SIAM J. OPTIM.
Vol. 36, No. 1, pp. 466–501
© 2026 Society for Industrial and Applied Mathematics

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

GLOBAL CONVERGENCE OF AN AUGMENTED LAGRANGIAN METHOD FOR NONLINEAR PROGRAMMING VIA RIEMANNIAN OPTIMIZATION*

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

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

Key words. safeguarded augmented Lagrangian method, constrained nonlinear programming, constraint qualifications, embedded submanifold

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

DOI. 10.1137/24M1692319

1. Introduction.

This paper advances the understanding of findings concerning constraint qualifications and convergence properties inherent in an augmented Lagrangian method designed for Riemannian manifolds, as initially outlined in [5]. It aims to demonstrate how the theoretical framework based on Riemannian concepts can introduce innovative perspectives and viable alternative solutions to problems traditionally addressed through Euclidean theory. Additionally, this study highlights the capacity of modern Riemannian geometry concepts to enrich conventional Euclidean theory, thereby refining theoretical paradigms within Euclidean space. To achieve this objective, we introduce novel constraint qualifications and explore the applicability of Riemannian augmented Lagrangian methods to a specific category of constrained nonlinear programming problems characterized by both equality and inequality constraints, with the equality constraints further categorized into two distinct types. The constrained optimization problem under consideration is formally defined as follows:

*Received by the editors September 10, 2024; accepted for publication (in revised form) October 27, 2025; published electronically March 19, 2026.
https://doi.org/10.1137/24M1692319

†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, lfprudente@ufg.br).
¶Department of Applied Mathematics, University of São Paulo, São Paulo-SP, Brazil (ghaeser@ime.usp.br).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

$$\underset{q\in\mathbb{R}^n}{\text{Minimize}}\, f(q), \qquad \text{subject to} \quad h(q)=0,\ H(q)=0,\ G(q)\leq 0, \tag{1.1}$$

where the functions $f:\mathbb{R}^n\to\mathbb{R}$, $h:=(h_1,\dots,h_t)^\top:\mathbb{R}^n\to\mathbb{R}^t$, $H:=(H_1,\dots,H_s)^\top:\mathbb{R}^n\to\mathbb{R}^s$, and $G:=(G_1,\dots,G_m)^\top:\mathbb{R}^n\to\mathbb{R}^m$ are continuously differentiable. A standard approach to solving problem (1.1) is through the *augmented Lagrangian algorithm,* which involves the iterative unconstrained minimization of the *standard Powell–Hestenes–Rockafellar augmented Lagrangian function* given by

$$L_\rho(q,\eta,\lambda,\mu):=f(q)+\frac{\rho}{2}\Big(\Big\|h(q)+\frac{\eta}{\rho}\Big\|_2^2+\Big\|H(q)+\frac{\lambda}{\rho}\Big\|_2^2+\Big\|\Big[G(q)+\frac{\mu}{\rho}\Big]_+\Big\|_2^2\Big), \tag{1.2}$$

where $\rho>0$ is a fixed penalty parameter, and safeguarded Lagrange multipliers $\eta:=(\eta_1,\dots,\eta_t)\in\mathbb{R}^t$, $\lambda:=(\lambda_1,\dots,\lambda_s)\in\mathbb{R}^s$, and $\mu:=(\mu_1,\dots,\mu_m)\in\mathbb{R}^m_+$ are estimated in each (outer) iteration. Here $[u]_+$ stands for the projection of $u\in\mathbb{R}^m$ onto the nonnegative orthant $\mathbb{R}^m_+$. An alternative approach to addressing constrained optimization problems in the format (1.1), previously utilized in [4, 19], involves considering the so-called lower-level constraints:

$$\mathbb{M}:=\{q\in\mathbb{R}^n \mid h(q)=0\}. \tag{1.3}$$

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

$$\mathbb{L}_\rho(q,\lambda,\mu):=f(q)+\frac{\rho}{2}\Big(\Big\|H(q)+\frac{\lambda}{\rho}\Big\|_2^2+\Big\|\Big[G(q)+\frac{\mu}{\rho}\Big]_+\Big\|_2^2\Big), \tag{1.4}$$

subject to the lower-level set $\mathbb{M}$. The idea behind this division arises from the strategic advantage that augmented Lagrangian methods offer in solving nonlinear programming problems. By partitioning the equality constraints, a level of flexibility is introduced, allowing for the prioritization of constraints based on their relevance to the current problem or the ease with which they can be managed. Consequently, this approach within the augmented Lagrangian framework enables the incorporation of certain constraints directly into the objective functional via penalty terms, while other constraints are inherently enforced by restricting the optimization domain to a lower-level set, ensuring their satisfaction without requiring penalization. As a result, subproblems are formulated as minimizing $\mathbb{L}_\rho(\cdot,\eta,\mu)$ subject to $\mathbb{M}$. For instance, if the goal is to maintain feasibility for a set $\mathbb{M}$, these subproblems can be addressed using methods that keep the (inner) iterates in $\mathbb{M}$. In this manner, the sequence generated by these constrained augmented Lagrangian methods remains feasible with respect to $\mathbb{M}$. Additionally, a notable aspect of this approach applies to scenarios where the objective function and/or constraints are defined solely at points belonging to $\mathbb{M}$, rendering the equality constraint $h(q)=0$ ineligible for penalization.

Understanding optimality conditions and constraint qualifications is crucial in the study of nonlinear programing problems. The *Karush/Kuhn–Tucker (KKT) conditions* play a pivotal role in identifying optimal solutions, while *constraint qualifications* ensure that these solutions satisfy the KKT conditions. Over time, modern nonlinear programming theory has witnessed the evolution of KKT conditions and the emergence of new constraint qualifications. This evolution has significantly broadened the theoretical framework of nonlinear optimization, allowing the application of augmented Lagrangian methods across a wide range of problem classes. Notably, constraint qualifications such as the *constant rank constraint qualification* (CRCQ) [37],

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

constant positive linear dependence condition (CPLD) [47], *relaxed CRCQ* (RCRCQ) [45], *relaxed CPLD* (RCPLD) [10], *constant rank of the subspace component* (CRSC) [9] and *quasinormality constraint qualification* (QN) [35], have been introduced to enhance the understanding and application of optimization techniques. Moreover, the introduction of sequential optimality conditions, such as the *approximate KKT* (AKKT) [7] and *positive AKKT* (PAKKT) [6] conditions, has provided additional flexibility by relaxing the KKT conditions without assuming a constraint qualification. These developments represent important advancements in the field and are essential for advancing the state-of-the-art in nonlinear programming research. For example, within the framework of safeguarded augmented Lagrangian methods, their strength lies in their ability to generate PAKKT sequences for constrained nonlinear programming problems. Under any of the aforementioned constraint qualifications, this ensures that all limit points of such sequences adhere to the KKT conditions, a topic extensively explored in the literature (see, for instance, [6, 8, 9]). We use the adjective *strict* to distinguish the constraint qualifications with the aforementioned sequential property; see [11].

To address nonlinear optimization problems in the format (1.1), we introduce new strict constraint qualifications termed *lower strict constraint qualifications* (lower-SCQs) to take into account the lower-level approach of considering augmented Lagrangian subproblems constrained to the lower-level set $\mathbb{M}$. These constraint qualifications serve as less restrictive counterparts to CRCQ, CPLD, RCRCQ, RCPLD, CRSC, and QN. In this new scenario, it is no longer guaranteed that the limit points of the sequence generated by classic augmented Lagrangian methods satisfy the KKT conditions. Therefore, by considering the equality constraints (1.3) as a Riemannian manifold, we employ tools from Riemannian geometry to establish a connection between the lower-SCQs and their Riemannian counterparts recently introduced in [5], referred to as *Riemannian strict constraint qualifications* (Riemannian SCQs). Furthermore, by introducing the concept of *lower-AKKT* and *lower-PAKKT* for problem (1.1), which serve as counterparts to AKKT and PAKKT, respectively, we show that the *Riemannian adaptation of the classic safeguarded augmented Lagrangian algorithm,* an intrinsic algorithm presented in [55], is able to produce lower-PAKKT sequences that are feasible for $\mathbb{M}$. Moreover, under any lower-SCQ we show that all limit points of this sequence satisfy the KKT conditions for problem (1.1). Additionally, as we establish a link between these lower-SCQs and the Riemannian-SCQs, we highlight the effectiveness of the theory within Riemannian manifolds. This effectiveness offers valuable support for the convergence analysis of algorithms in nonlinear programming, especially when compared to those formulated in Euclidean spaces. In particular, the equivalence between lower-SCQs and Riemannian SCQs not only provides deeper theoretical insights but also creates new possibilities for developing algorithms in Euclidean settings that benefit from the weaker assumptions and structural features of the Riemannian framework, thereby expanding their applicability. Our numerical experiments further illustrate these practical advantages. Moreover, our contributions broaden both the theoretical foundations and practical scope of augmented Lagrangian methods by offering effective strategies for handling lower-level constraints. As previously noted, the augmented Lagrangian method incorporating general lower-level constraints was introduced in [4], where the software ALGENCAN was designed to handle box constraints. Our work can be viewed as a natural continuation of these ideas, establishing the theoretical basis necessary to support modern implementations of augmented Lagrangian methods in more general constrained optimization problems. This underscores that there are

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

various subtle aspects concerning constraint qualifications in Riemannian settings that would be overlooked if the problem were solely addressed with the existing Euclidean theory. In this sense, as mentioned earlier, this paper also serves as a complement to aid in understanding the range of applications of the theory presented in [5]. It is worth noting that many sets of the form (1.3) define a Riemannian manifold and naturally arise in various applications. For further examples, see [2, 23]. Finally, we note that augmented Lagrangian methods on Riemannian manifolds have attracted growing interest in recent years (see [29, 52, 53, 57]), and a rigorous treatment of strict constraint qualifications is likely to play a central role in this development.

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

1.1. Notations, terminology, and basics results. For a finite subset $\mathcal{K} \subset \mathbb{N} = \{1, 2, \dots\}$, we denote its cardinality by $|\mathcal{K}|$. We use $\subset$ to denote inclusion and this does not preclude the sets from being equal. The set of all $m \times n$ matrices with real entries is denoted by $\mathbb{R}^{m\times n}$. For $M \in \mathbb{R}^{m\times n}$, the matrix $M^\top \in \mathbb{R}^{n\times m}$ is the *transpose* of M. Let $\mathbb{R}^n \equiv \mathbb{R}^{n\times 1}$ be the m-dimensional Euclidean space with the norm denoted by $\|\cdot\|_2$. We denote the infinity norm in $\mathbb{R}^m$ by $\|\cdot\|_\infty$. The *open* and *closed balls* of radius $r > 0$ in $\mathbb{R}^m$, centered at p, are, respectively, defined by $B_r(p) := \{q \in \mathbb{R}^m | \|p - q\|_2 < r\}$ and $B_r[p] := \{q \in \mathbb{R}^m \mid \|p - q\|_2 \le r\}$. For all $p, q \in \mathbb{R}^m$, $\min\{p, q\} \in \mathbb{R}^m$ is the componentwise minimum of p and q. We denote by $[q]_+$ the Euclidean projection of q onto the nonnegative orthant $\mathbb{R}^m_+$. The subspace spanned by a set $\mathcal{C} \subset R^m$ is denoted by $\mathrm{Span}(\mathcal{C})$. Given a differentiable function $\varphi := (\varphi_1, \dots, \varphi_m)^\top : \mathbb{R}^n \to \mathbb{R}^m$, we denote by $\varphi'(q) \in \mathbb{R}^{n\times m}$ the *transpose of the Jacobian matrix of* φ *at* q, that is, $\varphi'(q) = [\,\varphi_1'(q)\ \varphi_2'(q)\ \cdots\ \varphi_m'(q)\,]$, where each column $\varphi_j'(q) \in \mathbb{R}^n$ is the Euclidean gradient of the scalar-valued function φ_j. In the special case $m = 1$, the vector $\varphi'(q) \in \mathbb{R}^n$ itself denotes the *Euclidean gradient* of the scalar function φ. Throughout this paper, whenever we write a set of vectors $\{u_1, u_2, \dots, u_m\}$, it should be understood as a multiset, meaning that repetitions of vectors are allowed. For a given subspace $V \subset \mathbb{R}^m$, its *orthogonal subspace* is defined by $V^\perp := \{z \in \mathbb{R}^m \mid v^\top z = 0\ \ \forall v \in V\}$ and the *Euclidean projection operator* onto $V^\perp$ is denoted by $\mathrm{Proj}_{V^\perp}$. Let $C \subset \mathbb{R}^n$ be a given cone. The *polar cone* of C, denoted by C°, is defined as $C^\circ := \{v \in \mathbb{R}^n \mid v^\top u \le 0\ \forall u \in C\}$.

DEFINITION 1.1. *Let $V = \{v_1, \dots, v_s\}$ and $W = \{w_1, \dots, w_m\}$ be two finite multisets on $\mathbb{R}^n$. The multiset $V \cup W$ is said to be positive-linearly dependent with respect to W if there exist $\alpha = (\alpha_1, \dots, \alpha_s) \in \mathbb{R}^s$ and $\beta = (\beta_1, \dots, \beta_m) \in \mathbb{R}^m_+$ such that $(\alpha, \beta) \ne 0$ and $\sum_{i=1}^s \alpha_i v_i + \sum_{j=1}^m \beta_j w_j = 0$. Otherwise, $V \cup W$ is said to be positive-linearly independent with respect to W.*

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

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

LEMMA 1.1. *Let* $\mathcal{C}_1 = \{u_1, \ldots, u_t\}$, $\mathcal{C}_2 = \{v_1, \ldots, v_s\}$, *and* $\mathcal{C}_3 = \{w_1, \ldots, w_m\}$ *be finite multisets of vectors in* $\mathbb{R}^n$. *Suppose that* $\mathcal{C}_1$ *is linearly independent, and let* $V := \mathrm{Span}(\mathcal{C}_1)$ *and* $V^\perp$ *its orthogonal complement. Define* $\mathrm{Proj}_{V^\perp} \mathcal{C}_2 := \{\mathrm{Proj}_{V^\perp} v_i \mid i = 1, \ldots, s\}$ *and* $\mathrm{Proj}_{V^\perp} \mathcal{C}_2 := \{\mathrm{Proj}_{V^\perp} w_j \mid j = 1, \ldots, m\}$. *Then, the following statements are equivalent:*

(i) *The multiset* $(\mathcal{C}_1 \cup \mathcal{C}_2) \cup \mathcal{C}_3$ *is linearly independent (respectively, positive-linearly independent with respect to* $\mathcal{C}_3$*).*

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

LEMMA 1.2. *Let* $\mathcal{C}_1 := \{v_\ell \in \mathbb{R}^n \mid \ell = 1, \ldots, t\}$, $\mathcal{C}_2 := \{w_j \in \mathbb{R}^n \mid j = 1, \ldots, m\}$, $\mathcal{K} \subset \{1, \ldots, m\}$, *and* $\mathcal{C}_{\mathcal{K}} := \{w_j \in \mathbb{R}^n \mid j \in \mathcal{K}\}$. *Let* $V := Span(\mathcal{C}_1)$ *and* $V^\perp$ *be its orthogonal subspace. Define* $\mathrm{Proj}_{V^\perp} \mathcal{C}_{\mathcal{K}} := \{\mathrm{Proj}_{V^\perp} w_j \mid j \in \mathcal{K}\}$ *and* $\mathrm{Proj}_{V^\perp} \mathcal{C}_2 := \{\mathrm{Proj}_{V^\perp} w_j \mid j = 1, \ldots, m\}$. *Assume that* $\mathcal{C}_1$ *is linearly independent. Then, the following statements are equivalent:*

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

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

Since $h = (h_1, \ldots, h_t)^\top : \mathbb{R}^n \to \mathbb{R}^t$ is continuously differentiable on $\mathbb{R}^n$, by *assuming that the set of gradients* $\{h'_\ell(q) \mid \ell = 1, \ldots, t\}$ *is linearly independent for all* $q \in \mathbb{R}^n$, *we conclude that the set* (1.3) *is an embedded submanifold of* $\mathbb{R}^n$ *of dimension* $n - t$. The *open* and *closed balls* of radius $r > 0$ in $\mathbb{M}$, centered at p, are, respectively, defined by $\mathbb{B}_r(p) := \{q \in \mathbb{M} \mid d(p,q) < r\}$ and $\mathbb{B}_r[p] := \{q \in \mathbb{M} \mid d(p,q) \le r\}$, where $d(\cdot,\cdot)$ *is the Riemannian distance* associated with the induced metric from $\mathbb{R}^n$. The *tangent plane at* $q \in \mathbb{M}$ is given by

$$T_q\mathbb{M} := \left\{v \in \mathbb{R}^n \mid h'(q)^\top v = 0\right\} = \left\{v \in \mathbb{R}^n \mid h'_\ell(q)^\top v = 0,\ \ell = 1, \ldots, t\right\}. \tag{1.5}$$

To simplify the notation we also denote the metric in $T_q\mathbb{M}$ by $\|\cdot\|$. It follows from (1.5) that

$$T_q\mathbb{M} := \mathrm{Ker}\, h'(q)^\top, \qquad T_q\mathbb{M}^\perp = \mathrm{Im}\, h'(q), \qquad \mathbb{R}^n = T_q\mathbb{M} \oplus T_q\mathbb{M}^\perp. \tag{1.6}$$

Therefore, (1.5) and the second equality in (1.6) imply that

$$T_q\mathbb{M}^\perp = \left\{ h'(q)^\top \eta = \sum_{i=1}^{t} \eta_\ell h'_\ell(q) \mid \eta = (\eta_1, \ldots, \eta_t) \in \mathbb{R}^t \right\}. \tag{1.7}$$

For the sake of simplicity, for a given $q \in \mathbb{M}$, let $\mathrm{Proj}_q : \mathbb{R}^n \to T_q\mathbb{M}$ denote the *projection operator,* which is given by

$$\mathrm{Proj}_q\, v = \left(I - h'(q)\big(h'(q)^\top h'(q)\big)^{-1} h'(q)^\top\right) v; \tag{1.8}$$

see, for example, [43, p. 377]. Hence, by using (1.8), the *intrinsic gradient* of a differentiable function $\varphi : \mathbb{M} \to \mathbb{R}$ is given by

$$\mathrm{grad}\, \varphi(q) = \mathrm{Proj}_q\, \varphi'(q). \tag{1.9}$$

For a more detailed discussion on embedded submanifolds, see [23]. We will need the following lemma, and we provide its proof for the sake of completeness.

