii), and (iii) of Definition 4.5 are satisfied. For that, set $\gamma_k = \left\|(1, \lambda^k, \mu^k)\right\|_\infty$ as in Definition 4.1 and take an infinite subsequence indexed by K_1 such that $\lim_{k\in K_1} \gamma_k = +\infty$. To simplify the notations let us define the auxiliary sequence

$$U^k := (1, \lambda^k, \mu^k) \in \mathbb{R} \times \mathbb{R}^s \times \mathbb{R}^m_+ \qquad \forall k \in \mathbb{N},$$

with $\lim_{k\in K_1} \|U^k\|_2 = \infty$. Take an infinite subset $K_2 \subset K_1$ such that the sequence $(U^k/\|U^k\|_2)_{k\in K_2}$ converges to some $(0, \lambda, \mu) \in \mathbb{R} \times \mathbb{R}^s \times \mathbb{R}^m_+$, with $\|(0, \lambda, \mu)\| = 1$. Thus, considering that $(p^k)_{k\in\mathbb{N}}$ is a primal PAKKT sequence, we conclude that

$$\lim_{k\in K_2} \frac{\operatorname{grad}L(p^k, \lambda^k, \mu^k)}{\gamma_k}$$
$$= \lim_{k\in K_2} \left(\frac{\operatorname{grad}f(p^k)}{\gamma_k} + \sum_{i=1}^s \frac{\lambda_i^k}{\gamma_k} \operatorname{grad}h_i(p^k) + \sum_{j=1}^m \frac{\mu_j^k}{\gamma_k} \operatorname{grad}g_j(p^k) \right) = 0.$$

Hence, taking into account that $\mu_j = 0$ for $j \notin \mathcal{A}(p)$, we obtain that item (i) of Definition 4.5 is satisfied at p. In addition, since $(\lambda, \mu) \neq 0$, item (ii) of Definition 4.5 is also satisfied at p. From (4.1) and (4.2), we have that $\lambda_i^k h_i(p^k) > 0$ whenever $\lambda_i \neq 0$, and $\mu_j^k g_j(p^k) > 0$ whenever $\mu_j > 0$, which gives precisely item (iii) of Definition 4.5. Therefore, QN fails.   ∎

Finally, it remains for us to show that Algorithm 2.1 generates PAKKT sequences, which gives its global convergence result under QN.

THEOREM 4.7. *Assume Algorithm 2.1 generates an infinite sequence $(p^k)_{k\in\mathbb{N}}$ with a feasible accumulation 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 2.1. In particular, p is a KKT point, and any limit point of $(\lambda^k, \mu^k)_{k\in K}$ is a Lagrange multiplier associated with p.*

Proof. By Step 1 and Step 2 of the algorithm, we have

$$\lim_{k\in K} \operatorname{grad}L(p^k, \lambda^k, \mu^k) = \lim_{k\in K} \operatorname{grad}\mathcal{L}_{\rho_k}(p^k, \bar\lambda^k, \bar\mu^k) = 0,$$

with $\mu_j^k = 0$ for sufficiently large $k \in K$ if $j \notin \mathcal{A}(p)$. To see this, note that when $(\rho_k)_{k\in K}$ is unbounded, this follows from the definition of μ_j^k, the boundedness of $(\bar\mu_j^k)_{k\in K}$, and the fact that $\rho_k g_j(p^k) \to -\infty$. When $(\rho_k)_{k\in K}$ is bounded, we must have from Step 3 that $v^k \to 0$. In particular, $\max\{0, \frac{\bar\mu_j^k}{\rho_k} + g_j(p^k)\} - \frac{\bar\mu_j^k}{\rho_k} \to 0$. Since $g_j(p^k) < 0$ is bounded

away from zero when $j \notin \mathcal{A}(p)$, we must have $\mu_j^k = 0$ for sufficiently large $k \in K$. Thus (i), (ii), and (iii) of Definition 4.1 hold.

To prove (iv) of Definition 4.1, let $\gamma_k := \left\|(1, \lambda^k, \mu^k)\right\|_\infty$ and assume that $\lim_{k \in K} \gamma_k = +\infty$. Let $i \in \{1, \ldots, s\}$ be such that $\lim_{k \in K} \frac{\lambda_i^k}{\gamma_k} = \lambda_i \neq 0$ and $j \in \mathcal{A}(p)$ be such that $\lim_{k \in K} \frac{\mu_j^k}{\gamma_k} = \mu_j > 0$. This implies that $\lambda_i^k = \bar{\lambda}_i^k + \rho_k h_i(p^k)$ is unbounded for $k \in K$, with $\lambda_i^k \lambda_i > 0$. Since $(\bar{\lambda}_i^k)_{k \in K}$ is bounded, the only possibility is that $\rho_k \to +\infty$ and $\lambda_i h_i(p^k) > 0$ for sufficiently large $k \in K$. Similarly, we have $\mu_j g_j(p^k) > 0$ for sufficiently large $k \in K$, which completes the proof. $\square$

We conclude by providing another property of QN in connection with Algorithm 2.1. Instead of considering (2.5) in Step 1 of Algorithm 2.1, one may consider a more flexible criterion for solving the correspondent subproblem. That is, instead of requiring the iterate p^k to satisfy $\left\|\operatorname{grad} \mathcal{L}_{\rho_k}(p^k, \bar{\lambda}^k, \bar{\mu}^k)\right\| \leq \epsilon_k$, one may require the looser criterion $\left\|\frac{\operatorname{grad} \mathcal{L}_{\rho_k}(p^k, \bar{\lambda}^k, \bar{\mu}^k)}{\gamma_k}\right\| \leq \epsilon_k$, where $\gamma_k := \left\|(1, \lambda^k, \mu^k)\right\|_\infty$ with λ^k and μ^k given as in Step 2 of the algorithm. That is, one abdicates robustness of the solution of the subproblem in place of an easier computable iterate. For instance, this is the approach considered in the well-known interior point method IPOPT [57], even though it tends to generate unbounded dual sequences [34]. This modification gives rise to the so-called scaled-PAKKT condition [11] which we present in the Riemannian setting as follows.

DEFINITION 4.8. *The scaled-PAKKT condition for problem* (2.1) *is satisfied at a point* $p \in \Omega$ *if there exist sequences* $(p^k)_{k \in \mathbb{N}} \subset \mathcal{M}$, $(\lambda^k)_{k \in \mathbb{N}} \subset \mathbb{R}^s$, *and* $(\mu^k)_{k \in \mathbb{N}} \subset \mathbb{R}_+^m$ *such that the following hold:*
  (i) $\lim_{k \to \infty} p^k = p$;
  (ii) $\lim_{k \to \infty} \left\|\frac{\operatorname{grad} L(p^k, \lambda^k, \mu^k)}{\gamma_k}\right\| = 0$, *where* $\gamma_k := \left\|(1, \lambda^k, \mu^k)\right\|_\infty$;
  (iii) $\mu_j^k = 0$ *for all* $j \notin \mathcal{A}(p)$ *and sufficiently large* k;
  (iv) *If* $\gamma_k \to +\infty$, *then*

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

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

It is easy to see that Algorithm 2.1 with the looser criterion in Step 1 as described previously generates scaled-PAKKT sequences. Now, it is easy to see that QN is still sufficient for guaranteeing boundedness of the dual scaled-PAKKT sequence, following the proof of Theorem 4.6; however, let us show that QN is somewhat the weakest condition with this property.

THEOREM 4.9. *If, for each continuously differentiable function* $f : \mathcal{M} \to \mathbb{R}$ *such that* $p \in \Omega$ *is a scaled-PAKKT point, the KKT conditions also hold, then* p *satisfies* QN *or* $\mathcal{L}(p)^\circ = T_p \mathcal{M}$.

Proof. Assume that the point p does not satisfy the quasinormality condition and $\mathcal{L}(p)^\circ \neq T_p \mathcal{M}$. We will show that there exists a continuously differentiable function $f : \mathcal{M} \to \mathbb{R}$ such that p is a scaled-PAKKT point, but p is not a KKT point. Since $\mathcal{L}(p)^\circ \neq T_p \mathcal{M}$, taking into account that $\mathcal{L}(p)^\circ \subset T_p \mathcal{M}$ and $0 \in \mathcal{L}(p)^\circ$, we concluded that there exists $v \in T_p \mathcal{M}$ with $v \neq 0$ such that $v \notin \mathcal{L}(p)^\circ$. Thus, by the definition of $\mathcal{L}(p)^\circ$ in (2.4), we have

$$(4.5) \qquad -v + \sum_{i=1}^{s} \bar{\lambda}_i \operatorname{grad} h_i(p) + \sum_{j \in \mathcal{A}(p)} \bar{\mu}_j \operatorname{grad} g_j(p) \neq 0, \quad \forall \bar{\mu}_j \geq 0, \forall \bar{\lambda}_i \in \mathbb{R}.$$