LEMMA 1.3. *Let* $X_1, \ldots, X_m$ *be continuous vector fields on a Riemannian manifold* $\mathbb{M}$. *Let* $p \in \mathbb{M}$ *and assume that* $\{X_1(p), \ldots, X_m(p)\}$ *are linearly independent on* $T_p\mathbb{M}$. *Then, there exists* $\epsilon > 0$ *such that* $\{X_1(q), \ldots, X_m(q)\}$ *are also linearly independent on* $T_q\mathbb{M}$, *for all* $q \in \mathbb{B}_\epsilon(p)$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Proof. Suppose by contradiction that for every $k \in \mathbb{N}$ there exists a point $p^k \in \mathbb{B}_{1/k}(p)$ at which the vectors $X_1(p^k), \ldots, X_m(p^k)$ are linearly dependent. Then, for each k there exist scalars $\alpha_1^k, \ldots, \alpha_m^k$, not all zero, such that $\sum_{i=1}^m \alpha_i^k X_i(p^k) = 0$. Without loss of generality, we may normalize these scalars so that $\sum_{i=1}^m |\alpha_i^k|^2 = 1$ for each k. Since the unit sphere in $\mathbb{R}^m$ is compact, there exists a subsequence $(\alpha_1^{k_j}, \ldots, \alpha_m^{k_j})$ that converges to some vector $(\alpha_1, \ldots, \alpha_m)$ with $\sum_{i=1}^m |\alpha_i|^2 = 1$. Moreover, as $\lim_{j\to+\infty} p^{k_j} = p$ and $X_1, \ldots, X_m$ are continuous, we have

$$0 = \lim_{j\to\infty} \sum_{i=1}^m \alpha_i^{k_j} X_i(p^{k_j}) = \sum_{i=1}^m \alpha_i X_i(p).$$

This implies that the nonzero vector $(\alpha_1, \ldots, \alpha_m)$ is a nontrivial linear combination of $X_1(p), \ldots, X_m(p)$ that vanishes, contradicting the linear independence of these vectors at p. Therefore, there must exist some $\epsilon > 0$ such that for all $q \in \mathbb{B}_\epsilon(p)$ the set $\{X_1(q), \ldots, X_m(q)\}$ is linearly independent on $T_q\mathbb{M}$. ∎

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

2. Preliminaries.

This section defines essential notations and concepts in Euclidean and Riemannian geometry, reviews the basics for addressing the Euclidean problem (1.1), and uses a submanifold concept to rewrite it as an intrinsic nonlinear optimization problem, yielding new results.

2.1. Nonlinear optimization problems on Euclidean space.

The *feasible set* $\Omega \subset \mathbb{R}^n$ of problem (1.1) and the *set of indices of active inequality constraints* at $p \in \Omega$, denoted by $\mathcal{A}(p)$, are defined, respectively, as follows:

(2.1)
$$\Omega := \{q \in \mathbb{R}^n \mid h(q) = 0,\ H(q) = 0,\ G(q) \le 0\}, \qquad \mathcal{A}(p) := \{j \in \{1, \ldots, m\} \mid G_j(p) = 0\}.$$

It is easy to see that Ω is closed, as h, H, and G are continuous functions. We say that the KKT conditions are satisfied at $p \in \Omega$ when there exist Lagrange multipliers $(\eta, \lambda, \mu) \in \mathbb{R}^t \times \mathbb{R}^s \times \mathbb{R}^m_+$ such that the following two conditions hold:

(i) $L'(p, \eta, \lambda, \mu) = 0$,

(ii) $\mu_j = 0$ for all $j \notin \mathcal{A}(p)$,

where $L(\cdot, \eta, \lambda, \mu)\colon \mathbb{R}^n \to \mathbb{R}$ is the *Lagrangian function* associated with problem (1.1), defined by

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

and $L'(q, \eta, \lambda, \mu)$ is its *gradient.*[1] For $p \in \Omega$, the *linearized cone* $\mathcal{L}(p)$ associated with Ω at p is defined by

$$\mathcal{L}(p) := \left\{v \in \mathbb{R}^n \mid h_\ell(p)^\top v = 0,\ \ell = 1, \ldots, t;\ H_i'(p)^\top v = 0,\ i = 1, \ldots, s;\ G_j'(p)^\top v \le 0,\ j \in \mathcal{A}(p)\right\},$$

and, by the definition of the polar cone, its *polar* $\mathcal{L}(p)^\circ$ is given by

(2.2)
$$\mathcal{L}(p)^\circ = \left\{v \in \mathbb{R}^n \mid v = \sum_{\ell=1}^t \eta_\ell h_\ell'(p) + \sum_{i=1}^s \lambda_i H_i'(p) + \sum_{j=1}^m \mu_j G_j'(p),\ \mu_j \ge 0,\ \eta_\ell, \lambda_i \in \mathbb{R}\right\}.$$

[1] Although the Lagrangian L is a function of four variables, to simplify the notation, we denote by $L'(q, \eta, \lambda, \mu)$ the gradient with respect to the first variable.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

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

$$\mathcal{T} := \{1,\dots,t\}, \qquad \mathcal{S} := \{1,\dots,s\}, \tag{2.3}$$

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

$$[h'_{\bar{\mathcal{T}}}, H'_{\mathcal{I}}, G'_{\mathcal{J}}](q) := \left(\{h'_\ell(q) \mid \ell \in \bar{\mathcal{T}}\} \cup \{H'_i(q) \mid i \in \mathcal{I}\}\right) \cup \{G'_j(q) \mid j \in \mathcal{J}\}. \tag{2.4}$$

If one or two of the sets $\bar{\mathcal{T}}$, $\mathcal{I}$, or $\mathcal{J}$ are empty, the corresponding terms will be omitted from (2.4). For instance, for $\bar{\mathcal{T}} = \varnothing$, the set in (2.4) will be denoted by $[H'_{\mathcal{I}}, G'_{\mathcal{J}}](q) := \{H'_i(q) \mid i \in \mathcal{I}\} \cup \{G'_j(q) \mid j \in \mathcal{J}\}$. In addition, for the sake of simplicity, we set $h' := h'_{\mathcal{T}}$ and $H' := H'_{\mathcal{S}}$. Two constraint qualifications that will be used later are stated below.

Definition 2.1. *A point $p \in \Omega$ is said to satisfy the linear independence constraint qualification (LICQ) if the set $[h', H', G'_{\mathcal{A}(p)}](p)$ is linearly independent. It satisfies the Mangasarian–Fromovitz constraint qualification (MFCQ) if the set $[h', H', G'_{\mathcal{A}(p)}](p)$ is positive-linearly independent with respect to the multiset $[G'_{\mathcal{A}(p)}](p)$.*

We end this section by recalling the *(Euclidean) safeguarded augmented Lagrangian algorithm* for solving problem (1.1), which uses the standard Powell–Hestenes–Rockafellar augmented Lagrangian function given in (1.2); see [4, 8, 21, 38].

Algorithm 2.1 Euclidean safeguarded augmented Lagrangian algorithm.

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

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

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

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

Step 2. *(Estimate new multipliers)* Compute

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

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

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

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

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

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

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

In practice, Step 1 of the algorithm may fail if the subproblem $\min_{q\in\mathbb{R}^n} L_{\rho^k}(q,\eta^k, \lambda^k,\mu^k)$ cannot be solved to the required tolerance ε_k. This may occur due to numerical issues such as divergence of the chosen optimization method, ill-conditioning, limitations in the computational solver, or simply because the subproblem has no solution. In practical implementations, we monitor whether the inner solver produces a minimizer p^k satisfying $\|L'_{\rho^k}(p^k,\eta^k,\lambda^k,\mu^k)\|_\infty \le \varepsilon_k$. If no such p^k is found within the available computational budget, the algorithm declares failure. It is worth noting that Step 1, as stated, may not always be directly implementable, which must be addressed in practical implementations. However, to proceed with our theoretical analysis, we assume that it is successful. Algorithm 2.1 is widely recognized for its ability to generate AKKT sequences (see [7, 20]). Under strict constraint qualifications such as CRSC, or even weaker conditions, all limit points of such a sequence satisfy the KKT conditions (see, for example, [9, 12]). In the subsequent section, we introduce new strict constraint qualifications for problem (1.1). In this new scenario, it is no longer guaranteed that the limit points of the sequence generated by Algorithm 2.1 satisfy the KKT conditions. Therefore, we will employ tools from Riemannian geometry to establish a connection between these new strict constraint qualifications and the Riemannian strict constraint qualifications introduced in [5]. Consequently, the Riemannian version of Algorithm 2.1, an intrinsic algorithm introduced in [55], generates AKKT sequences for problem (1.1). Under these new strict constraint qualifications, we will show that all its limit points satisfy the KKT conditions.

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

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

The assumption (H1) ensures that the set defined by the equality constraints $h(p)=0$ is an embedded submanifold of dimension $n-t$ of $\mathbb{R}^n$. This guarantees a constant dimension for the tangent spaces associated with the submanifold, thus ensuring that projection operators onto these tangent spaces are well-defined and continuous. This assumption is central to our theoretical developments. Additionally, various sets of the form (1.3) satisfying (H1) arise naturally in many practical applications; see, for instance, [23] and [2] for further discussions and examples. It is worth noting that (H1) only needs to be satisfied locally, specifically in a neighborhood of the KKT point. Unless all gradients of the equality constraints vanish (i.e., unless the problem is highly degenerate), one can select a maximally linearly independent set of gradients at that point. The corresponding equality constraints then serve as coordinates for h. By applying Lemma 1.3, the function h defined in this manner satisfies (H1) in a neighborhood of the point under consideration. In this way, the specific submanifold under consideration is as follows:

$$\mathbb{M} := \{q\in\mathbb{R}^n \mid h(q)=0\}, \tag{2.5}$$

where $h=(h_1,\dots,h_t)\colon \mathbb{R}^n\to\mathbb{R}^t$ is continuously differentiable on $\mathbb{R}^n$. We denote by $\langle\cdot,\cdot\rangle$ the metric in $\mathbb{M}$ induced from the Euclidean metric in $\mathbb{R}^n$, and by $\|\cdot\|$ the associated norm. Although a different metric could be used, we adopt this approach for the sake of simplicity.

We use (2.5) to rewrite problem (1.1) in a more convenient form as an intrinsic nonlinear optimization problem, stated equivalently as follows,

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

$$\underset{q\in\mathbb{M}}{\text{minimize}}\, f(q) \qquad \text{subject to} \qquad H(q)=0,\; G(q)\le 0, \tag{2.6}$$

where $f\colon \mathbb{M}\to\mathbb{R}$, $H=(H_1,\dots,H_s)\colon \mathbb{M}\to\mathbb{R}^s$, and $G=(G_1,\dots,G_m)\colon \mathbb{M}\to\mathbb{R}^m$ are continuously differentiable on $\mathbb{M}$. We denote the *intrinsic feasible set* with respect to the submanifold $\mathbb{M}$ for problem (2.6) as $\Omega_{\mathbb{M}}\subset\mathbb{M}$, and we denote the *set of indices of active inequality constraints* at $p\in\Omega_{\mathbb{M}}$ by $\mathcal{A}_{\mathbb{M}}(p)$, given by

$$\Omega_{\mathbb{M}} := \{q\in\mathbb{M} \mid H(q)=0,\; G(q)\le 0\}, \qquad \mathcal{A}_{\mathbb{M}}(p) := \{j\in\{1,\dots,m\} \mid G_j(p)=0\}\,. \tag{2.7}$$

Remark 1. Problems (1.1) and (2.6) are topologically identical and, in particular, have the same solutions. Additionally, from (2.1) and (2.7), we have $\Omega=\Omega_{\mathbb{M}}$ and $\mathcal{A}(p)=\mathcal{A}_{\mathbb{M}}(p)$. However, we emphasize that the functions f, H, and G in problem (1.1) are conceptually different from those in problem (2.6), as they are now defined as functions on the Riemannian manifold. Consequently, the gradients of the functions f, H_i, and G_j are computed using formula (1.9), specifically $\operatorname{grad} f(q)=\operatorname{Proj}_q f'(q)$, $\operatorname{grad} H_i(q)=\operatorname{Proj}_q H_i'(q)$, and $\operatorname{grad} G_j(q)=\operatorname{Proj}_q G_j'(q)$.

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

(i) $\operatorname{grad}\mathbb{L}(p,\lambda,\mu)=0$,

(ii) $\mu_j=0$, for all $j\notin\mathcal{A}_{\mathbb{M}}(p)$,

where $\mathbb{L}(\cdot,\lambda,\mu)\colon\mathbb{M}\to\mathbb{R}$ is the Lagrangian function associated with problem (2.6) and is defined as follows,

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

and its intrinsic gradient,[2] denoted by $\operatorname{grad}\mathbb{L}(q,\lambda,\mu)\in T_q\mathbb{M}$, is given by

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

Similarly to section 2.1, we introduce some notations for conciseness. Let $\mathcal{S}$ as in (2.3), and consider $\mathcal{I}\subset\mathcal{S}$ and $\mathcal{J}\subset\mathcal{A}_{\mathbb{M}}(p)$. For a given $q\in\Omega_{\mathbb{M}}$, define the following sets of vectors:

$$[\operatorname{grad} H_{\mathcal{I}},\operatorname{grad} G_{\mathcal{J}}](q) := \{\operatorname{grad} H_i(q) \mid i\in\mathcal{I}\}\cup\{\operatorname{grad} G_j(q)\mid j\in\mathcal{J}\}. \tag{2.8}$$

If one of the sets $\mathcal{I}$ or $\mathcal{J}$ is empty, then the corresponding set will not appear in (2.8). For instance, for $\mathcal{I}=\varnothing$, the set in (2.4) will be denoted by $[\operatorname{grad} G_{\mathcal{J}}](q) := \{\operatorname{grad} G_j(q)\mid j\in\mathcal{J}\}$. In addition, for the sake of simplicity, we set $\operatorname{grad} H := \operatorname{grad} H_{\mathcal{S}}$. Next, we recall two intrinsic constraint qualifications for problem (2.6), which were introduced in [56] and [15], respectively.

DEFINITION 2.2. *A point $p\in\Omega_{\mathbb{M}}$ is said to satisfy the LICQ if* $[\operatorname{grad} H,\operatorname{grad} G_{\mathcal{A}_{\mathbb{M}}(p)}](p)$ *is linearly independent. It satisfies the MFCQ if* $[\operatorname{grad} H,\operatorname{grad} G_{\mathcal{A}_{\mathbb{M}}(p)}](p)$ *is positive-linearly independent with respect to the multiset* $[\operatorname{grad} G_{\mathcal{A}_{\mathbb{M}}(p)}](p)$.

For the sake of convenience, we recall the following strict constraint qualifications which were originally introduced and studied on a general Riemannian manifold in [5].

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

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

DEFINITION 2.3. *A point $p \in \Omega_{\mathbb{M}}$ is said to satisfy the following:*

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

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

(iii) *the RCRCQ if there exists $\epsilon > 0$ such that the following two conditions hold:*
  (a) *the rank of $[\operatorname{grad} H](q)$ is constant for all $q \in \mathbb{B}_\epsilon(p)$;*
  (b) *let $\mathcal{K} \subset \mathcal{S}$ such that $[\operatorname{grad} H_{\mathcal{K}}](p)$ is a basis for $\operatorname{Span}([\operatorname{grad} H](p))$. For all $\mathcal{J} \subset \mathcal{A}_{\mathbb{M}}(p)$, if $[\operatorname{grad} H_{\mathcal{K}}, \operatorname{grad} G_{\mathcal{J}}](p)$ is linearly dependent, then $[\operatorname{grad} H_{\mathcal{K}}, \operatorname{grad} G_{\mathcal{J}}](q)$ is linearly dependent for all $q \in \mathbb{B}_\epsilon(p)$.*

(iv) *the RCPLD if there exists $\epsilon > 0$ such that the following two conditions hold:*
  (a) *the rank of $[\operatorname{grad} H](q)$ is constant for all $q \in \mathbb{B}_\epsilon(p)$;*
  (b) *Let $\mathcal{K} \subset \mathcal{S}$ such that $[\operatorname{grad} H_{\mathcal{K}}](p)$ is a basis for $\operatorname{Span}([\operatorname{grad} H](p))$. For all $\mathcal{J} \subset \mathcal{A}_{\mathbb{M}}(p)$, if $[\operatorname{grad} H_{\mathcal{K}}, \operatorname{grad} G_{\mathcal{J}}](p)$ is positive-linearly dependent with respect to $[\operatorname{grad} G_{\mathcal{J}}(p)]$, then $[\operatorname{grad} H_{\mathcal{K}}, \operatorname{grad} G_{\mathcal{J}}](q)$ is linearly dependent for all $q \in \mathbb{B}_\epsilon(p)$.*

In Definition 2.3, the key condition is the local persistence of linear dependence or positive linear dependence under small perturbations. It is important to emphasize that while the *local persistence of linear independence* naturally holds under small perturbations (as formally shown in Lemma 1.4), the same does not generally apply to the *local persistence of linear dependence or positive-linear dependence.* Below we recall an intrinsic version of the sequential optimality conditions, which are satisfied at a local minimizer of problem (2.6) in the absence of constraint qualifications. Specifically, we consider the AKKT conditions, introduced in the general context of Riemannian manifolds in [55], and the PAKKT conditions, proposed in [5]. Let $p \in \Omega_{\mathbb{M}}$, $(p^k)_{k\in\mathbb{N}} \subset \mathbb{M}$, $(\lambda^k)_{k\in\mathbb{N}} \subset \mathbb{R}^s$, $(\mu^k)_{k\in\mathbb{N}} \subset \mathbb{R}^m_+$, and consider the following statements:

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

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

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

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

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

and

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

It is worth noting that there is nothing special about the constant "1" in the expression $\gamma_k := \|(1, \lambda^k, \mu^k)\|_\infty$; this is simply a choice that ensures it is always strictly positive, thereby preventing division by zero in algebraic manipulations involving it.