To proceed, we take $0 < \delta < r_p$, the injectivity radius, and define $f : B_\delta(p) \to \mathbb{R}$ by $f(q) := \langle -v, -\exp_p^{-1} q \rangle$, which is continuously differentiable, and $\operatorname{grad} f(p) = -v$. Hence, (4.5) implies that p is not a KKT point. It remains to show that p is a scaled-PAKKT point. Since p does not satisfy QN, there exist $\lambda \in \mathbb{R}^s$ and $\mu \in \mathbb{R}_+^m$ that satisfy items (i), (ii), and (iii) of Definition 4.5. In particular, by item (iii), let $(p^k)_{k \in \mathbb{N}} \subset \mathcal{M}$ be such that $\lim_{k \to \infty} p^k = p$, with

$$(4.6) \quad \lambda_i h_i(p^k) > 0 \quad \forall i \in \{1, \ldots, s\}, \; \lambda_i \neq 0, \quad and \quad \mu_j g_j(p^k) > 0 \quad \forall j \in \mathcal{A}(p), \; \mu_j > 0.$$

By the continuity of $\operatorname{grad} h$ and $\operatorname{grad} g$ and item (i) of Definiton 4.5, we have

$$(4.7) \qquad \lim_{k \to \infty} \left(\sum_{i=1}^{s} \lambda_i \operatorname{grad} h_i(p^k) + \sum_{j \in \mathcal{A}(p)} \mu_j \operatorname{grad} g_j(p^k) \right) = 0.$$

It follows from (4.7) and from the continuity of $\operatorname{grad} f$, provided that $\lim_{k \to \infty} p^k = p$, that

$$(4.8) \qquad \lim_{k \to \infty} \frac{1}{k} \left(\operatorname{grad} f(p^k) + \sum_{i=1}^{s} k \lambda_i \operatorname{grad} h_i(p^k) + \sum_{j \in \mathcal{A}(p)} k \mu_j \operatorname{grad} g_j(p^k) \right) = 0.$$

Taking into account that any positive multiple of (λ, μ) also satisfies the three items of Definition 4.5, we can suppose without loss of generality that $\|(\lambda, \mu)\|_\infty = 1$. Thus, setting $\lambda^k := k\lambda$, $\mu^k := k\mu$, and $\gamma_k := \|(1, \lambda^k, \mu^k)\|_\infty$, we have $\gamma_k = k$. Hence, (4.8) becomes

$$(4.9) \qquad \lim_{k \to \infty} \frac{1}{\gamma_k} \left(\operatorname{grad} f(p^k) + \sum_{i=1}^{s} \lambda_i^k \operatorname{grad} h_i(p^k) + \sum_{j \in \mathcal{A}(p)} \mu_j^k \operatorname{grad} g_j(p^k) \right) = 0.$$

We conclude that p is a scaled-PAKKT point and the proof is complete. $\qquad \square$

5. Examples of CRO. While unconstrained Riemannian optimization has achieved a certain degree of establishment, the available literature concerning CRO problems, specifically formulated as problem (2.1), remains relatively limited. This section aims to present some problems, as described by (2.1), from the existing body of literature on Riemannian optimization which fits our framework.

Let us begin with an example of an optimization problem constrained on a sphere intersected with the positive orthant.

Example 7. Let $V \in \mathbb{R}^{n \times n}$ be a matrix and $\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n : x^T x = 1\}$ the unit sphere, which is an embedded Riemannian submanifold of $\mathbb{R}^n$. Consider the following constrained optimization problem on $\mathbb{S}^{n-1}$:

$$(5.1) \qquad \begin{aligned} &\underset{x \in \mathbb{S}^{n-1}}{\text{Minimize}} \, x^T V x, \\ &\text{subject to } x \geq 0, \end{aligned}$$

where $x \geq 0$ represents the fact that x lies within the nonnegative orthant. It is important to emphasize that problem (5.1) involves the minimization of a quadratic form, which may be nonconvex on the feasible set. The mathematical model presented in (5.1) encompasses various practical challenges, notably addressing nonnegative