DEFINITION 2.4. *Assume that $p \in \Omega_{\mathbb{M}}$. If there exist sequences $(p^k)_{k\in\mathbb{N}} \subset \mathbb{M}$, $(\lambda^k)_{k\in\mathbb{N}} \subset \mathbb{R}^s$, and $(\mu^k)_{k\in\mathbb{N}} \subset \mathbb{R}^m_+$ such that*

(1) *conditions (i), (ii), and (iii) hold, then $p \in \Omega_{\mathbb{M}}$ is called an AKKT point for problem (2.6);*

(2) *conditions (i), (ii), (iii), and (iv) hold, then $p \in \Omega_{\mathbb{M}}$ is called a PAKKT point for problem (2.6).*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Note that PAKKT is a necessary optimality condition, as shown in [5] for general Riemannian manifolds, not necessarily embedded submanifolds of $\mathbb{R}^n$. It is also worth noting that, for embedded submanifolds of $\mathbb{R}^n$, this result follows directly from [17, 46]. Consequently, AKKT is also a necessary condition, as any point satisfying PAKKT automatically fulfills AKKT.

THEOREM 2.1. *Let $p \in \Omega_{\mathbb{M}}$ be a local minimizer for problem* (2.6). *Then p is a PAKKT point.*

Next, we recall two other constraint qualifications introduced for general Riemannian manifolds in [5]. To facilitate our discussion, let us first establish the following definition: For $p \in \Omega_{\mathbb{M}}$, we denote by $\mathcal{L}_{\mathbb{M}}(p)$ the *linearized cone associated with $\Omega_{\mathbb{M}}$ at p* which is defined as

$$\mathcal{L}_{\mathbb{M}}(p) = \big\{ v \in T_p\mathbb{M} \mid \langle \operatorname{grad} H_i(p), v\rangle = 0,\ i = 1, \ldots, s;\ \langle \operatorname{grad} G_j(p), v\rangle \leq 0\ \ \forall j \in \mathcal{A}_{\mathbb{M}}(p) \big\},$$

and its polar is given by

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

The linearized cone is fundamental in the study of constraint qualifications in both Euclidean and Riemannian settings. Its relationship with the tangent cone, as well as its theoretical foundations in the Riemannian context, is discussed in detail in [15].

DEFINITION 2.5. *A point $p \in \Omega_{\mathbb{M}}$ is said to satisfy the CRSC if there exists $\epsilon > 0$ such that for $\mathcal{J}^-_{\mathbb{M}}(p) = \{j \in \mathcal{A}_{\mathbb{M}}(p) | - \operatorname{grad} G_j(p) \in \mathcal{L}_{\mathbb{M}}(p)^\circ\}$, the rank of the set $[\operatorname{grad} H, \operatorname{grad} G_{\mathcal{J}^-(p)}](q)$ remains constant for all $q \in \mathbb{B}_\epsilon(p)$.*

Next, we revisit the lower constraint qualification previously introduced for general Riemannian manifolds in [5], which serves as the counterpart to its Euclidean version first introduced in [35]; see also [16, 17, 46, 27].

DEFINITION 2.6. *A point $p \in \Omega_{\mathbb{M}}$ satisfies the QN if there do not exist $\lambda \in \mathbb{R}^s$ and $\mu \in \mathbb{R}^m_+$ such that*

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

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

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

In the following, we present an augmented Lagrangian algorithm to address problem (1.1) by refraining from penalizing the constraint set $\{q \in \mathbb{R}^n \mid h(q) = 0\}$ and instead focusing on penalizing the constraint set $\{q \in \mathbb{R}^n \mid H(q) = 0,\ G(q) \leq 0\}$. For this purpose, we recall the intrinsic safeguarded augmented Lagrangian algorithm introduced in [55], specifically designed for optimization on Riemannian manifolds. This algorithm was initially designed to solve problem (2.6), which represents the Riemannian version of problem (1.1). The formulation of the algorithm involves the partial Powell–Hestenes–Rockafellar augmented Lagrangian function defined by (1.4). The *intrinsic safeguarded augmented Lagrangian algorithm* is stated as follows.

In practice, Step 1 of Algorithm 2.2 can fail for similar reasons as described for Algorithm 2.1, such as numerical difficulties or the absence of a feasible solution to the subproblem. Although practical implementations must handle such potential failures, for theoretical purposes, we assume that Step 1 is successfully completed. The capability of Algorithm 2.2 to generate AKKT sequences for problem (2.6) was demonstrated

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Algorithm 2.2 Intrinsic safeguarded augmented Lagrangian algorithm.

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

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

$$\left\|\operatorname{grad} \mathbb{L}_{\rho_k}(p^k, \bar{\lambda}^k, \bar{\mu}^k)\right\| \le \epsilon_k. \tag{2.9}$$

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

Step 2. *(Estimate new multipliers)* Compute

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

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

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

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

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

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

in [55], while in [5] it was proven that it also produces PAKKT sequences. For the reader's convenience and future reference, we revisit the main convergence results of Algorithm 2.2 established in [5].

THEOREM 2.2. *Suppose that $p \in \Omega_{\mathbb{M}}$ satisfies RCPLD or CRSC. If p is an AKKT point, then p is a KKT point for problem* (2.6).

THEOREM 2.3. *Let $p \in \Omega_{\mathbb{M}}$ be a PAKKT point with associated primal sequence $(p^k)_{k\in\mathbb{N}}$ and dual sequence $(\lambda^k, \mu^k)_{k\in\mathbb{N}}$. Assume that p satisfies QN. Then $(\lambda^k, \mu^k)_{k\in\mathbb{N}}$ is a bounded sequence. In particular, p satisfies the KKT conditions for problem* (2.6) *and any limit point of $(\lambda^k, \mu^k)_{k\in\mathbb{N}}$ is a Lagrange multiplier associated with p.*

THEOREM 2.4. *Assume that Algorithm* 2.2 *generates an infinite sequence $(p^k)_{k\in\mathbb{N}}$ with a feasible limit point p, say, $\lim_{k\in K} p^k = p$, where $K \subset \mathbb{N}$ is an infinity sunset. Then, p is a PAKKT point with corresponding primal sequence $(p^k)_{k\in K}$ and dual sequence $(\lambda^k, \mu^k)_{k\in K}$ as generated by Algorithm* 2.2*. In particular, if RCPLD or CRSC hold, p is a KKT point for problem* (2.6)*. Alternatively, if QN holds, p is a KKT point for problem* (2.6) *and $(\lambda^k, \mu^k)_{k\in K}$ is bounded with any of its limit points being a Lagrange multiplier associated with p.*

Note that Algorithm 2.2 tends to generate feasible limit points. This stems from the fact that limit points are stationary to an infeasibility measure; see [55, Theorem 3].

3. Lower SCQ. Let us now recall the Euclidean nonlinear programming problem (1.1). Inspired by the Riemannian approach, we will propose new weak constraint

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

qualifications for problem (1.1). We will show that the new conditions are able to provide standard global convergence results to a constrained augmented Lagrangian method where a subset of linearly independent equality constraints are kept within the subproblems. These are termed lower-level constraints, which inspire the name of the conditions. Let us start by the extension of the sequential optimality conditions AKKT and PAKKT, which will be generated by the constrained algorithm we propose.

That is, in the absence of constraint qualifications, the following definition introduces sequential optimality conditions, which will be shown to be fulfilled by a local minimizer of problem (1.1). Consider the nonlinear programming problem (1.1) under assumption (H1). Let $p \in \Omega$, $(p^k)_{k\in\mathbb{N}} \subset \{q \in \mathbb{R}^n \mid h(q) = 0\}$, $(\eta^k)_{k\in\mathbb{N}} \subset \mathbb{R}^t$, $(\lambda^k)_{k\in\mathbb{N}} \subset \mathbb{R}^s$, $(\mu^k)_{k\in\mathbb{N}} \subset \mathbb{R}^m_+$, and consider the following statements:

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

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

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

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

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

and

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

Definition 3.1. *Assume that $p \in \Omega$. If there exist sequences $(p^k)_{k\in\mathbb{N}} \subset \{q \in \mathbb{R}^n \mid h(q) = 0\}$, $(\eta^k)_{k\in\mathbb{N}} \subset \mathbb{R}^t$, $(\lambda^k)_{k\in\mathbb{N}} \subset \mathbb{R}^s$, and $(\mu^k)_{k\in\mathbb{N}} \subset \mathbb{R}^m_+$ such that*

(1) *conditions (i), (ii), and (iii) hold, then $p \in \Omega$ is called a lower-AKKT point for problem* (1.1)*;*

(2) *conditions (i), (ii), (iii), and (iv) hold, then $p \in \Omega$ is called a lower-PAKKT point for problem* (1.1)*.*

The difference with respect to the standard definition is that the sequence $(x^k)_{k\in\mathbb{N}}$ must be feasible with respect to the equality constraints that satisfy assumption (H1) while the sign control *(iv)* is not required for these constraints; see [17, 46]. The companion lower constraint qualifications are defined as follows.

Definition 3.2. *Consider the nonlinear programming problem* (1.1) *under assumption* (H1)*. Let Ω and $\mathcal{A}(p)$, where $p \in \Omega$, be given by* (2.1)*. The point $p \in \Omega$ is said to satisfy*

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

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

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

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

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

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

(iv) *the lower-RCPLD, if there exists* $\epsilon > 0$ *such that the following two conditions hold:*
  (a) *the rank of* $[h', H'](q)$ *is constant for all* $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$*;*
  (b) *let* $\mathcal{K} \subset \mathcal{S}$ *be such that* $[h', H'_{\mathcal{K}}](p)$ *is a basis for* $\operatorname{Span}([h', H'](p))$*. For all* $\mathcal{J} \subset \mathcal{A}(p)$*, if* $[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](p)$ *is positive-linearly dependent, then the set* $[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](q)$ *is linearly dependent for all* $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$*.*

As in Definition 2.3, the central condition in Definition 3.2 is the local persistence of linear or positive-linear dependence under small perturbations. It is worth emphasizing again that, while the *local persistence of linear independence* naturally holds under small perturbations (as established by Lemma 1.4), the same does not generally hold true for the *local persistence of linear or positive-linear dependence.* As previously mentioned, the motivation behind the development of Definition 3.2 arises from the application of penalty methods or augmented Lagrangian methods for solving nonlinear programming problems where a subset of equality constraints are kept within the subproblems. This approach provides the flexibility to preselect constraints—referred to as lower-level constraints—that align with specific interests or are simpler to handle, while penalizing only the more challenging constraints. Additionally, it ensures that the sequence generated by the chosen minimization method remains feasible for these lower-level constraints. This guarantees that when the stopping criterion for this method is satisfied at some point, the feasibility of those constraints is maintained.

The difference between the new strict constraint qualifications introduced in Definition 3.2 and the standard strict constraint qualifications lies in the requirement that the condition be satisfied at a smaller number of points. Specifically, the new conditions must be satisfied in a neighborhood of the point restricted to a previously chosen set of constraints, namely, $B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$. In contrast, the standard strict constraint qualifications require the point to satisfy the defining condition in a full neighborhood, i.e., $B_\epsilon(p)$, which is naturally more challenging to fulfill. Additionally, the definitions of (R)CRCQ and (R)CPLD are simplified by taking into account assumption (H1). That is, CRCQ and CPLD require taking into consideration $h'_{\bar{\mathcal{I}}}$, where $\bar{\mathcal{I}} \subset \mathcal{T}$, while lower-CRCQ and lower-CPLD require only h'. Also, in item (b) of RCRCQ and RCPLD, $h'_{\bar{\mathcal{K}}}$ is required where $\bar{\mathcal{I}} \subset \mathcal{S}$ while only h' is considered in lower-RCRCQ and lower-RCPLD. Thus it is clear that these definitions imply the usual ones. An example where the implication is strict will be given considering a definition of lower-CRSC which we provide next:

Definition 3.3. *A point* $p \in \Omega$ *is said to satisfy the lower-CRSC) if there exists* $\epsilon > 0$ *such that for* $\mathcal{J}^-(p) = \{j \in \mathcal{A}(p) \mid \pm G'_j(p) \in \mathcal{L}(p)^\circ\}$*, the rank of the set* $[h', H', G'_{\mathcal{J}^-(p)}](q)$ *is constant for all* $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$*.*

Note that the standard CRSC condition requires the matrix $[h', H', G'_{\mathcal{J}^-(p)}](q)$ to have constant rank for all $q \in B_\epsilon(p)$. In contrast, the lower-CRSC condition imposes the same constant rank requirement but only for points satisfying the lower-level constraint, i.e., for $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$. Consequently, we obtain the following theorem.

Theorem 3.1. *Let* $p \in \Omega$ *satisfy CRSC for problem* (1.1) *and assume* (H1)*. Then,* p *also satisfies lower-CRSC.*

The following example demonstrates that the implication in Theorem 3.1 is strict and cannot be reversed.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

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

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

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

$$h'(x) = u, \quad H_1'(x) = 2(u^\top x)(v^\top x)u + (u^\top x)^2 v, \quad G_1'(x) = 2(u^\top x)u - w, \quad G_2'(x) = w. \tag{3.1}$$

We claim that not all $x \in \Omega$ satisfy the usual CRSC constraint qualification. Indeed, it follows from (2.2) and (3.1) that $\mathcal{J}^-(x) = \{1, 2\}$ for all $x \in \Omega$. In addition, the following two statements hold:

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

Thus, no points $x \in \Omega$ satisfy the usual CRSC, as claimed. On the other hand, all $x \in \Omega$ satisfy lower-CRCQ. Indeed, we have $rank(\{h'(y), H_1'(y), G_1'(y), G_2'(y)\}) = 2$ for all $y \in \mathbb{R}^n$ with $h(y) = 0$.

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

DEFINITION 3.4. *A point $p \in \Omega$ satisfies the lower-QN if there do not exist $\eta \in \mathbb{R}^t$, $\lambda \in \mathbb{R}^s$, and $\mu \in \mathbb{R}^m_+$ such that*

(i) $\sum_{i=1}^t \eta_\ell h_\ell(p) + \sum_{i=1}^s \lambda_i H_i'(p) + \sum_{j\in\mathcal{A}(p)} \mu_j G_j'(p) = 0$;
(ii) $\mu_j = 0$ *for all* $j \notin \mathcal{A}(p)$ *and* $(\eta, \lambda, \mu) \neq 0$;
(iii) *for all* $\epsilon > 0$, *there exists* $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ *such that* $\lambda_i H_i(q) > 0$ *for all* $i \in \mathcal{S}$ *with* $\lambda_i \neq 0$ *and* $\mu_j G_j(q) > 0$ *for all* $j \in \mathcal{A}(p)$ *with* $\mu_j > 0$.

It is worth mentioning that lower-QN differs from the usual QN definition introduced in [35] (see also [16, 17, 46, 27]), only in item (iii). In the standard QN definition for problem (1.1), this item is stated as follows:

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

The proof that the usual QN implies lower-QN for problem (1.1) under (H1) is not immediate as in the case of the other lower-SCQs we introduced. Let us prove this.