principal component analysis (PCA) as exemplified in [50]. Furthermore, it is applicable in tackling problems within the domain of copositive programming, a renowned NP-hard problem, as elaborated in [36]. To delve deeper into these concepts, we shall provide an overview based on the discussions in [50, 36]. Classical PCA seeks to estimate the direction of maximum variability, denoted as v_0, within a high-dimensional dataset $\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_n \subset \mathbb{R}^d$. One approach used to deal with this problem is the so-called spiked model, which assumes that the dataset has a special structure and associates with it a symmetric matrix V. Under this framework, classical PCA involves the computation of $v_1 := v_1(V)$, the eigenvector that corresponds to the largest eigenvalue of $-V$. This eigenvector, in a certain sense, serves as an approximation to v_0. Thus, classical PCA is effectively represented as problem (5.1), removing the positivity constraint. However, it is well known that in the high-dimensional regime where $n = O(d)$, this approach breaks down dramatically due to the fact that v_1 becomes asymptotically less indicative of v_0. A potential solution to overcome this issue is to incorporate structural constraints, such as enforcing sparsity within the solution space. In the work [50], the authors leverage the prior knowledge of nonnegativity in the data entries. Specifically, they propose the solution of the nonnegative PCA, as outlined in (5.1), to approximate v_0 by computing v_1. Now, let us establish the connection between problem (5.1) and copositive programming, a field with several practical applications. To begin, let us recall the concept of copositivity. A real symmetric matrix $V \in \mathbb{R}^{n \times n}$ is copositive if its corresponding quadratic form $\mathbb{R}^n \ni v \mapsto q_V(x) := x^T V x$ takes only nonnegative values on the nonnegative orthant. Copositive programming fundamentally involves the determination of whether a matrix V is copositive, see the review paper [36] for more details. Taking into account the homogeneity of the quadratic form q_V, the determination of whether a matrix X is copositive or not can be achieved by solving problem (5.1) and verifying whether its minimum value is nonnegative.

To proceed with the presentation of the following two examples, we need to introduce some notations. Let $I_k \in \mathbb{R}^{k \times k}$ be the identity matrix. For a matrix $X = (X_{ij}) \in \mathbb{R}^{n \times k}$, the notation $X \geq 0$ means entrywise nonnegativity. Considering $k \leq n$, $\mathbb{R}^{n \times k}$ is equipped with the standard inner product $\langle X, Y \rangle := \operatorname{tr}(Y^T X)$ and the corresponding Frobenius norm $\| \cdot \|_F$, where tr stands for the matrix trace function, the set

$$\mathbb{S}^{n,k} := \{ X \in \mathbb{R}^{n \times k} : X^T X = I_k \},$$

denotes the Stiefel manifold, which is an embedded Riemannian submanifold of $\mathbb{R}^{n \times k}$. In particular, it should be noted that $\mathbb{S}^{n,1} = \mathbb{S}^{n-1}$. An important subset of $\mathbb{S}^{n,k}$ is its nonnegative part, which finds applications in various contexts. It is defined as follows:

$$(5.2) \qquad \mathbb{S}_+^{n,k} := \{ X \in \mathbb{R}^{n \times k} : X^T X = I_k, X \geq 0 \}.$$

For a better comprehension of the set $\mathbb{S}_+^{n,n}$ and its applications, it is often beneficial to rewrite it as follows:

$$(5.3) \qquad \mathbb{S}_+^{n,k} := \left\{ X \in \mathbb{R}^{n \times n} : X E_n = E_n, X^T E_n = E_n, X_{ij} \in \{0, 1\} \forall i, j \right\},$$

were $E_n \in \mathbb{R}^{n \times 1}$ are vectors with all entries as ones. See [61]. Consequently, if $X \in \mathbb{S}_+^{n,n}$, then it implies that each row and column of X contains exactly one positive entry, which is equal to 1. In particular, the set $\mathbb{S}_+^{n,n}$ represents a collection of permutation matrices of order n. For a comprehensive study about $\mathbb{S}_+^{n,k}$, see, for example [24, section 7.3]. In the next example, we examine an optimization

problem constrained to $\mathbb{S}_+^{n,k}$, which encompasses various significant specific cases, notably including the problem highlighted in Example 7. In our presentation, we follow the approach outlined in [39].

Example 8. Consider the following constrained optimization problem on the Stiefel manifold $\mathbb{S}^{n,k}$:

$$(5.4) \qquad \begin{aligned} &\underset{X \in \mathbb{S}^{n,k}}{\text{Minimize}} f(X), \\ &\text{subject to } X \geq 0, \end{aligned}$$

where $f : \mathbb{R}^{n \times k} \to \mathbb{R}$ is continuously differentiable function. It is important to point out that, despite its seemingly simple statement, problem (5.4) has many challenges, primarily because the constraint set is only geodesically convex for $k = 1$. As illustrated by [38, 39], and the references therein, the mathematical model presented in (5.4) addresses a wide range of applications and interests. For instance, for $k = n$, when considering two sets of correlated objects associated with the rows and columns of a square matrix, each permutation matrix represents an assignment from objects in one group to objects in the other group. Hence, problem (5.4) is commonly referred to as the nonlinear assignment problem. It aims to find the optimal assignment between two groups, referred to as "facilities" and "locations," with the objective of minimizing the nonlinear cost. One notable variant is the quadratic assignment problem, a fundamental combinatorial optimization problem within the field of operations research in mathematics, categorized as a facilities location problem. By using the equivalence between (5.2) and (5.3), this problem can be formally defined using the notation $\mathbb{S}^{n,n}$ as follows:

$$(5.5) \qquad \begin{aligned} &\underset{X \in \mathbb{S}^{n,n}}{\text{Minimize}} \operatorname{tr}(A^T X B X^T), \\ &\text{subject to } X \geq 0, \end{aligned}$$

where $A, B \in \mathbb{R}^{n \times n}$. Another particular instance of problem (5.4) is the so-called trace minimization with nonnegative orthogonality constraints, formulated as

$$(5.6) \qquad \begin{aligned} &\underset{X \in \mathbb{S}^{n,k}}{\text{Minimize}} \operatorname{tr}(X^T M X), \\ &\text{subject to } X \geq 0, \end{aligned}$$

where $M \in \mathbb{R}^{n \times n}$ is a symmetric matrix. Note that for $k = 1$, problem (5.6) reduces to problem (5.1). Finally, it is worth noting that the task of projecting a matrix $M \in \mathbb{R}^{n \times k}$ onto $\mathbb{S}_+^{n,k}$ represents a specific case of problem (5.4). Using the equality $\|X - M\|_F^2 = k - 2\operatorname{tr}(X^T M) + \|M\|_F^2$ for all $X \in \mathbb{S}^{n,k}$, this can be reformulated as a constrained optimization problem on the nonnegative section of $\mathbb{S}^{n,k}$ as follows:

$$(5.7) \qquad \begin{aligned} &\underset{X \in \mathbb{S}^{n,k}}{\text{Minimize}} -2\operatorname{tr}(X^T M), \\ &\text{subject to } X \geq 0. \end{aligned}$$

For additional specific instances of problem (5.4), see [38, 39].

In the following example, we present a clustering problem in machine learning, following the approach outlined in [27]; see also [24].

Example 9. Consider a set of data points $\mathcal{C} := \{x_1, \ldots, x_\ell\} \subset \mathbb{R}^{n \times 1}$. The k-means clustering problem involves partitioning these data point set $\mathcal{C}$ into k subsets, denoted as $\mathcal{C}_1, \ldots, \mathcal{C}_k$, with the objective of minimizing the sum of squared distances between each data point and the centroid of the corresponding subset to which it belongs.

Let $\mathbb{J}_j := \{i \in \{1, \ldots, \ell\} : x_i \in \mathcal{C}_j\}$, and let $|\mathbb{J}_j|$ denote the cardinality of the set $\mathbb{J}_j$, which also represents the number of data points in set $\mathcal{C}_j$. Additionally, $c_j := \frac{1}{|\mathbb{J}_j|} \sum_{i \in \mathbb{J}_j} x_i$ stands for the centroid of the set $\mathcal{C}_i$. Formally, the k-means clustering problem can be stated as follows:

$$
\text{(5.8)} \qquad \begin{aligned} &\text{Minimize} \sum_{j=1}^{k} \sum_{i \in \mathbb{J}_j} \|x_i - c_j\|_F^2, \\ &\text{subject to } \mathbb{J}_1 \cup \cdots \cup \mathbb{J}_k = \{1, \ldots, \ell\}. \end{aligned}
$$

According to calculations outlined in [27], problem (5.8) can be reformulated equivalently as an optimization problem with nonnegative orthogonality constraints, expressed as follows:

$$
\text{(5.9)} \qquad \begin{aligned} &\underset{X \in \mathbb{S}^{n,k}}{\text{Minimize}} \, \mathrm{tr}(DXX^T), \\ &\text{subject to } XX^T E_n = E_n, \, X \geq 0, \end{aligned}
$$

where $D := (D_{ij}) \in \mathbb{R}^{n \times n}$ with $D_{ij} = \|x_i - x_j\|_F^2$ for all $i, j = 1, \ldots, n$.