THEOREM 3.2. *Let $p \in \Omega$ satisfy QN for problem* (1.1) *and assume* (H1). *Then, p satisfies lower-QN.*

Proof. Assume, by contradiction, that $p \in \Omega$ does not satisfy lower-QN. Then, there exist $\eta \in \mathbb{R}^t$, $\lambda \in \mathbb{R}^s$, and $\mu \in \mathbb{R}^m_+$ satisfying conditions (i), (ii), and (iii) in Definition 3.4. Condition (iii) implies the existence of a sequence $(p^k)_{k\in\mathbb{N}} \subset \{q \in \mathbb{R}^n \mid h(q) = 0\}$ such that $\lim_{k\to+\infty} p^k = p$. Furthermore, $\lambda_i H_i(p^k) > 0$ for all $i \in \mathcal{S}$ with $\lambda_i \neq 0$, and $\mu_j G_j(p^k) > 0$ for all $j \in \mathcal{A}(p)$ with $\mu_j > 0$. Define the set $\bar{\mathcal{T}} := \{i \in \mathcal{T} \mid \eta_\ell \neq 0\}$. If $\bar{\mathcal{T}} = \varnothing$, then p also does not satisfy the classical definition of QN. Now, assume $\bar{\mathcal{T}} \neq \varnothing$. Consider the submatrix $[h'_{\bar{\mathcal{T}}}](p)$ of $[h'](p)$, where $[h'](p)$ is the Jacobian matrix of h, and the rows of $[h'_{\bar{\mathcal{T}}}](p)$ correspond to $h_\ell(p)$ for $i \in \bar{\mathcal{T}}$. Since assumption (H1) implies that $h_1'(p), \ldots, h_t'(p)$ are linearly independent, it follows that there is no vector $\beta_{\bar{\mathcal{T}}}$ of order $|\bar{\mathcal{T}}|$, with entries β_i for $i \in \bar{\mathcal{T}}$, such that $[h'_{\bar{\mathcal{T}}}](p)^\top \beta_{\bar{\mathcal{T}}} = 0$ with

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

$\beta_{\bar{\mathcal{T}}} \geq 0$ and $\beta_{\bar{\mathcal{T}}} \neq 0$. By Gordan's alternative theorem (see, for example, [13, p. 51]), it follows that there exists a vector $d \in \mathbb{R}^n$ such that $[h'_{\bar{\mathcal{T}}}](p)d > 0$ or, equivalently, $h_\ell(p)^\top d > 0$ for all $i \in \bar{\mathcal{T}}$. Since we can make this construction replacing any $h'_\ell(p)$ by $-h'_\ell(p)$, we will assume that d is such that $\eta_\ell h_\ell(p)^\top d > 0$ for all $i \in \bar{\mathcal{T}}$. On the other hand, since $\lambda_i H_i(p^k) > 0$ for all $i \in \mathcal{S}$ with $\lambda_i \neq 0$ and $\mu_j G_j(p^k) > 0$ for all $j \in \mathcal{A}(p)$ with $\mu_j > 0$, define sequences $(q^k)_{k\in\mathbb{N}}$ and $(\varepsilon_k)_{k\in\mathbb{N}}$ such that $q^k := p^k + \varepsilon_k d$ with $\lim_{k\to+\infty} \varepsilon_k = 0^+$ and $\varepsilon_k > 0$ such that $\lambda_i H_i(q^k) > 0$ for all $i \in \mathcal{S}$ with $\lambda_i \neq 0$, and $\mu_j G_j(q^k) > 0$ for all $j \in \mathcal{A}(p)$ with $\mu_j > 0$. Given that $\eta_\ell \neq 0$, $\lim_{k\to+\infty} p^k = p$, and $\lim_{k\to+\infty} \varepsilon_k = 0^+$, we have $\lim_{k\to+\infty}(\eta_\ell h_\ell(p^k + \varepsilon_k d) - \eta_\ell h_\ell(p^k))/\varepsilon_k = \eta_\ell h'_\ell(p)^\top d > 0$ for all $i \in \bar{\mathcal{T}}$. Therefore, since $h_\ell(p^k) = 0$ for all k, there exists $k_i \in \mathbb{N}$ such that $\eta_\ell h_\ell(q^k) = \eta_\ell h_\ell(p^k + \varepsilon_k d) > 0$ for all $k > k_i$. Let $\bar{k} = \max\{k_i \mid i \in \bar{\mathcal{T}}\}$. Consequently, since $\lim_{k\to+\infty} p^k = p$, for any $\epsilon > 0$, there exists $q^k \in B_\epsilon(p)$ such that $\eta_\ell h_\ell(q^k) > 0$ for all $i \in \mathcal{T}$ with $\eta_\ell \neq 0$, $\lambda_i H_i(q^k) > 0$ for all $i \in \mathcal{S}$ with $\lambda_i \neq 0$, and $\mu_j G_j(q^k) > 0$ for all $j \in \mathcal{A}(p)$ with $\mu_j > 0$ and all $k \geq \bar{k}$. Since $\eta \in \mathbb{R}^t$, $\lambda \in \mathbb{R}^s$, and $\mu \in \mathbb{R}^m_+$ satisfy conditions (i) and (ii) in Definition 3.4, this contradicts the classical definition of QN. ∎

Let us now show that the above implication is strict. In fact, the following example illustrates that there exist cases where the standard QN condition does not hold, while the lower-QN condition holds. This confirms that the implication from the standard QN to the lower-QN condition is strict and not reversible.

Example 2. Let $n \geq 3$ be a positive integer and $u, v \in \mathbb{R}^{n\times 1}$ be such that $u \neq 0$, $v \neq 0$, $u^\top v = 0$, and $h, H_1, G_1 \colon \mathbb{R}^n \to \mathbb{R}$ be functions defined, respectively, by

$$h(x) := u^\top x - v^\top x, \qquad H_1(x) := (u^\top x)^2 + (v^\top x)^2, \qquad G_1(x) := (u^\top x)^2 - (v^\top x)^2. \tag{3.2}$$

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

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\, f(x) \qquad \text{subject to} \quad h(x) = 0,\ H_1(x) = 0,\ G_1(x) \leq 0. \tag{3.3}$$

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

$$h'(x) = u - v, \qquad H'_1(x) = 2(u^\top x)u + 2(v^\top x)v, \qquad G'_1(x) = 2(u^\top x)u - 2(v^\top x)v. \tag{3.4}$$

Denote by $\Omega = \{x \in \mathbb{R}^n \mid h(x) = 0,\ H_1(x) = 0,\ G_1(x) \leq 0\}$ the feasible set of problem (3.3). Thus, using (3.2) we have $\Omega = \operatorname{Span}\{u, v\}^\perp$ and $\mathcal{A}(x) = \{1\}$ for all $x \in \Omega$. Thus, by using (3.4), note that for items (i) and (ii) of Definition 3.4 to be satisfied at $x \in \Omega$ we must take $\eta = 0$ and $\lambda_1 \in \mathbb{R}$ and $\mu_1 \in \mathbb{R}_+$ are arbitrary. In addition, given $\epsilon > 0$ and $x \in \Omega$, we can choose $y \in B_\epsilon(x)$ with $y \notin \Omega$ satisfying $H_1(y) = (u^\top y)^2 + (v^\top y)^2 > 0$ and $G_1(y) = (u^\top y)^2 - (v^\top y)^2 > 0$. For instance, for $y := x + \alpha u$ we can choose $\alpha > 0$ such that $y \in B_\epsilon(x)$ and $H_1(y) = \alpha^2\|u\|^4 > 0$ and $G_1(y) = \alpha^2\|u\|^4 > 0$. Thus, choosing $\eta_1 = 0$, $\lambda_1 > 0$, and $\mu_1 > 0$, all three items of the usual definition of QN are satisfied. Hence, all $x \in \Omega$ do not satisfy the usual QN. On the other hand, since $G_1(y) = 0$ for all $y \in B_\epsilon(x) \cap \{y \in \mathbb{R}^n :\ h(y) = 0\}$, there is no $\mu_1 > 0$ satisfying item (iii) of Definition 3.4 such that $\mu_1 G_1(y) > 0$. Therefore, all $x \in \Omega$ satisfy lower-QN.

This section concludes by noting that the satisfaction of constraint qualifications can depend on the chosen representation of the optimization problem. However, determining an "optimal" or "ideal" representation in advance is generally challenging or

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

even infeasible in practice. Despite this difficulty, the constraint qualifications introduced in this section explicitly acknowledge and clarify this representation-dependent nature by emphasizing their advantages. Through theoretical analysis and the computational experiments presented in section 5, the practical benefits of the intrinsic formulation are demonstrated, particularly in simplifying the satisfaction of SCQ compared to traditional formulations, which are widely recognized in the literature as fundamental to the study of nonlinear programming problems [7, 10, 9]. It is worth emphasizing that constraint qualifications collectively form a hierarchy, each having its own particular theoretical and computational advantages. We briefly summarize the main characteristics of the most important constraint qualifications. The LICQ is equivalent to the uniqueness of the Lagrange multiplier for any objective function minimized at a given feasible point [51]. Despite this favorable property, LICQ is considered overly restrictive, failing, for example, when constraints are repeated or redundant. In contrast, the MFCQ is widely adopted, with multiple equivalent formulations and significant theoretical implications. Notably, MFCQ ensures compactness of the multiplier set [30] and stability of the feasible set under perturbations [48]. Despite its popularity, MFCQ is sensitive to seemingly innocuous reformulations, such as replacing an equality constraint by two opposing inequalities, potentially resulting in an unbounded multiplier set and thus causing MFCQ to fail [51]. The CRCQ, originally introduced by [37] to study differentiability of value functions, addresses these drawbacks by remaining invariant under redundant constraints or the splitting of equality constraints. Moreover, CRCQ naturally encompasses the linear constraint case, eliminating the need for a separate analysis. However, CRCQ does not incorporate the positivity requirement for multipliers associated with active inequalities. This gap was filled by the CPLD condition introduced by [47]. The CPLD condition is strictly weaker than both MFCQ and CRCQ combined and has gained prominence primarily through its key role in establishing global convergence of optimization methods, such as sequential quadratic programming algorithms and the widely used ALGENCAN solver [20]. It has also found significant applications in bilevel optimization [54, 44], switching constraints [40], and exact penalty methods [31]. Subsequent relaxed versions of CRCQ and CPLD, namely, RCRCQ [45] and RCPLD [10], were introduced to further refine these conditions, retaining their beneficial properties. Additionally, classical and weaker conditions such as Abadie's constraint qualification (ACQ) and Slater's condition also have important theoretical roles. ACQ states that the linearized cone coincides with the tangent cone, a crucial property for establishing duality results [1]. Slater's condition, widely applied in convex optimization, guarantees strong duality and the existence of multipliers by requiring strict feasibility for inequality constraints [50]. Last, the QN constraint qualification, initially proposed by [35], is strictly weaker than MFCQ. It uniquely ensures boundedness of approximate multiplier sequences generated by primal-dual algorithms, even when the actual set of multipliers may be unbounded [7]. This characteristic makes QN particularly relevant in algorithmic convergence analysis [9].

4. Connecting the extrinsic and intrinsic approaches. This section establishes connections between the extrinsic concepts discussed earlier in section 3, related to the nonlinear optimization problem presented in (1.1), and the ideas addressed in subsection 2.2 regarding the nonlinear optimization problem presented intrinsically in (2.6). Our goal is to establish new global convergence results of an augmented Lagrangian algorithm for the Euclidean problem by means of the equivalent optimization problem on an embedded manifold. In order to do this we will first show that the

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Euclidean KKT conditions for problem (1.1) and the Riemmanian KKT conditions for the equivalent problem (2.6) are in fact equivalent. This appears to be new as we only found a proof of this fact under convexity assumptions; see [56].

Recall that we are under assumption (H1), hence, it follows from (1.8) that the mapping Proj_q is well-defined for all $q \in \mathbb{R}^n$.

THEOREM 4.1. *A point* $p \in \Omega$ *is a KKT (respectively, lower-AKKT or lower-PAKKT) point of problem* (1.1) *if and only if* $p \in \Omega_{\mathbb{M}}$ *is a KKT (respectively, AKKT or PAKKT) point of problem* (2.6).

Proof. First, we establish the equivalence for KKT points. Assume that $p \in \Omega$ is a KKT point for problem (1.1). Thus, there exist $(\eta, \lambda, \mu) \in \mathbb{R}^t \times \mathbb{R}^s \times \mathbb{R}^m_+$ such that $L'(p, \eta, \lambda, \mu) = 0$ and $\mu_j = 0$ for all $j \notin \mathcal{A}(p)$, i.e.,

$$f'(p) + \sum_{\ell=1}^{t} \eta_\ell h'_\ell(p) + \sum_{i=1}^{s} \lambda_i H'_i(p) + \sum_{j=1}^{m} \mu_j G'_j(p) = 0, \qquad \mu_j = 0 \ \forall\, j \notin \mathcal{A}(p). \tag{4.1}$$

From (1.7), it follows that $h'(p)^\top \eta = \sum_{i=1}^{t} \eta_\ell h'_\ell(p) \in T_p\mathbb{M}^\perp$. Consequently, considering (1.8), we conclude that $\mathrm{Proj}_p\, h'(p)^\top \eta = 0$. Thus, by applying Proj_p to (4.1) and using (1.9) along with the fact that $\mathcal{A}(p) = \mathcal{A}_{\mathbb{M}}(p)$, we obtain

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

Therefore, $\mathrm{grad}\, \mathbb{L}(p, \lambda, \mu) = 0$ and $\mu_j = 0$ for all $j \notin \mathcal{A}_{\mathbb{M}}(p)$. Hence, since we also have $p \in \Omega_{\mathbb{M}}$, it follows that p is a KKT point for problem (2.6) as well.

Reciprocally, suppose that $p \in \Omega_{\mathbb{M}}$ is a KKT point for problem (2.6). Thus, there exist $(\lambda, \mu) \in \mathbb{R}^s \times \mathbb{R}^m_+$ such that $\mathrm{grad}\, \mathbb{L}(p, \lambda, \mu) = 0$ and $\mu_j = 0$ for all $j \notin \mathcal{A}_{\mathbb{M}}(p)$, i.e.,

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

Therefore, using the projection formula Proj_p and (1.9), we have

$$\mathrm{Proj}_p \left(f'(p) + \sum_{i=1}^{s} \lambda_i H'_i(p) + \sum_{j=1}^{m} \mu_j G'_j(p) \right) = 0, \qquad \mu_j = 0 \ \forall\, j \notin \mathcal{A}_{\mathbb{M}}(p).$$

Hence, by (1.6), we conclude that $f'(p) + \sum_{i=1}^{s} \lambda_i H'_i(p) + \sum_{j=1}^{m} \mu_j G'_j(p) \in T_p\mathbb{M}^\perp$. Thus, (1.7) implies that there exists $\eta \in \mathbb{R}^t$ such that

$$f'(p) + \sum_{i=1}^{s} \lambda_i H'_i(p) + \sum_{j=1}^{m} \mu_j G'_j(p) = -\sum_{i=1}^{t} \eta_\ell h'_\ell(p).$$

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

$$f'(p) + \sum_{\ell=1}^{t} \eta_\ell h'_\ell(p) + \sum_{i=1}^{s} \lambda_i H'_i(p) + \sum_{j=1}^{m} \mu_j G'_j(p) = 0, \qquad \mu_j = 0 \ \forall\, j \notin \mathcal{A}_{\mathbb{M}}(p).$$

Therefore, $L'(p, \eta, \lambda, \mu) = 0$ and $\mu_j = 0$ for all $j \notin \mathcal{A}_{\mathbb{M}}(p)$. Since $p \in \Omega_{\mathbb{M}} = \Omega$ and $\mathcal{A}_{\mathbb{M}}(p) = \mathcal{A}(p)$, the point p satisfies the KKT conditions for problem (1.1), concluding the proof for KKT points.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

To establish the statements regarding other equivalences, we first note that the projection map Proj_q, defined in (1.8), is continuous. Consequently, the part of the proof involving the Lagrangian follows arguments similar to those used for establishing the statement related to KKT points. Given that $\mathbb{M} = \{q \in \mathbb{R}^n \mid h(q) = 0\}$, $\Omega_{\mathbb{M}} = \Omega$, and $\mathcal{A}_{\mathbb{M}}(p) = \mathcal{A}(p)$, we conclude the proof by observing that the additional conditions required are directly equivalent. ∎

The next theorem shows that Euclidean LICQ and MFCQ are equivalent to their Riemannian counterparts. The proof for the LICQ can be found in [56]. Since both proofs follow directly from Lemma 1.1, we omit them.

THEOREM 4.2. *A point $p \in \Omega$ satisfies LICQ (respectively, MFCQ) for problem* (1.1) *if and only if $p \in \Omega_{\mathbb{M}}$ satisfies LICQ (respectively, MFCQ) for problem* (2.6).

We now begin the discussion where we establish the connection between the extrinsic Definition 3.2 and the intrinsic Definition 2.3. Our discussion commences by establishing the connection between the first two items of these definitions.

THEOREM 4.3. *A point $p \in \Omega$ satisfies lower-CRCQ (respectively, lower-CPLD) for problem* (1.1) *if and only if $p \in \Omega_{\mathbb{M}}$ satisfies CRCQ (respectively, CPLD) for problem* (2.6).

Proof. Suppose that $p \in \Omega$ satisfies lower-CRCQ (respectively, lower-CPLD) for problem (1.1). Assume, by contradiction, that $p \in \Omega_{\mathbb{M}}$ does not satisfies CRCQ (respectively, CPLD) for problem (2.6). According to Definition 2.3, there exist $\mathcal{I} \subset \mathcal{S}$ and $\mathcal{J} \subset \mathcal{A}_{\mathbb{M}}(p) = \mathcal{A}(p)$ such that $[\operatorname{grad} H_{\mathcal{I}}, \operatorname{grad} G_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent), and for each $k \in \mathbb{N}$, there exists $q_k \in \mathbb{B}_{1/k}(p)$ such that $[\operatorname{grad} H_{\mathcal{I}}, \operatorname{grad} G_{\mathcal{J}}](q_k)$ is linearly independent. Taking into account that $[h'](p)$ is linearly independent and $q_k \in \mathbb{B}_{1/k}(p)$ for all $k \in \mathbb{N}$, we can assume without loss of generality that $[h'](q_k)$ is also linearly independent. Consequently, by applying Lemma 1.1 and considering (1.5) and (1.9), we conclude that $[h', H'_{\mathcal{I}}, G'_{\mathcal{J}}](q_k)$ is also linearly independent for each $q_k \in \mathbb{B}_{1/k}(p)$. On the other hand, since $[\operatorname{grad} H_{\mathcal{I}}, \operatorname{grad} G_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent), applying again Lemma 1.1 and taking into account (1.5) and (1.9), we obtain that $[h', H'_{\mathcal{I}}, G'_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent). Thus, since $p \in \Omega$ satisfies lower-CRCQ (respectively, lower-CPLD) for problem (1.1), item (i) (respectively, item (ii)) of Definition 3.2 implies that there exists $\epsilon > 0$ such that $[h', H'_{\mathcal{I}}, G'_{\mathcal{J}}](q)$ is linearly dependent (respectively, positive-linearly dependent) for all $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$. Hence, as $B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ is an open subset of $\mathbb{M}$ and $\lim_{k \to +\infty} q_k = p$, there exists $\bar{k} \in \mathbb{N}$ such that $q_{\bar{k}} \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\} \cap \mathbb{B}_{1/\bar{k}}(p)$ and $[h', H'_{\mathcal{I}}, G'_{\mathcal{J}}](q_{\bar{k}})$ is linearly dependent (respectively, positive-linearly dependent), which is a contradiction. Therefore, $p \in \Omega_{\mathbb{M}}$ satisfies CRCQ (respectively, CPLD) for problem (2.6).

Reciprocally, suppose that $p \in \Omega_{\mathbb{M}}$ satisfies CRCQ (respectively, CPLD) for problem (2.6). Assume, by contradiction, that $p \in \Omega$ does not satisfy lower-CRCQ (respectively, lower-CPLD) for problem (1.1). Thus, there exist $\mathcal{I} \subset \mathcal{S}$ and $\mathcal{J} \subset \mathcal{A}(p) = \mathcal{A}_{\mathbb{M}}(p)$ such that $[h', H'_{\mathcal{I}}, G'_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent) and, for all $k \in \mathbb{N}$, there exists $q_k \in B_{1/k}(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ such that $[h', H'_{\mathcal{I}}, G'_{\mathcal{J}}](q_k)$ is linearly independent. Hence, by using Lemma 1.1 and considering (1.5) and (1.9), the set $[\operatorname{grad} H_{\mathcal{I}}, \operatorname{grad} G_{\mathcal{J}}](q_k)$ is linearly independent. Since $[h', H'_{\mathcal{I}}, G'_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent), by employing Lemma 1.1 and considering (1.5) and (1.9), the set $[\operatorname{grad} H_{\mathcal{I}}, \operatorname{grad} G_{\mathcal{J}}](p)$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

is linearly dependent (respectively, positive-linearly dependent). Thus, considering that $p \in \Omega_{\mathbb{M}}$ satisfies CRCQ (respectively, CPLD) for problem (2.6), there exists $\epsilon > 0$ such that $[\operatorname{grad} H_{\mathcal{I}}, \operatorname{grad} G_{\mathcal{J}}](q)$ is linearly dependent for all $q \in \mathbb{B}_\epsilon(p)$. Since $B_{1/k}(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ is an open subset of $\mathbb{M}$ and $\lim_{k\to+\infty} q_k = p$, there exists $\bar{k} \in \mathbb{N}$ such that $q_{\bar{k}} \in \mathbb{B}_\epsilon(p) \cap B_{1/\bar{k}}(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ and $[\operatorname{grad} H_{\mathcal{I}}, \operatorname{grad} G_{\mathcal{J}}](q_{\bar{k}})$ is linearly dependent, which is a contradiction. Therefore, $p \in \Omega$ satisfies lower-CRCQ (respectively, lower-CPLD) for problem (1.1), which concludes the proof. ∎

The following theorem establishes the connection between lower-RCRCQ and lower-RCPLD for problem (1.1), and RCRCQ and RCPLD for problem (2.6).

THEOREM 4.4. *A point $p \in \Omega$ satisfies lower-RCRCQ (respectively, lower-RCPLD) for problem* (1.1) *if and only if $p \in \Omega_{\mathbb{M}}$ satisfies RCRCQ (respectively, RCPLD) for problem* (2.6).

Proof. Suppose that $p \in \Omega$ satisfies lower-RCRCQ (respectively, Lower-RCPLD) for problem (1.1). Assume, by contradiction, that $p \in \Omega_{\mathbb{M}}$ does not satisfy RCRCQ (respectively, RCPLD) for problem (2.6). Thus, taking $\mathcal{K} \subset \mathcal{S}$ such that $[\operatorname{grad} H_{\mathcal{K}}](p)$ is a basis for $\operatorname{Span}([\operatorname{grad} H](p))$, at least one of the following two conditions holds for each $k \in \mathbb{N}$:

(a) there exists $q_k \in \mathbb{B}_{1/k}(p)$ such that $|\mathcal{K}| = \operatorname{rank}([\operatorname{grad} H](q)) \neq \operatorname{rank}([\operatorname{grad} H](q_k))$;

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

First, assume that (a) holds for infinitely many $k \in \mathbb{N}$. Using Lemma 1.3, it follows that there exists a subsequence $(q_{k_j})_{j\in\mathbb{N}}$ of $(q_k)_{k\in\mathbb{N}}$ such that $|\mathcal{K}| < \operatorname{rank}([\operatorname{grad} H](q_{k_j}))$ for all $j \in \mathbb{N}$. Thus, since $\mathcal{S}$ is finite, there exists $\bar{\mathcal{K}} \subset \{1, \ldots, s\}$ satisfying

$$|\mathcal{K}| < |\bar{\mathcal{K}}| := \operatorname{rank}([\operatorname{grad} H_{\bar{\mathcal{K}}}](q_{k_j})) \qquad \forall j \in \mathbb{N}. \tag{4.2}$$

In particular, the definition of $\bar{\mathcal{K}}$ implies that $[\operatorname{grad} H_{\bar{\mathcal{K}}}](q_{k_j})$ is linearly independent. Since $[h'](q_{k_j})$ is linearly independent, applying Lemma 1.1 and using (1.5) and (1.9), we conclude that $[h', H'_{\bar{\mathcal{K}}}](q_{k_j})$ is also linearly independent for all $j \in \mathbb{N}$. Hence, due to $\bar{\mathcal{K}} \subset \mathcal{S}$, we have

$$t + |\bar{\mathcal{K}}| \leq \operatorname{rank}([h', H'](q_{k_j})) \qquad \forall j \in \mathbb{N}. \tag{4.3}$$

Since $p \in \Omega$ satisfies lower-RCRCQ (respectively, lower-RCPLD) for problem (1.1), there exists $\epsilon > 0$ such that $\operatorname{rank}([h', H'](q))$ is constant for all $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$. Thus, considering that $[h'](q)$ is linearly independent for all $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$, there exists $\hat{\mathcal{K}} \subset \mathcal{S}$ such that $[H'_{\hat{\mathcal{K}}}](p)$ is linearly independent and

$$t + |\hat{\mathcal{K}}| = \operatorname{rank}([h', H'](q)) \qquad \forall q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}. \tag{4.4}$$

Since $\lim_{j\to+\infty} q_{k_j} = p$, there exist j_0 such that $q_{k_j} \in B_\epsilon(p) \cap \{p \in \mathbb{R}^n \mid h(p) = 0\}$ for all $j \geq j_0$. Hence, (4.3) and (4.4) imply that

$$|\bar{\mathcal{K}}| \leq |\hat{\mathcal{K}}|. \tag{4.5}$$

Thus, utilizing Lemma 1.1 and considering (1.5) and (1.9), it follows that $[\operatorname{grad} H_{\hat{\mathcal{K}}}](p)$ is also linearly independent. Hence, considering that $[\operatorname{grad} H_{\mathcal{K}}](p)$ is a basis for $\operatorname{Span}([\operatorname{grad} H](p))$, we conclude that $|\hat{\mathcal{K}}| \leq |\mathcal{K}|$. Thus, by (4.2), we obtain $|\hat{\mathcal{K}}| \leq$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

$|\mathcal{K}| < |\bar{\mathcal{K}}|$, contradicting (4.5). Therefore, (a) must hold only for a finite number of $k \in \mathbb{N}$. Hence, it follows that (b) holds for all sufficiently large $k \in \mathbb{N}$. We may assume, without loss of generality, that (b) holds for all $k \in \mathbb{N}$. Note that by applying Lemma 1.2 and using (1.5) and (1.9), we obtain that $[h', H'_{\mathcal{K}}](p)$ is a basis for $\mathrm{Span}([h', H'](p))$. Let $\mathcal{J} \subset \mathcal{A}(p)$ be such that $[\operatorname{grad} H_{\mathcal{K}}, \operatorname{grad} G_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent) and $(q_n)_{k\in\mathbb{N}} \subset \mathbb{B}_{1/k}(p)$ be a sequence such that $[\operatorname{grad} H_{\mathcal{K}}, \operatorname{grad} G_{\mathcal{J}}](q_k)$ is linearly independent for all $k \in \mathbb{N}$. Given that $[h'](q_k)$ is linearly independent for all $k \in \mathbb{N}$, by applying Lemma 1.1 and considering (1.5) and (1.9), we have

$$[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](q_k) \tag{4.6}$$

is also linearly independent for all $k \in \mathbb{N}$. Now, as $[\operatorname{grad} H_{\mathcal{K}}, \operatorname{grad} G_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent), using Lemma 1.1 and taking into account (1.5) and (1.9), we obtain that $[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](p)$ is also linearly dependent (respectively, positive-linearly dependent). Since the point $p \in \Omega$ satisfies lower-RCRCQ (respectively, lower-RCPLD) for problem (1.1), there exists $\epsilon > 0$ such that $[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](q)$ is linearly dependent for all $q \in B_\epsilon(p) \cap \{p \in \mathbb{R}^n \mid h(p) = 0\}$. Thus, due to $\lim_{k\to+\infty} q_k = p$, there exist $q_k \in B_\epsilon(p)$ such that the set in (4.6) is linearly dependent, which is also a contradiction. Therefore, $p \in \Omega_{\mathbb{M}}$ satisfies RCRCQ (respectively, RCPLD) for problem (2.6).

Reciprocally, suppose that $p \in \Omega_{\mathbb{M}}$ satisfies RCRCQ (respectively, RCPLD) for problem (2.6). Assume, by contradiction, that $p \in \Omega$ does not satisfy lower-RCRCQ (respectively, lower-RCPLD) for problem (1.1). Thus, taking $\mathcal{K} \subset \mathcal{S}$ such that $[h', H'_{\mathcal{K}}](p)$ is a basis for $\mathrm{Span}([h', H'](p))$, at least one of the following two conditions holds for each $k \in \mathbb{N}$:

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

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

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

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

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

$$t + |\mathcal{K}| < t + |\bar{\mathcal{K}}| := \mathrm{rank}([h', H'_{\bar{\mathcal{K}}}](q_{k_j})) \qquad \forall j \in \mathbb{N}. \tag{4.7}$$

Hence, $[h', H'_{\bar{\mathcal{K}}}](q_{k_j})$ is linearly independent for all $j \in \mathbb{N}$. Thus, using Lemma 1.1 and considering (1.5) and (1.9), we conclude that $[\operatorname{grad} H_{\bar{\mathcal{K}}}](q_{k_j})$ is also linearly independent for all $j \in \mathbb{N}$. In particular, we have

$$|\bar{\mathcal{K}}| \le \mathrm{rank}([\operatorname{grad} H](q_{k_j})) \qquad \forall j \in \mathbb{N}. \tag{4.8}$$

Since $p \in \Omega_{\mathbb{M}}$ satisfies RCRCQ (respectively, RCPLD) for problem (2.6), there exists $\epsilon > 0$ such that $\mathrm{rank}([\operatorname{grad} H](q))$ is constant for all $q \in \mathbb{B}_\epsilon(p)$. Let $\hat{\mathcal{K}} \subset \{1, \dots, s\}$ be such that $[\operatorname{grad} H_{\hat{\mathcal{K}}}](p)$ is a basis of $\mathrm{Span}([\operatorname{grad} H](p))$. Thus,

$$|\hat{\mathcal{K}}| = \mathrm{rank}([\operatorname{grad} H](q)) \qquad \forall q \in \mathbb{B}_\epsilon(p). \tag{4.9}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

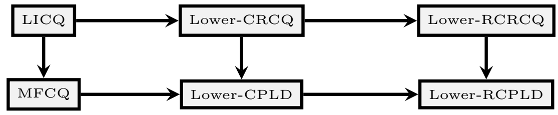

Fig. 1. *Lower SCQs for problem* (1.1).

Hence, due to $\lim_{j\to+\infty} q_{k_j} = p$, there exist $q_{k_j} \in \mathbb{B}_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$, which, together with (4.8) and (4.9), implies that $|\bar{\mathcal{K}}| \leq |\hat{\mathcal{K}}|$. On the other hand, taking into account Lemma 1.1, we conclude that $[h', H'_{\hat{\mathcal{K}}}](p)$ is linearly independent. Since $[h', H'_{\mathcal{K}}](p)$ is a basis for $\mathrm{Span}([h', H'](p))$, we have $|\hat{\mathcal{K}}| \leq |\mathcal{K}|$. The latter two inequalities imply that $|\bar{\mathcal{K}}| \leq |\hat{\mathcal{K}}| \leq |\mathcal{K}|$, contradicting (4.7). Therefore, (c) must hold only for a finite number of indexes $k \in \mathbb{N}$. Hence, without loss of generality, we may assume that (d) holds for all $k \in \mathbb{N}$. Since $[h', H'_{\mathcal{K}}](p)$ is a basis for $\mathrm{Span}([h', H'](p))$, applying Lemma 1.2 and using (1.5) and (1.9), we obtain that $[\mathrm{grad}\, H_{\mathcal{K}}](p)$ is also a basis for $\mathrm{Span}([\mathrm{grad}\, H](p))$. Let $\mathcal{J} \subset \mathcal{A}(p)$ be such that $[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent), and consider a sequence $(q_k)_{k\in\mathbb{N}} \subset B_{1/k}(p)$ such that $[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](q_k)$ is linearly independent for all $k \in \mathbb{N}$. Since $[h'](q_k)$ is linearly independent, applying Lemma 1.1 and considering (1.5) and (1.9), we have

$$[\mathrm{grad}\, H_{\mathcal{K}}, \mathrm{grad}\, G_{\mathcal{J}}](q_k) \tag{4.10}$$

is also linearly independent. Now, as $[h', H'_{\mathcal{K}}, G'_{\mathcal{J}}](p)$ is linearly dependent (respectively, positive-linearly dependent), applying Lemma 1.1 and considering (1.5) and (1.9), we have that $[\mathrm{grad}\, H_{\mathcal{K}}, \mathrm{grad}\, G_{\mathcal{J}}](p)$ is also linearly dependent (respectively, positive-linearly dependent). Taking into account that $p \in \Omega_{\mathbb{M}}$ satisfies RCRCQ (respectively, RCPLD) for problem (2.6), there exists $\epsilon > 0$ such that $[\mathrm{grad}\, H_{\mathcal{K}}, \mathrm{grad}\, G_{\mathcal{J}}](q)$ is linearly dependent for all $q \in \mathbb{B}_\epsilon(p)$. Given that $\lim_{k\to+\infty} q_k = p$, there exist $q_k \in \mathbb{B}_\epsilon(p)$ such that the set in (4.10) is linearly dependent, which is a contradiction. ∎

As a consequence of Theorems 4.2, 4.3, and 4.4, along with the relationships established for SCQs in [5] for a general Riemannian manifold, the diagram in Figure 1 illustrates the relationship among the lower SCQs given in Definition 3.2.

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

Lemma 4.1. *Let $\mathcal{J}^-(p)$ and $\mathcal{J}^-_{\mathbb{M}}(p)$ be as in Definitions* 3.3 *and* 2.5, *respectively. Then, it holds that $\mathcal{J}^-(p) = \mathcal{J}^-_{\mathbb{M}}(p)$.*

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

Theorem 4.5. *A point $p \in \Omega$ satisfies lower-CRSC for problem* (1.1) *if and only if $p \in \Omega_{\mathbb{M}}$ satisfies CRSC for problem* (2.6).