Concluding this session, we turn our attention to specific computational aspects related to the implementation of augmented Lagrangian methods. While the primary focus of this paper revolves around deriving new theoretical properties of this method, it is important to highlight the historical progression of implementing augmented Lagrangian methods over the years. With a well-documented history of successful applications in Euclidean space, the augmented Lagrangian method has demonstrated its efficacy in addressing various real-world challenges, exhibiting robust computational performance, as indicated in prior studies [2, 22]. Considering its strong performance within the Euclidean space, there is a plausible indication of its potential applicability within the Riemannian context. This notion has been supported by other researchers who have explored the augmented Lagrangian method within the Riemannian framework, yielding promising results, as presented in recent research [59]. For a specific instance of the implementation of Algorithm 2.1 without a safeguard and its application in resolving Examples 7 and 9, along with the minimum balanced cut problem, the seminal work by [45] serves as the primary reference, marking the inception of the study of this algorithm within the Riemannian domain.

6. Conclusions. In this paper we presented a detailed global convergence analysis of a safeguarded augmented Lagrangian method defined on a complete Riemannian manifold. In order to do this, we presented several weak CQs that can be used to obtain stationarity of all limit points of a primal sequence generated by the algorithm, despite the fact that the dual sequence may be unbounded. In doing so, we described several properties of these conditions well known in the Euclidean setting, which should foster further developments in the Riemannian setting. By means of a stronger sequential optimality condition, we were able to present a weak CQ which guarantees boundedness of the dual sequence, even when the true set of Lagrange multipliers is unbounded. In presenting our conditions, we provided illustrative examples to prove that our conditions are strictly weaker than previously known ones in *any* complete Riemannian manifold with dimension $n \geq 2$.

Note that when defining the sequential optimality conditions, we chose to present the simplest complementarity measure, namely, item (ii) of Theorem 2.1 and item (iii) of Definition 4.1, while in [59], they considered a slightly stronger complementarity measure known as approximate gradient projection. See the recent discussion about

several different complementarity measures in [4] in the context of Euclidean conic optimization. We foresee significant progress in this topic in the nearby future; in particular, several stronger first- and second-order global convergence results of augmented Lagrangian methods and other algorithms should be expected to be extended to the Riemannian setting.

Finally, in the particular case where the manifold $\mathcal{M}$ can be embedded in an Euclidean space, one can treat $x \in \mathcal{M}$ as a subproblem/lower level constraint as described in [21]. It is clear that one should exploit the Riemannian structure in order to solve the subproblems more efficiently, while it is also clear that an intrinsic formulation of the theory is well justified [19]; however, in this context, it is not clear whether the pure Euclidean theory differs from the one formulated in the Riemannian setting. This topic will be the subject of a forthcoming paper.

REFERENCES

[1] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ, 2008.

[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 (2007), pp. 1286–1309.

[3] 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 (2019), pp. 743–766.

[4] R. Andreani, W. Gómez, G. Haeser, L. Mito, and A. Ramos, *On optimality conditions for nonlinear conic programming*, Math. Oper. Res., 3 (2022), pp. 2160–2185.

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

[6] R. Andreani, G. Haeser, L. Mito, and H. Ramírez, *Sequential constant rank constraint qualifications for nonlinear semidefinite programming with applications*, Set-Valued Var. Anal., 31 (2023), 3.

[7] R. Andreani, G. Haeser, L. Mito, H. Ramírez, and T. Silveira, *Global convergence of algorithms under constant rank conditions for nonlinear second-order cone programming*, J. Optim. Theory Appl., 195 (2022), pp. 42–78.

[8] R. Andreani, G. Haeser, L. M. Mito, and H. Ramírez C, *A minimal face constant rank constraint qualification for reducible conic programming*, Optimization Online, (2023).

[9] R. Andreani, G. Haeser, L. M. Mito, and H. Ramírez C, *Weak notions of nondegeneracy in nonlinear semidefinite programming*, Math. Program., 205 (2024), pp. 1–32.