Proof. Suppose that $p \in \Omega$ satisfies lower-CRSC for problem (1.1). Assume, by contradiction, that $p \in \Omega_{\mathbb{M}}$ does not satisfy CRSC for problem (2.6). Thus, for each $k \in \mathbb{N}$, there exists $q_k \in \mathbb{B}_{1/k}(p)$ such that $|\mathcal{K}| := \mathrm{rank}([\mathrm{grad}\, H, \mathrm{grad}\, G_{\mathcal{J}^-_{\mathbb{M}}(p)}](p)) \neq \mathrm{rank}([\mathrm{grad}\, H, \mathrm{grad}\, G_{\mathcal{J}^-_{\mathbb{M}}(p)}](q_k))$. Using Lemma 1.3, we may assume that there exists a subsequence $(q_{k_j})_{j\in\mathbb{N}}$ of $(q_k)_{k\in\mathbb{N}}$ such that

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

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

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

$$|\mathcal{K}| < |\bar{\mathcal{K}}| + |\bar{\mathcal{K}}^-|, \tag{4.11}$$

and $[\operatorname{grad} H_{\bar{\mathcal{K}}}, \operatorname{grad} G_{\bar{\mathcal{K}}^-}](q_{k_j})$ is linearly independent. Hence, considering that $[h'](q_{k_j})$ is linearly independent, applying Lemma 1.1 and using (1.5) and (1.9), we conclude that $[h', H'_{\bar{\mathcal{K}}}, G'_{\bar{\mathcal{K}}^-}](q_{k_j})$ is also linearly independent for all $j \in \mathbb{N}$. By using Lemma 4.1, we obtain that $\mathcal{J}^-_{\mathbb{M}}(p) = \mathcal{J}^-(p)$. Therefore, due to $\bar{\mathcal{K}} \subset \mathcal{S}$ and $\bar{\mathcal{K}}^- \subset \mathcal{J}^-_{\mathbb{M}}(p) = \mathcal{J}^-(p)$, we have

$$t + |\bar{\mathcal{K}}| + |\bar{\mathcal{K}}^-| \le \operatorname{rank}([h', H', G'_{\mathcal{J}^-(p)}](q_{k_j})) \qquad \forall j \in \mathbb{N}. \tag{4.12}$$

Since $p \in \Omega$ satisfies lower-CRSC for problem (1.1), there exists $\epsilon > 0$ such that $\operatorname{rank}([h', H', G'_{\mathcal{J}^-(p)}](q))$ is constant for all $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$. Thus, there exist $\hat{\mathcal{K}} \subset \mathcal{S}$ and $\hat{\mathcal{K}}^- \subset \mathcal{J}^-(p)$ such that $[h', H'_{\hat{\mathcal{K}}}, G'_{\hat{\mathcal{K}}^-}](q)$ is linearly independent for all $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$, and

$$t + |\hat{\mathcal{K}}| + |\hat{\mathcal{K}}^-| = \operatorname{rank}([h', H', G'_{\mathcal{J}^-(p)}](q)) \tag{4.13}$$

for all $q \in B_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ (perhaps decreasing ϵ if necessary). Given that $\lim_{j\to+\infty} q_{k_j} = p$, there exists $\hat{j}$ such that $q_{k_j} \in B_\epsilon(p) \cap \{p \in \mathbb{R}^n \mid h(p) = 0\}$ for all $j \ge \hat{j}$. Hence, (4.12) and (4.13) imply that

$$|\bar{\mathcal{K}}| + |\bar{\mathcal{K}}^-| \le |\hat{\mathcal{K}}| + |\hat{\mathcal{K}}^-|. \tag{4.14}$$

On the other hand, since $[h', H'_{\hat{\mathcal{K}}}, G'_{\hat{\mathcal{K}}^-}](q)$ is linearly independent, applying Lemma 1.1 and taking into account (1.5) and (1.9), we obtain that $[\operatorname{grad} H'_{\hat{\mathcal{K}}}, \operatorname{grad} G_{\hat{\mathcal{K}}^-}](q)$ is also linearly independent. Thus, considering that $|\mathcal{K}| := \operatorname{rank}([\operatorname{grad} H, \operatorname{grad} G_{\mathcal{J}^-_{\mathbb{M}}(p)}](p))$, we have $|\hat{\mathcal{K}}| + |\hat{\mathcal{K}}^-| \le |\mathcal{K}|$. This, together with (4.11), implies $|\hat{\mathcal{K}}| + |\hat{\mathcal{K}}^-| \le |\mathcal{K}| < |\bar{\mathcal{K}}| + |\bar{\mathcal{K}}^-|$, leading to a contradiction with (4.14). Therefore, $p \in \Omega_{\mathbb{M}}$ satisfies CRSC for problem (2.6).

Reciprocally, suppose that $p \in \Omega_{\mathbb{M}}$ satisfies CRSC for problem (2.6). Consider $\mathcal{J}^-(p) = \{j \in \mathcal{A}(p) \mid -G'_j(p) \in \mathcal{L}(p)^\circ\}$ and assume, by contradiction, that $p \in \Omega$ does not satisfy CRSC for problem (1.1). Thus, for each $k \in \mathbb{N}$, there exists $q_k \in B_{1/k}(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$ such that $t + |\mathcal{K}| := \operatorname{rank}([h', H', G'_{\mathcal{J}^-(p)}](q)) \neq \operatorname{rank}([h', H', G'_{\mathcal{J}^-(p)}](q_k))$. Therefore, by using Lemma 1.3, we conclude that there exists a subsequence $(q_{k_j})_{j\in\mathbb{N}}$ of $(q_k)_{k\in\mathbb{N}}$ such that $t + |\mathcal{K}| < \operatorname{rank}([h', H', G'_{\mathcal{J}^-(p)}](q_{k_j}))$ for all $j \in \mathbb{N}$. Considering that $\mathcal{S}$ and $\mathcal{J}^-(p)$ are finite, there exist $\bar{\mathcal{K}} \subset \mathcal{S}$ and $\bar{\mathcal{K}}^- \subset \mathcal{J}^-(p) = \mathcal{J}^-_{\mathbb{M}}(p)$ such that $t + |\mathcal{K}| < t + |\bar{\mathcal{K}}| + |\bar{\mathcal{K}}^-| := \operatorname{rank}([h', H'_{\bar{\mathcal{K}}}, G'_{\bar{\mathcal{K}}^-}](q_{k_j}))$ and $[h', H'_{\bar{\mathcal{K}}}, G'_{\bar{\mathcal{K}}^-}](q_{k_j})$ are linearly independent for all $j \in \mathbb{N}$. Thus, by applying Lemma 1.1 and considering (1.5) and (1.9), we conclude that $[\operatorname{grad} H_{\bar{\mathcal{K}}}, \operatorname{grad} G_{\bar{\mathcal{K}}^-}](q_{k_j})$ are also linearly independent for all $j \in \mathbb{N}$. In particular, we have

$$|\bar{\mathcal{K}}| + |\bar{\mathcal{K}}^-| \le \operatorname{rank}([\operatorname{grad} H, \operatorname{grad} G_{\mathcal{J}^-_{\mathbb{M}}(p)}](q_{k_j})) \qquad \forall j \in \mathbb{N}. \tag{4.15}$$

Since $p \in \Omega_{\mathbb{M}}$ satisfies CRSC for problem (2.6), there exists $\epsilon > 0$ such that rank $([\operatorname{grad} H, \operatorname{grad} G_{\mathcal{J}^-_{\mathbb{M}}(p)}](q))$ remains constant for all $q \in \mathbb{B}_\epsilon(p)$. Let $\hat{\mathcal{K}} \subset \mathcal{S}$ and $\hat{\mathcal{K}}^- \subset \mathcal{J}^-(p) = \mathcal{J}^-_{\mathbb{M}}(p)$ be sets such that $[\operatorname{grad} H_{\hat{\mathcal{K}}}, \operatorname{grad} G_{\hat{\mathcal{K}}^-}](p)$ is a basis of $\operatorname{Span}([\operatorname{grad} H, \operatorname{grad} G_{\mathcal{J}^-_{\mathbb{M}}(p)}](p))$. Thus,

$$|\hat{\mathcal{K}}| + |\hat{\mathcal{K}}^-| = \operatorname{rank}([\operatorname{grad} H, \operatorname{grad} G_{\mathcal{J}^-_{\mathbb{M}}(p)}](q)) \qquad \forall q \in \mathbb{B}_\epsilon(p). \tag{4.16}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Hence, given that $\lim_{j\to+\infty} q_{k_j} = p$, there exist $q_{k_j} \in \mathbb{B}_\epsilon(p) \cap \{q \in \mathbb{R}^n \mid h(q) = 0\}$, which together with (4.15) and (4.16) imply that

$$|\bar{\mathcal{K}}| + |\bar{\mathcal{K}}^-| \leq |\hat{\mathcal{K}}| + |\hat{\mathcal{K}}^-|. \tag{4.17}$$

On the other hand, by applying Lemma 1.1 and taking into account (1.5) and (1.9), we obtain that $[h', H'_{\hat{\mathcal{K}}}, G'_{\hat{\mathcal{K}}^-}](p)$ is linearly independent. Since $t + |\mathcal{K}| := \operatorname{rank}([h', H', G'_{\mathcal{J}^-(p)}](p))$, we have $|\hat{\mathcal{K}}| + |\hat{\mathcal{K}}^-| \leq |\mathcal{K}|$, which together with (4.17) implies that $|\bar{\mathcal{K}}| + |\bar{\mathcal{K}}^-| \leq |\hat{\mathcal{K}}| + |\hat{\mathcal{K}}^-| \leq |\mathcal{K}|$. Considering that $|\mathcal{K}| < |\bar{\mathcal{K}}| + |\bar{\mathcal{K}}^-|$, we have a contradiction. ∎

As a consequence of Theorems 2.2, 4.1, 4.4, and 4.5, we obtain that lower-RCPLD and lower-CRSC are constraint qualifications for problem (1.1). In particular, all other conditions shown in Figure 1 are also constraint qualifications. More specifically, Theorem 2.2 translated to problem (1.1) gives the following.

Corollary 4.1. *Suppose that $p \in \Omega$ satisfies lower-RCPLD or lower-CRSC. If p is a lower-AKKT point for problem* (1.1)*, then p is a KKT point for problem* (1.1)*.*

In [55], the capability of Algorithm 2.2 to generate AKKT sequences for problem (2.6) was demonstrated. This finding, supported by Theorem 4.1, indicates that Algorithm 2.2 also produces lower-AKKT sequences for the related problem (1.1). It is noteworthy that a comprehensive global convergence analysis of Algorithm 2.2, specifically designed to address problem (2.6), was presented in [5]. Given that problem (2.6) is essentially the Riemannian version of problem (1.1), Algorithm 2.2 is applicable to solving both problem instances. The next theorem establishes that, subject to any lower SCQ in Definition 3.2, any feasible limit point of the sequence generated by Algorithm 2.2 is a KKT point for problem (1.1). It is important to highlight that the sequence $(p^k)_{k\in\mathbb{N}}$ generated by Algorithm 2.2 is feasible for the constraint $\{q \in \mathbb{R}^n \mid h(q) = 0\}$.

Theorem 4.6. *Let $p \in \Omega$ be a limit point of a sequence $(p^k)_{k\in\mathbb{N}}$ generated by Algorithm* 2.2*. Assume that p satisfies lower-RCPLD or lower-CRSC. Then, p satisfies the KKT conditions for problem* (1.1)*.*

Proof. Let $p \in \Omega$ satisfy lower-RCPLD (respectively, lower-CRSC). By Theorem 4.4 (respectively, Theorem 4.5), it follows that $p \in \Omega_{\mathbb{M}}$ and also satisfies RCPLD (respectively, CRSC). According to [55, Theorem 3], p is an AKKT point for problem (2.6). Using Theorem 4.1, we conclude that p is a lower-AKKT point for problem (1.1). Therefore, by Corollary 4.1, p is a KKT point for problem (1.1). ∎

The next theorem establishes the connection between lower-QN and QN for problem (2.6).

Theorem 4.7. *A point $p \in \Omega$ satisfies lower-QN for problem* (1.1) *if and only if $p \in \Omega_{\mathbb{M}}$ satisfies QN for problem* (2.6)*.*

Proof. First, assume that $p \in \Omega$ satisfies lower-QN for problem (1.1). Assume, by contradiction, that $p \in \Omega_{\mathbb{M}}$ does not satisfies QN for problem (2.6). Since $\Omega = \Omega_{\mathbb{M}}$, we have $p \in \Omega_{\mathbb{M}}$, and item (i) of Definition 2.6 implies that there exist $\lambda \in \mathbb{R}^s$ and $\mu \in \mathbb{R}^m_+$ such that

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

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Given that $\mathcal{A}_{\mathbb{M}}(p)=\mathcal{A}(p)$, and using similar arguments as in the proof of Theorem 4.1, we can conclude that there exist $(\eta,\lambda,\mu)\in\mathbb{R}^t\times\mathbb{R}^s\times\mathbb{R}^m_+$ such that

$$\sum_{i=1}^{t}\eta_\ell h'_\ell(p)+\sum_{i=1}^{s}\lambda_i H'_i(p)+\sum_{j\in\mathcal{A}(p)}\mu_j G'_i(p)=0. \tag{4.18}$$

Additionally, considering item (ii) of Definition 2.6 and the fact that $\mathcal{A}_{\mathbb{M}}(p)=\mathcal{A}(p)$, we obtain

$$\mu_j=0\quad\forall j\notin\mathcal{A}(p),\quad(\eta,\lambda,\mu)\neq 0. \tag{4.19}$$

Finally, take $\epsilon>0$. It follows from item (iii) of Definition 2.6 that, for each $k\in\mathbb{N}$, there exists $q_k\in\mathbb{B}_{1/k}(p)$ such that $\lambda_i H_i(q_k)>0$ for all $i\in\mathcal{S}$ with $\lambda_i\neq 0$, and $\mu_j G_j(q_k)>0$ for all $j\in\mathcal{A}_{\mathbb{M}}(p)$ with $\mu_j>0$. Given that $\lim_{k\to+\infty}q_k=p$, there exists $\hat{k}\in\mathbb{N}$ such that $q_k\in B_\epsilon(p)\cap\{p\in\mathbb{R}^n\mid h(p)=0\}$ for all $k\geq\hat{k}$. Hence, considering that $\mathcal{A}_{\mathbb{M}}(p)=\mathcal{A}(p)$, it follows that there exists $q\in B_\epsilon(p)\cap\{q\in\mathbb{R}^n\mid h(q)=0\}$ such that $\lambda_i H_i(q)>0$ for all $i\in\mathcal{S}$ with $\lambda_i\neq 0$, and $\mu_j G_j(q)>0$ for all $j\in\mathcal{A}(p)$ with $\mu_j>0$. This, together with (4.18) and (4.19), contradicts Definition 3.4. Therefore, $p\in\Omega_{\mathbb{M}}$ satisfies QN for problem (2.6).

Reciprocally, suppose that $p\in\Omega_{\mathbb{M}}$ satisfies QN for problem (2.6). Assume, by contradiction, that $p\in\Omega$ does not satisfy lower-QN for problem (1.1). According to item (i) of Definition 3.4, there exist $\eta\in\mathbb{R}^t$, $\lambda\in\mathbb{R}^s$, and $\mu\in\mathbb{R}^m_+$ such that

$$\sum_{i=1}^{t}\eta_\ell h'_\ell(p)+\sum_{i=1}^{s}\lambda_i H'_i(p)+\sum_{j\in\mathcal{A}(p)}\mu_j G'_i(p)=0. \tag{4.20}$$

Given that $\mathcal{A}_{\mathbb{M}}(p)=\mathcal{A}(p)$, and using similar arguments as in the proof of Theorem 4.1, we can conclude that there exist $\lambda\in\mathbb{R}^s$ and $\mu\in\mathbb{R}^m_+$ such that

$$\sum_{i=1}^{s}\lambda_i\,\mathrm{grad}\,H_i(p)+\sum_{j\in\mathcal{A}_{\mathbb{M}}(p)}\mu_j\,\mathrm{grad}\,G_j(p)=0. \tag{4.21}$$

Additionally, item (ii) of Definition 3.4 implies that $\mu_j=0$ for all $j\notin\mathcal{A}(p)=\mathcal{A}_{\mathbb{M}}(p)$ and $(\eta,\lambda,\mu)\neq 0$. Moreover, since $\{h'_\ell(q)\mid\ell=1,\dots,t\}$ is linearly independent, (4.20) implies that if $\lambda=\mu=0$, then $\eta=0$, which means $(\eta,\lambda,\mu)=0$. Given that $(\eta,\lambda,\mu)\neq 0$ we have $(\lambda,\mu)\neq 0$. Thus,

$$\mu_j=0\quad\forall j\notin\mathcal{A}_{\mathbb{M}}(p),\quad(\lambda,\mu)\neq 0. \tag{4.22}$$

Now, take $\epsilon>0$. It follows from item (iii) of Definition 3.4 that for each $k\in\mathbb{N}$, there exist $q_k\in B_{1/k}(p)\cap\{q\in\mathbb{R}^n\mid h(q)=0\}$ such that $\lambda_i H_i(q_k)>0$ for all $i\in\mathcal{S}$ with $\lambda_i\neq 0$, and $\mu_j G_j(q_k)>0$ for all $j\in\mathcal{A}(p)$ with $\mu_j>0$. Since $\mathcal{A}_{\mathbb{M}}(p)=\mathcal{A}(p)$, $B_{1/k}(p)\cap\{q\in\mathbb{R}^n\mid h(q)=0\}$ is an open subset of $\mathbb{M}$, and $\lim_{k\to+\infty}q_k=p$, there exists $\bar{k}\in\mathbb{N}$ such that $q_{\bar{k}}\in\mathbb{B}_\epsilon(p)\cap B_{1/\bar{k}}(p)\cap\{q\in\mathbb{R}^n\mid h(q)=0\}$, with $\lambda_i H_i(q_{\bar{k}})>0$ for all $i\in\mathcal{S}$ with $\lambda_i\neq 0$, and $\mu_j G_j(q_{\bar{k}})>0$ for all $j\in\mathcal{A}_{\mathbb{M}}(p)$ with $\mu_j>0$. This, together with (4.21) and (4.22), contradicts Definition 2.6. Therefore, $p\in\Omega$ satisfies lower-QN for problem (1.1). ◻

THEOREM 4.8. *Let $p\in\Omega$ be a lower-PAKKT point with associated primal sequence $(p^k)_{k\in\mathbb{N}}$ and dual sequence $(\eta^k,\lambda^k,\mu^k)_{k\in\mathbb{N}}$. Assume that p satisfies lower-QN. Then, $(\eta^k,\lambda^k,\mu^k)_{k\in\mathbb{N}}$ is a bounded sequence. In particular, p satisfies the KKT conditions, and any limit point of $(\eta^k,\lambda^k,\mu^k)_{k\in\mathbb{N}}$ is a Lagrange multiplier associated with p.*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Proof. Using Theorems 2.3, 4.1, and 4.7, it follows that $(\lambda^k,\mu^k)_{k\in\mathbb{N}}$ is bounded. Assume, by contradiction, that (η^k) is unbounded. Let $K_1\subset\mathbb{N}$ be an infinite set and $\eta\in\mathbb{R}^t$ with $\|\eta\|_2=1$, such that $\lim_{k\in K_1}\|\eta^k\|_2=+\infty$ and $\lim_{k\in K_1}(\eta^k/\|\eta^k\|_2)=\eta$. Since $p\in\Omega_{\mathbb{M}}$ is lower-PAKKT with associated primal sequence $(p^k)_{k\in\mathbb{N}}$ and dual sequence $(\eta^k,\lambda^k,\mu^k)_{k\in\mathbb{N}}$, we have $\lim_{k\to\infty}L'(p^k,\eta^k,\lambda^k,\mu^k)=0$. Therefore,

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

Consequently, we obtain $\sum_{i=1}^{t}\eta_\ell h_\ell(p)=0$ with $\eta\neq 0$, contradicting assumption (H1). Thus, we conclude that $(\eta^k,\lambda^k,\mu^k)_{k\in\mathbb{N}}$ is a bounded sequence. ∎

Next, we show that Algorithm 2.2 produces lower-PAKKT sequences, thereby establishing its global convergence under the lower-QN condition.

THEOREM 4.9. *Assume that Algorithm* 2.2 *generates an infinite sequence* $(p^k)_{k\in\mathbb{N}}$ *with a feasible limit* $p\in\Omega$*, say,* $\lim_{k\in K}p^k=p$*. Then,* p *is a lower-PAKKT point with the correspondent primal sequence* $(p^k)_{k\in K}$ *and dual sequence* $(\eta^k,\lambda^k,\mu^k)_{k\in K}$*, where* $(\eta^k)_{k\in K}$ *can be determined from* $(\lambda^k,\mu^k)_{k\in K}$ *which is generated by the algorithm. In particular,* p *is a KKT point, and any limit point of* $(\eta^k,\lambda^k,\mu^k)_{k\in K}$ *is a Lagrange multiplier associated with* p*.*

Proof. By Theorem 2.4, we obtain that p is a PAKKT point with the corresponding primal sequence $(p^k)_{k\in K}$ and dual sequence $(\lambda^k,\mu^k)_{k\in K}$ as generated by Algorithm 2.2. Using Theorem 4.1, we conclude that p is a lower-PAKKT point. Furthermore, similarly to the proof of Theorem 4.1, we find that $(\eta^k)_{k\in K}$ can be determined from $(\lambda^k,\mu^k)_{k\in K}$. ∎

In [5], the authors introduced the concept of scaled-PAKKT point for problem (2.6). Essentially, it coincides with the definition of PAKKT with the condition $\lim_{k\to\infty}\operatorname{grad}\mathbb{L}(p^k,\lambda^k,\mu^k)/\gamma_k=0$, where $\gamma_k:=\|(1,\lambda^k,\mu^k)\|_\infty$, replacing item (ii) in Definition 2.4. Similarly to lower-PAKKT, we can also define the concept of *lower-scaled-PAKKT* and establish its connection with scaled-PAKKT. It was demonstrated in [5] that Algorithm 2.2 is capable of generating scaled-PAKKT sequences by ensuring $\|\operatorname{grad}\mathbb{L}_{\rho_k}(p^k,\bar\lambda^k,\bar\mu^k)/\gamma_k\|\leq\epsilon_k$, rather than (2.9) in Step 1. Moreover, QN ensures that the dual scaled-PAKKT sequence is bounded. Therefore, Theorems 4.1 and 4.7 imply that lower-QN is sufficient to guarantee that the dual lower-scaled-PAKKT sequence is bounded.

We conclude this section by emphasizing that the equivalence established between lower-SCQs and Riemannian-SCQs offers important theoretical insights, clarifying and extending classical constraint qualifications in nonlinear programming. This equivalence explicitly shows how intrinsic Riemannian geometry conditions naturally translate into weaker constraint qualifications applicable in the Euclidean setting. Such insights enable the development of optimization methods capable of addressing constrained problems that might otherwise pose challenges within traditional Euclidean frameworks. The numerical experiments discussed in the next section illustrate these practical advantages, further demonstrating the benefits of incorporating intrinsic Riemannian ideas into classical Euclidean optimization methodologies.

5. Numerical experiments. This section presents numerical results to illustrate the practical advantages of exploiting Riemannian techniques in solving certain classes of optimization problems. The experiments were conducted using MATLAB version 9.11.0 (R2021b) on a computer with a 3.7 GHz Intel Core i5 6-Core processor

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

and 8 GB 2667 MHz DDR4 RAM, running macOS Sonoma 14.4.1. All codes are available at https://github.com/lfprudente/RiemannianAL. We compare the performance of the Riemannian and Euclidean safeguarded augmented Lagrangian methods as follows:

- *Riemannian-AL* (Riemannian augmented Lagrangian)*:* Algorithm 2.2 with Manopt [24] to solve the subproblem in Step 1. We use the RLBFGS solver, a Riemannian limited memory BFGS algorithm [36].
- *Euclidean-AL* (Euclidean augmented Lagrangian)*:* Algorithm 2.1 with ASA [33] to solve the subproblem in Step 1. In the Euclidean version, we treat the box constraints ($l \le q \le u$, for given vectors $l, u \in \mathbb{R}^n$) as *lower-level constraints,* i.e., they are not penalized. This is computationally more efficient because ASA, an active-set code in C for box-constrained optimization, already handles such bounds directly via gradient projection [28] and the preconditioned CG_DESCENT algorithm (with an L-BFGS–type preconditioner) [32, 34], both included in ASA. We use the MATLAB interface provided in [14].

In contrast to the Euclidean version, where box constraints are naturally handled by the subproblem solver, incorporating bounds directly on a Riemannian manifold (defined by $h(q) = 0$) is not straightforward. Consequently, in the Riemannian approach we penalize any such constraints.

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

$$\|\operatorname{grad} \mathbb{L}(p^k, \lambda^k, \mu^k)\| \le \varepsilon_{\rm opt} \quad \text{(respectively, } \|L'(p^k, \eta^k, \lambda^k, \mu^k)\| \le \varepsilon_{\rm opt}),$$

$$\max\{\|H(p^k)\|_\infty, \|G(p^k)_+\|_\infty\} \le \varepsilon_{\rm feas} \quad \text{(respectively, } \max\{\|h(p^k)\|_\infty, \|H(p^k)\|_\infty, \|G(p^k)_+\|_\infty\} \le \varepsilon_{\rm feas}),$$

$$\min\{-G_i(p^k), \mu^k_\ell\} \le \varepsilon_{\rm compl} \quad \text{for all} \quad i = 1, \dots, m.$$

These conditions correspond to the approximate fulfillment of the KKT conditions. We also stopped the execution of the algorithms if the penalty parameter became too large ($\rho_k > 10^{20}$) or if the algorithms exceeded the maximum number of 50 outer iterations allowed. These stopping criteria are practical safeguards: an excessively large penalty parameter is typically a symptom of either infeasibility or severe numerical difficulties, while exceeding the maximum iterations indicates that the algorithm is not converging in a reasonable amount of time.

For both algorithms, we used the following parameters: $\tau = 0.5$, $\gamma = 10$, $\lambda_{\min} = -10^{20}$, $\lambda_{\max} = \mu_{\max} = 10^{20}$, $\bar\lambda^1 = \bar\mu^1 = 0$, $\varepsilon_{\rm opt} = \varepsilon_{\rm feas} = \varepsilon_{\rm compl} = 10^{-5}$, $\varepsilon_k = \max\{\varepsilon_{\rm opt}, \sqrt{\varepsilon_{\rm opt}}/10^{k-1}\}$ for all $k \ge 1$, and

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

as suggested in [4, 20]. Our choice for the parameters $\lambda_{\min}$, $\lambda_{\max}$, and $\mu_{\max}$ is made to avoid interfering with the natural scaling of the multipliers, ensuring that the box for the safeguarded multiplier estimates is large enough to contain the "true" Lagrange multipliers. Moreover, the adaptive strategy to set ρ_1 balances the goals of driving the iterates toward feasibility and making progress in reducing the objective. For each experiment, both solvers used the same starting point.

In the following subsections, we describe three applications that can be modeled as problem (2.6), along with the corresponding results. Although we report CPU times

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

in our experiments, these values should be interpreted with caution. Indeed, each version (Riemannian versus Euclidean) employs a different internal solver (written in different languages), chosen to give the best possible performance in its respective setting. Therefore, the CPU times do not provide a strict one-to-one comparison. Instead, the primary focus of these experiments is on the solution quality and the robustness of each approach.

5.1. Greediness phenomenon. The *greediness* phenomenon is the tendency of some nonlinear programming methods to find highly infeasible points with very small functional values, typically in the initial iterations; see [26]. The phenomenon can occur in problems for which the objective function assumes significantly lower values at infeasible points than in the feasible region. In an augmented Lagrangian algorithm, unconstrained minimizers may attract the iterates at early stages of the calculations, causing the penalty parameter to grow excessively. This excessive growth leads to ill-conditioning, which harms the overall convergence. Consider the following example [26]:

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad -\sum_{i=1}^{n}(x_i^8+x_i) \quad \text{subject to} \quad \|x\|_2^2=1, \quad x_2+\sum_{i=1}^{n} x_i \le 0. \tag{5.1}$$

Since $\|x\|^2=1$ defines the unit sphere $\mathbb{S}^{n-1}\subset\mathbb{R}^n$, problem (5.1) can be rewritten as

$$\underset{x\in\mathbb{S}^{n-1}}{\text{minimize}} \quad -\sum_{i=1}^{n}(x_i^8+x_i) \quad \text{subject to} \quad x_2+\sum_{i=1}^{n} x_i \le 0. \tag{5.2}$$

We applied Euclidean-AL and Riemannian-AL algorithms to problems (5.1) and (5.2), respectively. We set $n = 50$ and randomly generated 100 starting points on the sphere $\mathbb{S}^{n-1}$. The Euclidean-AL algorithm failed in all instances, typically generating the first iterate with $\|x^1\|_\infty \approx 10^{43}$ and $f(x^1) = -\infty$. From there, two situations were observed: either NaNs were generated, resulting in algorithm failure, or the penalty parameter grew beyond the maximum allowed. In contrast, the Riemannian-AL algorithm successfully solved the problem in all instances. The maximum number of iterations was 6, and the greatest final penalty parameter was less than 20. For illustrative purposes, Figure 2 shows the behavior of the Riemannian-AL algorithm in the case where $n=2$, $x^0=(\sqrt{2}/2,\sqrt{2}/2)$, and $\rho_1\approx 7$. The Riemannian-AL algorithm converged to the global solution $x^*=(2\sqrt{5}/5,-\sqrt{5}/5)$ in 4 iterations. Visually, the second iterate x^2 is virtually identical to the solution x^*. In contrast, as for the larger instances (i.e., when $n=50$), the Euclidean-AL algorithm failed to solve the problem.

We conclude that in the presence of greediness, by exploiting the intrinsic geometry of the manifold, the Riemannian-AL algorithm effectively maintains feasibility, controls the penalty parameter, and successfully overcomes the phenomenon. In contrast, the Euclidean-AL method fails to handle this issue, leading to divergent iterates and an unbounded penalty parameter.

5.2. Packing circles within ellipses. The circle packing problem involves finding the maximum radius r of N identical circles that can be fitted without overlapping into a two-dimensional fixed-size container [42]. This problem has a wide range of applications, as discussed in [42, 18] and the references therein. In this section, we consider the container to be an ellipse with semiaxes $a \ge b > 0$. Using continuous variables $(r;u,v,s)\in\mathbb{R}\times(\mathbb{R}^N)^3$, this problem can be modeled [18] as follows:

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

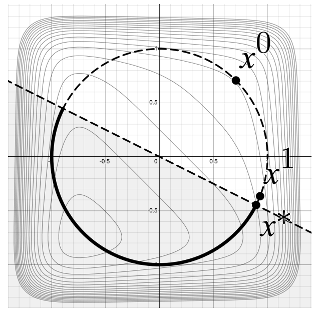

Fig. 2. *Iterates generated by the Riemannian-AL algorithm for problem* (5.2) *with* $n=2$. *The feasible set is highlighted in bold, and the level sets of the objective function are shown in light gray. The algorithm converges in* 4 *iterations, with the second iterate* x^2 *closely coinciding with the global solution* x^*.

$$
\begin{array}{lll}
 & \underset{(r;u,v,s)\in\mathbb{R}\times(\mathbb{R}^N)^3}{\text{maximize}} & r \\
 & \text{subject to} & u_i^2+v_i^2=1 \quad \forall i=1,\ldots,N, \\
 & & b^2(s_i-1)^2[(b^2/a^2)u_i^2+v_i^2]\ge r^2 \quad \forall i=1,\ldots,N, \\
(5.3) & & a^2\left[(1+(s_i-1)(b^2/a^2))u_i-(1+(s_j-1)(b^2/a^2))u_j\right]^2 \\
 & & +b^2(s_iv_i-s_jv_i)^2\ge 4r^2 \quad \forall i<j, \\
 & & 0\le s_i\le 1 \quad \forall i=1,\ldots,N, \\
 & & r\ge 0.
\end{array}
$$

In this formulation, the variables $u,v,s\in\mathbb{R}^N$ represent ellipse-based coordinates associated with the centers of the N circles: for each $i=1,\ldots,N$, the pair (u_i,v_i) defines a direction on the unit circle, resulting from a change of variables that maps the ellipse into a unit circle, and $s_i\in[0,1]$ is a scaling factor that determines how far along this direction the center of the ith circle is placed. The first constraint ensures that (u_i,v_i) lies on the unit circle; the second ensures that the ith circle remains inside the ellipse; the third prevents overlapping between any two circles; and the last two define valid ranges for s_i and r. For further modeling details, we refer to [18].

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

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

Since, for all $i=1,\ldots,N$, the constraint $u_i^2+v_i^2=1$ defines the unit circle $\mathbb{S}^1\subset\mathbb{R}^2$, problem (5.3) can be written in the format of (2.6) by taking $(r;(u,v);s)\in\mathbb{M}:=\mathbb{R}\times(\mathbb{S}^1)^N\times\mathbb{R}^N$, thereby omitting the first group of constraints (i.e., those enforcing $u_i^2+v_i^2=1$, which are now inherent to the definition of $\mathbb{S}^1$). The resulting problem is as follows:

$$
\begin{array}{ll}
\underset{(r;u,v,s)\in\mathbb{M}}{\text{maximize}} & r \\
\text{subject to} & b^2(s_i-1)^2[(b^2/a^2)u_i^2+v_i^2]\ge r^2 \quad \forall i=1,\ldots,N, \\
 & a^2\left[(1+(s_i-1)(b^2/a^2))u_i-(1+(s_j-1)(b^2/a^2))u_j\right]^2 \\
 & +b^2(s_iv_i-s_jv_i)^2\ge 4r^2 \quad \forall i<j, \\
 & 0\le s_i\le 1 \quad \forall i=1,\ldots,N, \\
 & r\ge 0.
\end{array}
$$

We considered eight instances of the problems with $(a,b)=(2,1)$ and $N\in\{3,10,20,30,40,50,80,100\}$. The starting points $(r^0;(u^0,v^0);s^0)$ were randomly generated in $[0,1]\times(\mathbb{S}^1)^N\times[0,1]^N$. Table 1 shows the results. In the table, "N" is the

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

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

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

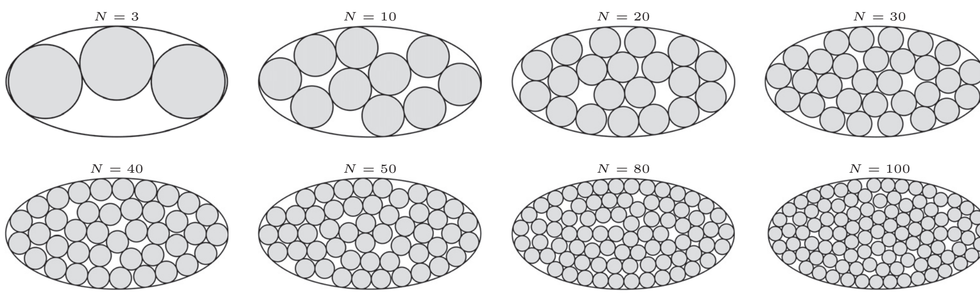

FIG. 3. *Illustration of the solutions found by the Riemannian-AL algorithm for packing N circles within an ellipse with* $(a,b) = (2,1)$.

number of circles to be packed, "n" denotes the total number of decision variables (i.e., r, (u,v), and s), and "#c." is the number of constraints. For the Riemannian-AL algorithm, the manifold constraints are excluded from the count, and for the Euclidean-AL algorithm, the box constraints are excluded. "It" is the number of augmented Lagrangian iterations, "Obj." is the final optimal radius, "Feas." is the feasibility measure at the final iterate p^* given by $\max\{\|h(p^*)\|_\infty, \|H(p^*)\|_\infty, \|G(p^*)_+\|_\infty\}$, and "Time" is the CPU time in seconds. The best reported datum for each instance is highlighted in bold. Figure 3 illustrates the "solutions" found by the Riemannian-AL algorithm.