[10] R. Andreani, G. Haeser, L. M. Mito, H. Ramírez C, and T. P. Silveira, *First- and second-order optimality conditions for second-order cone and semidefinite programming under a constant rank condition*, Math. Program., 202 (2023), pp. 473–513.

[11] 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 (2022), pp. 121–146.

[12] 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 (2012), pp. 255–273.

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

[14] R. Andreani, G. Haeser, L. D. Secchin, and P. J. S. Silva, *New sequential optimality conditions for mathematical problems with complementarity constraints and algorithmic consequences*, SIAM J. Optim., 4 (2019), pp. 3201–3230.

[15] R. Andreani, G. Haeser, and D. Viana, *Optimality conditions and global convergence for nonlinear semidefinite programming*, Math. Program., 1 (2020), pp. 203–235.

[16] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva, *A cone-continuity constraint qualification and algorithmic consequences*, SIAM J. Optim., 26 (2016), pp. 96–110.

[17] 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 (2005), pp. 473–485.

[18] R. ANDREANI, M. L. SCHUVERDT, AND L. D. SECCHIN, *On enhanced KKT optimality conditions for smooth nonlinear optimization*, SIAM J. Optim., 34 (2024), pp. 1515–1539.

[19] R. BERGMANN AND R. HERZOG, *Intrinsic formulation of KKT conditions and constraint qualifications on smooth manifolds*, SIAM J. Optim., 29 (2019), pp. 2423–2444.

[20] D. P. BERTSEKAS, *Nonlinear Programming*, 3rd ed., Athena Scientific Optimization and Computation Series, Athena Scientific, Belmont, MA, 2016.

[21] E. G. BIRGIN, G. HAESER, AND A. RAMOS, *Augmented Lagrangians with constrained subproblems and convergence to second-order stationary points*, Comput. Optim. Appl., 1 (2018), pp. 51–75.

[22] E. G. BIRGIN AND J. M. MARTÍNEZ, *Practical Augmented Lagrangian Methods for Constrained Optimization*, Fundam. Algorithms 10, SIAM, Philadelphia, PA, 2014.

[23] E. BÖRGENS, C. KANZOW, P. MEHLITZ, AND G. WACHSMUTH, *New constraint qualifications for optimization problems in Banach spaces based on asymptotic KKT conditions*, SIAM J. Control Optim., 30 (2020), pp. 2956–2982.

[24] N. BOUMAL, *An Introduction to Optimization on Smooth Manifolds*, Cambridge University Press, Cambridge, 2023.

[25] L. F. BUENO, G. HAESER, F. LARA, AND F. N. ROJAS, *Augmented Lagrangian method for the solution of quasi-equilibrium problems*, Comput. Optim. Appl., 76 (2020), pp. 737–766.

[26] L. F. BUENO, G. HAESER, AND F. N. ROJAS, *Optimality conditions and constraint qualifications for generalized Nash equilibrium problems and their practical implications*, SIAM J. Optim., 1 (2019), pp. 31–54.

[27] T. CARSON, D. G. MIXON, S. VILLAR, AND R. WARD, *Manifold optimization for k-means clustering*, in Proceedings of the 2017 International Conference on Sampling Theory and Applications (SampTA), IEEE, 2017, pp. 73–77.

[28] K. DENG AND Z. PENG, *A manifold inexact augmented Lagrangian method for nonsmooth optimization on Riemannian submanifolds in Euclidean space*, IMA J. Numer. Anal., 43 (2023), pp. 1653–1684.

[29] M. P. A. DO CARMO, *Riemannian Geometry*, Mathematics: Theory & Applications, transl. F. Flaherty, Birkhäuser Boston, Boston, MA, 1992.

[30] A. EDELMAN, T. A. ARIAS, AND S. T. SMITH, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 303–353.

[31] H. GFRERER AND B. S. MORDUKHOVICH, *Complete characterizations of tilt stability in nonlinear programming under weakest qualification conditions*, SIAM J. Optim., 25 (2015), pp. 2081–2119.

[32] G. GIORGI, B. JIMÉNEZ, AND V. NOVO, *Approximate Karush-Kuhn-Tucker condition in multiobjective optimization*, J. Optim. Theory Appl., 171 (2016), pp. 70–89.