As can be seen, in most instances the Riemannian-AL algorithm outperformed the Euclidean-AL algorithm in terms of the final optimal radius. The results indicate that for instances involving a smaller number of circles the Euclidean-AL algorithm tends to be faster, and the final optimal radius r differs only slightly from that obtained by the Riemannian-AL algorithm. However, as the number of circles increases, the Riemannian-AL algorithm demonstrates a clear advantage in achieving a larger final value of r. For larger instances, the Euclidean-AL algorithm often encountered issues where the centers of some circles moved to the ellipse boundary, resulting in a significantly smaller radius (r in the order of 10^{-3}). This phenomenon led to poor local minimizers and was never observed with the Riemannian-AL algorithm. For example, when packing 20 circles, the Euclidean-AL algorithm required 15 different starting points, taking a total of 183.1 s to find a solution with r in the order of 10^{-1}. Similarly, for packing 40 circles, the Euclidean-AL algorithm needed 7 initial points, taking a total of 1060.7 s. Notably, for packing 80 circles, the Euclidean-AL algorithm

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

used 100 starting points and spent over 13 hours without finding a solution with a final radius r in the order of 10^{-1}. These results show that while the Riemannian-AL algorithm efficiently solves all instances without needing multiple starting points, the Euclidean-AL algorithm frequently requires several initial points to find good solutions, resulting in more computational time.

5.3. *k*-Means clustering. Given a set of N data points, the k-means clustering problem involves partitioning these points into k clusters, with the goal of minimizing the sum of squared distances between each data point and the centroid of its corresponding cluster. This process helps in grouping similar data points together, uncovering underlying patterns or structures within the data. We refer to [3] for numerous applications across various domains. Let $\mathcal{P} := \{x_1, \ldots, x_N\} \subset \mathbb{R}^\delta$ represent the given data points, and denote the clusters by $\mathcal{C}_1, \ldots, \mathcal{C}_k \subset \mathcal{P}$. Here, δ denotes the number of features of each x_i. The k-means clustering problem can be formulated as follows,

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

where $\mu_j := \frac{1}{|\mathcal{C}_j|} \sum_{x_i \in \mathcal{C}_j} x_i$ and $|\mathcal{C}_j|$ is the cardinality of $\mathcal{C}_j$. In this formulation, the decision variables are indeed the clusters themselves, i.e., the subsets $\mathcal{C}_1, \ldots, \mathcal{C}_k$ that partition the data set. According to [25], this problem can be equivalently reformulated as a continuous optimization problem with nonnegative orthogonality constraints, expressed as follows,

$$\underset{Y \in \mathbb{R}^{N \times k}}{\text{minimize}} \quad -\operatorname{tr}(Y^\top D Y) \quad \text{subject to} \quad Y^\top Y = I, \quad Y \geq 0, \quad YY^\top e = e, \tag{5.4}$$

where $D := (D_{ij}) \in \mathbb{R}^{N \times N}$ with $D_{ij} = x_i^\top x_j$ for all $i, j = 1, \ldots, N$, I is the k-dimensional identity matrix, the inequality $Y \geq 0$ is componentwise, and $e \in \mathbb{R}^N$ is the vector of ones. The constraint $Y^\top Y = I$ is the Stiefel manifold embedded in the $N \times k$ real matrix space, denoted by $St_{N,k} := \{Y \in \mathbb{R}^{N \times k} \mid Y^\top Y = I\}$. Thus, problem (5.4) can be written in the format of (2.6) by taking $Y \in St_{N,k}$ and omitting the constraint $Y^\top Y = I$. A feasible point $Y \in \mathbb{R}^{N \times k}$ has exactly one nonnegative entry per row and all nonzero entries of a column are equal. A point x_i is assigned to cluster $\mathcal{C}_j$ if $Y_{ij} \neq 0$. This property ensures that the clustering structure can be directly recovered from the solution matrix Y.

As in [41], we considered some datasets from the UCI Machine Learning Repository [39]. The main characteristics of the considered problems are described in Table 2. The table also shows the number of variables related to problem (5.4), and the number of constraints of its Riemannian (#c(Riem.)) and Euclidean (#c(Eucl.)) versions. The starting points were generated corresponding to a random cluster classification. For all k-means problems, we used $\varepsilon_{\text{opt}} = 10^{-4}$ in the stopping criterion, since the tighter value $\varepsilon_{\text{opt}} = 10^{-5}$ often caused Manopt's RLBFGS solver to reach the minimum (inner) step size allowed, degrading the overall convergence.

Table 3 shows the results organized similarly to Table 1, with the addition of the "Acc." column, which reports the accuracy of correct classifications. The results indicate that the Riemannian-AL algorithm consistently produces higher clustering accuracy (Acc.) compared to the Euclidean-AL algorithm. Despite generally requiring more computational time, the Riemannian-AL algorithm achieves lower objective

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

TABLE 2
Main characteristics of the considered k-means problems.

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

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

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

values (Obj.) and maintains a high level of feasibility (Feas.). The Euclidean-AL algorithm often faced convergence difficulties and failed to provide results for several datasets. Specifically, for the Cloud, Iris, Pima diabetes, Raisin, and SPECTF problems, the Euclidean-AL algorithm got stuck at infeasible points that are stationary for an infeasibility measure (see [20, Theorem 6.3]), leading to excessive increases in the penalty parameter until the algorithm failed. The convergence to such infeasible points impairs the overall performance of an augmented Lagrangian algorithm and is often used as a stopping criterion related to failure; see [20, section 10.2.3].

We conclude this section by illustrating the solution found by the Riemannian-AL algorithm in a synthetic k-means problem with $\delta = 2$. We generated 500 randomly perturbed data points around the reference points $(3, 3)$, $(-3, -3)$, and $(6, -6)$, resulting in a problem with 1500 variables and 2000 constraints. The Riemannian-AL algorithm took 8 iterations and a total of 103.3 s to find a solution with Obj $= -9.919 \times 10^3$ and Feas $= 1.7 \times 10^{-8}$. Figure 4 visually demonstrates the algorithm's effectiveness, as it correctly clusters 92.2% of the data points when compared with the original reference points.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

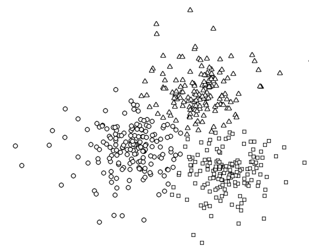

FIG. 4. *Illustration of the solution found by the Riemannian-AL algorithm in a synthetic k-means problem with* $\delta = 2$.

6. Conclusions. This paper examines the use of augmented Lagrangian methods for solving constrained nonlinear programming problems involving both equality and inequality constraints, with an emphasis on lower and upper-level constraints. We introduce and analyze lower-SCQs within this framework, demonstrating their reduced restrictiveness compared to traditional constraint qualifications. By applying Riemannian geometry, we connect lower-SCQs with their Riemannian counterparts, thereby enhancing the theoretical foundation of optimization on Riemannian manifolds. Our study reveals significant theoretical advancements and practical implications, including the introduction of new sequential optimality conditions in the safeguarded augmented Lagrangian algorithm. This algorithm generates lower-PAKKT sequences that adhere to manifold constraints and ensure all limit points satisfy the KKT conditions under any lower-SCQ. This finding highlights the effectiveness of our framework and the advantages of Riemannian optimization methods. The comparison between classic and Riemannian versions of the algorithm reveals that the intrinsic approach often outperforms the extrinsic one in certain cases. This advocates for the use of Riemannian methods in specific optimization problems and suggests new research directions beyond Euclidean spaces. In conclusion, this work advances both the theory and practical applications of nonlinear programming, emphasizing the dynamic nature of optimization research and encouraging further investigation into Riemannian methods across various theoretical and practical settings. Our analysis is confined to finite-dimensional constrained optimization; extending these results to infinite-dimensional Riemannian manifolds would require additional structural assumptions and represents a promising avenue for future research; see [22, 49] for preliminary approaches.

REFERENCES

[1] J. ABADIE, *On the Kuhn-Tucker theorem*, in Nonlinear Programming (NATO Summer School, Menton, 1964), North-Holland, Amsterdam, 1967, pp. 19–36.

[2] P.-A. ABSIL, R. MAHONY, AND R. SEPULCHRE, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ, 2008.

[3] M. AHMED, R. SERAJ, AND S. M. S. ISLAM, *The k-means algorithm: A comprehensive survey and performance evaluation*, Electronics, 9 (2020), 1295, https://doi.org/10.3390/electronics9081295.

[4] 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 (2008), pp. 1286–1309, https://doi.org/10.1137/060654797.

[5] R. ANDREANI, K. R. COUTO, O. P. FERREIRA, AND G. HAESER, *Constraint qualifications and strong global convergence properties of an augmented Lagrangian method on Riemannian manifolds*, SIAM J. Optim., 34 (2024), pp. 1799–1825, https://doi.org/10.1137/23M1582382.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[6] 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, https://doi.org/10.1137/17M1147330.

[7] R. Andreani, G. Haeser, and J. M. Martínez, *On sequential optimality conditions for smooth constrained optimization*, Optimization, 60 (2011), pp. 627–641, https://doi.org/10.1080/02331930903578700.

[8] 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, https://doi.org/10.1007/s12532-021-00207-9.

[9] 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, https://doi.org/10.1137/110843939.

[10] 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, https://doi.org/10.1007/s10107-011-0456-0.

[11] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva, *Strict constraint qualifications and sequential optimality conditions for constrained optimization*, Math. Oper. Res., 43 (2018), pp. 693–717, https://doi.org/10.1287/moor.2017.0879.

[12] 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, https://doi.org/10.1007/s10957-004-1861-9.

[13] M. S. Bazaraa and C. M. Shetty, *Nonlinear Programming*, Wiley, New York, 1979.

[14] S. Becker, *mex Interface for Bound Constrained Optimization via ASA*, https://www.mathworks.com/matlabcentral/fileexchange/35814-mex-interface-for-bound-constrained-optimization-via-asa (2024).

[15] R. Bergmann and R. Herzog, *Intrinsic formulation of KKT conditions and constraint qualifications on smooth manifolds*, SIAM J. Optim., 29 (2019), pp. 2423–2444, https://doi.org/10.1137/18M1181602.

[16] D. P. Bertsekas, *Nonlinear Programming*, 3rd ed., Athena Sci. Optim. Comput. Ser., Athena Scientific, Belmont, MA, 2016.

[17] D. P. Bertsekas and A. E. Ozdaglar, *Pseudonormality and a Lagrange multiplier theory for constrained optimization*, J. Optim. Theory Appl., 114 (2002), pp. 287–343, https://doi.org/10.1023/A:1016083601322.

[18] E. G. Birgin, L. H. Bustamante, H. F. Callisaya, and J. M. Martínez, *Packing circles within ellipses*, Int. Trans. Oper. Res., 20 (2013), pp. 365–389, https://doi.org/10.1111/itor.12006.

[19] E. G. Birgin, G. Haeser, and A. Ramos, *Augmented Lagrangians with constrained subproblems and convergence to second-order stationary points*, Comput. Optim. Appl., 69 (2018), pp. 51–75, https://doi.org/10.1007/s10589-017-9937-2.

[20] E. G. Birgin and J. M. Martínez, *Practical Augmented Lagrangian Methods for Constrained Optimization*, Fundamentals of Algorithms 10, SIAM, Philadelphia, 2014.

[21] E. G. Birgin and J. M. Martínez, *Complexity and performance of an augmented Lagrangian algorithm*, Optim. Methods Softw., 35 (2020), pp. 885–920, https://doi.org/10.1080/10556788.2020.1746962.

[22] 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. Optim., 30 (2020), pp. 2956–2982, https://doi.org/10.1137/19M1306804.

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

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

[25] T. Carson, D. G. Mixon, S. Villar, and R. Ward, *Manifold optimization for k-means clustering*, in 2017 International Conference on Sampling Theory and Applications (SampTA), IEEE, Piscataway, NJ, 2017, pp. 73–77.

[26] E. V. Castelani, A. L. M. Martinez, J. M. Martínez, and B. F. Svaiter, *Addressing the greediness phenomenon in nonlinear programming by means of proximal augmented Lagrangians*, Comput. Optim. Appl., 46 (2010), pp. 229–245, https://doi.org/10.1007/s10589-009-9271-4.

[27] J. J. Cochran, L. A. Cox Jr., P. Keskinocak, J. P. Kharoufeh, J. C. Smith, and M. V. Solodov, *Constraint qualifications in Wiley Encyclopedia of Operations Research and Management Science*, Wiley, New York, 2010.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

[28] Y.-H. Dai, W. W. Hager, K. Schittkowski, and H. Zhang, *The cyclic Barzilai–Borwein method for unconstrained optimization*, IMA J. Numer. Anal., 26 (07 2006), pp. 604–627, https://doi.org/10.1093/imanum/drl006.

[29] K. Deng, J. Hu, J. Wu, and Z. Wen, *Oracle complexities of Augmented Lagrangian methods for nonsmooth manifold optimization*, Math. Oper. Res., (2025), https://doi.org/10.1287/moor.2024.0498.

[30] J. Gauvin, *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming*, Math. Program., 12 (1977), pp. 136–138, https://doi.org/10.1007/BF01593777.

[31] L. Guo and J. Ye, *Necessary optimality conditions and exact penalization for non-Lipschitz nonlinear programs*, Math. Program., 168 (2018), pp. 571–598, https://doi.org/10.1007/s10107-017-1112-0.

[32] W. W. Hager and H. Zhang, *A new conjugate gradient method with guaranteed descent and an efficient line search*, SIAM J. Optim., 16 (2005), pp. 170–192, https://doi.org/10.1137/030601880.

[33] W. W. Hager and H. Zhang, *A new active set algorithm for box constrained optimization*, SIAM J. Optim., 17 (2006), pp. 526–557, https://doi.org/10.1137/050635225.

[34] W. W. Hager and H. Zhang, *The limited memory conjugate gradient method*, SIAM J. Optim., 23 (2013), pp. 2150–2168, https://doi.org/10.1137/120898097.

[35] M. R. Hestenes, *Optimization Theory: The Finite Dimensional Case*, Pure Appl. Math., Wiley Interscience, New York, 1975.

[36] W. Huang, K. A. Gallivan, and P.-A. Absil, *A Broyden class of quasi-Newton methods for Riemannian optimization*, SIAM J. Optim., 25 (2015), pp. 1660–1685, https://doi.org/10.1137/140955483.

[37] R. Janin, *Directional derivative of the marginal function in nonlinear programming*, in Sensitivity, Stability and Parametric Analysis, Math. Program. Stud. 21, Springer, Berlin, 1984, pp. 110–126, https://doi.org/10.1007/BFb0121214.

[38] C. Kanzow and D. Steck, *An example comparing the standard and safeguarded augmented Lagrangian methods*, Oper. Res. Lett., 45 (2017), pp. 598–603, https://doi.org/10.1016/j.orl.2017.09.005.

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

[40] 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, https://doi.org/10.1007/s10957-021-01879-y.

[41] C. Liu and N. Boumal, *Simple algorithms for optimization on Riemannian manifolds with constraints*, Appl. Math. Optim., 82 (2020), pp. 949–981, https://doi.org/10.1007/s00245-019-09564-3.

[42] C. López and J. Beasley, *A heuristic for the circle packing problem with a variety of containers*, European J. Oper. Res., 214 (2011), pp. 512–525, https://doi.org/10.1016/j.ejor.2011.04.024.

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

[44] 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, https://doi.org/10.1007/s11228-021-00578-0.

[45] L. Minchenko and S. Stakhovski, *On relaxed constant rank regularity condition in mathematical programming*, Optimization, 60 (2011), pp. 429–440, https://doi.org/10.1080/02331930902971377.

[46] A. E. Ozdaglar and D. P. Bertsekas, *The relation between pseudonormality and quasiregularity in constrained optimization*, Optim. Methods Softw., 19 (2004), pp. 493–506, https://doi.org/10.1080/10556780410001709420.

[47] 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, https://doi.org/10.1137/S1052623497326629.

[48] S. M. Robinson, *Stability theory for systems of inequalities, part* II*: Differentiable nonlinear systems*, SIAM J. Numer. Anal., 13 (1976), pp. 497–513, https://doi.org/10.1137/0713043.

[49] A. Schiela and J. Ortiz, *An SQP method for equality constrained optimization on Hilbert manifolds*, SIAM J. Optim., 31 (2021), pp. 2255–2284, https://doi.org/10.1137/20M1341325.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[50] M. Slater, *Lagrange multipliers revisited*, in Traces and Emergence of Nonlinear Programming, Birkhäuser, Basel, 2014, pp. 293–306.

[51] G. Wachsmuth, *On LICQ and the uniqueness of Lagrange multipliers*, Oper. Res. Lett., 41 (2013), pp. 78–80, https://doi.org/10.1016/j.orl.2012.11.009.

[52] Y. Wang, K. Deng, H. Liu, and Z. Wen, *A decomposition augmented Lagrangian method for low-rank semidefinite programming*, SIAM J. Optim., 33 (2023), pp. 1361–1390, https://doi.org/10.1137/22M1474539.

[53] M. Xu, B. Jiang, Y. Liu, and A. M.-C. So, *On the oracle complexity of a Riemannian inexact augmented Lagrangian method for nonsmooth composite problems over Riemannian submanifolds*, Optim. Lett., (2025), pp. 1–19, https://doi.org/10.1007/s11590-025-02210-8.

[54] 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, https://doi.org/10.1007/s10898-020-00907-x.

[55] 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, https://doi.org/10.1007/s10589-021-00336-w.

[56] 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.

[57] Y. Zhou, C. Bao, C. Ding, and J. Zhu, *A semismooth Newton based augmented Lagrangian method for nonsmooth optimization on matrix manifolds*, Math. Program., 201 (2023), pp. 1–61, https://doi.org/10.1007/s10107-022-01898-1.

Downloaded 03/19/26 to 200.137.198.84 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.