[33] L. GUO AND J. YE, *Necessary optimality conditions and exact penalization for non-Lipschitz nonlinear programs*, Math. Program., 168 (2018), pp. 571–598.

[34] G. HAESER, O. HINDER, AND Y. YE, *On the behavior of Lagrange multipliers in convex and non-convex infeasible interior point methods*, Math. Program., 186 (2021), pp. 257–288.

[35] M. R. HESTENES, *Optimization Theory: The Finite Dimensional Case*, John Wiley & Sons, New York, 1975.

[36] J.-B. HIRIART-URRUTY AND A. SEEGER, *A variational approach to copositive matrices*, SIAM Rev., 52 (2010), pp. 593–629.

[37] R. JANIN, *Directional derivative of the marginal function in nonlinear programming*, Math. Program. Stud., 21 (1984), pp. 110–126.

[38] B. JIANG, Y.-F. LIU, AND Z. WEN, L_p-*norm regularization algorithms for optimization over permutation matrices*, SIAM J. Optim., 26 (2016), pp. 2284–2313.

[39] B. JIANG, X. MENG, Z. WEN, AND X. CHEN, *An exact penalty approach for optimization with nonnegative orthogonality constraints*, Math. Program., 198 (2023), pp. 855–897.

[40] C. KANZOW, A. B. RAHARJA, AND A. SCHWARTZ, *Sequential optimality conditions for cardinality-constrained optimization problems with applications*, Comput. Optim. Appl., 80 (2021), pp. 185–211.

[41] A. KRUGER, L. MINCHENKO, AND J. OUTRATA, *On relaxing the Mangasarian-Fromovitz constraint qualification*, Positivity, 18 (2013), pp. 171–189.

[42] Z. LAI AND A. YOSHISE, *Riemannian interior point methods for constrained optimization on manifolds*, J. Optim. Theory Appl., 201 (2024), pp. 433–469.

[43] S. LANG, *Differential and Riemannian Manifolds*, 3rd ed., Grad. Texts in Math. 160, Springer-Verlag, New York, 1995.

[44] Y. Liang and J. Ye, *Optimality conditions and exact penalty for mathematical programs with switching constraints*, J. Optim. Theory Appl., 190 (2021), pp. 1–31.

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

[46] S. Lu, *Implications of the constant rank constraint qualification*, Math. Program., 126 (2011), pp. 365–392.

[47] P. Mehlitz and L. Minchenko, *R-regularity of set-valued mappings under the relaxed constant positive linear dependence constraint qualification with applications to parametric and bilevel optimization*, Set-Valued Var. Anal., 30 (2022), pp. 179–205.

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

[49] L. Minchenko and A. Tarakanov, *On error bounds for quasinormal programs*, J. Optim. Theory Appl., 148 (2011), pp. 571–579.

[50] A. Montanari and E. Richard, *Non-negative principal component analysis: Message passing algorithms and sharp asymptotics*, IEEE Trans. Inform. Theory, 62 (2016), pp. 1458–1484.

[51] M. Obara, T. Okuno, and A. Takeda, *Sequential quadratic optimization for nonlinear optimization problems on Riemannian manifolds*, SIAM J. Optim., 32 (2022), pp. 822–853.

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

[53] T. Sakai, *Riemannian Geometry*, Translations of Mathematical Monographs 160, American Mathematical Society, Providence, RI, 1996.

[54] L. W. Tu, *An Introduction to Manifolds*, 2nd ed., Universitext, Springer-Verlag New York, 2011.

[55] C. Udrişte, *Kuhn-Tucker theorem on Riemannian manifolds*, in Topics in Differential Geometry, Vol. II, Colloq. Math. Soc. János Bolyai 46, North-Holland, Amsterdam, 1988, pp. 1247–1259.

[56] G. Wachsmuth, *On LICQ and the uniqueness of Lagrange multipliers*, Oper. Res. Lett., 41 (2013), pp. 78–80.

[57] A. Wächter and L. Biegler, *On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming*, Math. Program., 1 (2006), pp. 25–57.

[58] M. Xu and J. Ye, *Relaxed constant positive linear dependence constraint qualification and its application to bilevel programs*, J. Global Optim., 78 (2020), pp. 181–205.

[59] 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 (2022), pp. 397–421.

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

[61] M. M. Zavlanos and G. J. Pappas, *A dynamical systems approach to weighted graph matching*, Automatica, 44 (2008), pp. 2817–2824.