Noname manuscript No.
(will be inserted by the editor)

On some perturbation properties of nonsmooth optimization on Riemannian manifolds with applications

Yuexin Zhou · Chao Ding · Yangjing Zhang

This version: Sep 30, 2025

Abstract This paper presents a perturbation analysis framework for nonsmooth optimization on connected Riemannian manifolds to bridge the gap between the rapid development of algorithmic approaches and a robust theoretical foundation. Using tangent-space local models, we transport core notions from Euclidean variational analysis, such as strong regularity, the Aubin property, and isolated calmness of the Karush–Kuhn–Tucker (KKT) solution mapping, to the manifold setting. Furthermore, we introduce the manifold (strong) variational sufficiency and show that its strong version is intrinsic, i.e., independent of the chosen retraction, and for polyhedral, second-order cone, and semidefinite programs, it coincides with the manifold strong second-order sufficient condition. These insights yield concrete algorithmic consequences. We show that the Riemannian sequential quadratic programming achieves local superlinear and, under mild additional assumptions, quadratic convergence without strict complementarity, while the Riemannian Augmented Lagrangian Method attains R-linear convergence even when Lagrange multipliers are nonunique. Moreover, the proposed condition guarantees positive definiteness of the generalized Hessians associated with the augmented Lagrangian, enabling superlinear semismooth Newton steps in inner solves. Numerical experiments on robust matrix completion and compressed modes validate the theoretical predictions.

Keywords nonsmooth optimizations · Riemannian manifold · perturbation analysis · local convergence rate

Mathematics Subject Classification (2010) 90C30 · 90C46 · 49J52 · 65K05

Y.X. Zhou
Engineering Systems and Design, Singapore University of Technology and Design, Singapore, Singapore.
Part of this work was completed while the author was with the Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, P.R. China, School of Mathematical Sciences, University of Chinese Academy of Science, Beijing, P.R. China.
E-mail: yuexin_zhou@sutd.edu.sg

C. Ding
Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, P.R. China.
E-mail: dingchao@amss.ac.cn

Y.J. Zhang
Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, P.R. China.
E-mail: yangjing.zhang@amss.ac.cn

1 Introduction

Consider the following nonsmooth optimization problems on Riemannian manifolds

$$\begin{array}{ll} \min & f(x) + \theta(g(x)) \\ \text{s.t.} & x \in \mathcal{M}, \end{array} \tag{1.1}$$

where $\mathcal{M}$ is a connected Riemannian manifold, $f : \mathcal{M} \to \mathbb{R}$ and $g : \mathcal{M} \to \mathbb{Y}$ are continuously differentiable functions, $\mathbb{Y}$ is a Euclidean space equipped with a scalar product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$, $\theta : \mathbb{Y} \to (-\infty, \infty]$ is a proper closed convex function. If θ is an indicator function of a closed convex set, then (1.1) is a constrained manifold optimization problem. Applications of (1.1) arise in various scenarios such as principal component analysis problems [65], low-rank matrix completion problems [13], orthogonal dictionary learning problems [19, 57], and compressed modes problems [45]. We refer readers to [1, 32] for further details.

Various algorithms have been developed to solve the manifold nonsmooth optimization problem (1.1), including subgradient methods [24, 26], proximal gradient methods [17, 34, 33], alternating direction methods of multipliers [36, 37], and proximal point algorithm (PPA) [16, 25]. More recently, the augmented Lagrangian method (ALM) has been extended to nonsmooth manifold optimization in [20, 64], further enriching the algorithmic landscape. In contrast with these advances on the computational side, the underlying theory remains relatively underdeveloped. In particular, fundamental aspects such as perturbation analysis have yet to be systematically explored for nonsmooth manifold problems, even though they are essential for understanding solution stability and for proving rigorous convergence guarantees.

Perturbation analysis is a cornerstone of optimization theory, as it offers critical insights into the stability of optimal solutions under small changes in the problem data. In the Euclidean setting, this topic has been systematically developed within an even broader framework of variational analysis. Among these, a significant line of research in the literature focuses on the regularity properties of Karush-Kuhn-Tucker (KKT) solution mappings, which describe how the set of stationary points changes under perturbations. The comprehensive treatise by Dontchev and Rockafellar [23] unifies several such properties: metric regularity, Aubin (pseudo-Lipschitz) property, calmness, isolated calmness, and strong regularity, each of which guarantees a Lipschitz-type response of the solution mapping to small data perturbations.

The regularity properties of KKT solution mappings have been characterized in various optimization settings. For nonlinear programming, strong regularity is characterized by the strong second-order sufficient condition (SSOSC), together with the linear independence constraint qualification (LICQ), as shown in [10, 11, 22]. In the context of conic optimization problems, strong regularity has similarly been characterized: for nonlinear second-order cone programs (NLSOC) and nonlinear semidefinite programming (NLSDP), [9] and [58] established that strong regularity is equivalent to the SSOSC combined with constraint nondegeneracy. Furthermore, for robust isolated calmness, [21] showed that this property holds if and only if the second-order sufficient condition (SOSC) and the strict Robinson constraint qualification (SRCQ) are satisfied for $\mathcal{C}^2$-cone reducible cases. These regularity properties not only deepen our theoretical understanding but also play a crucial role in applications, such as establishing error bounds for optimization methods. Moreover, they are instrumental in analyzing the convergence rates of numerical algorithms, particularly in proving the local superlinear convergence of Newton-type methods and ALM; see, for instance, [8, 35, 40, 59].

In the context of manifold optimization, perturbation analysis becomes more complicated due to the inherent geometric structure of the feasible set. Nevertheless, several recent works have begun to extend Euclidean theory to this setting. Hosseini and Pouryayevali [30] provided a sufficient condition for the weak metric regularity of real-valued functions on complete Riemannian manifolds under strict differentiability, and, as a by-product, derived new necessary optimality conditions. More recently, Bao, Ding, and Zhou [6] investigated the Riemannian robust isolated calmness of the KKT system and established its equivalence to the simultaneous satisfaction of the manifold strong Robinson constraint qualification (M-SRCQ) and the manifold second-order sufficient condition (M-SOSC). This result forms a crucial bridge between constraint qualifications and solution stability, closely mirroring the classical theory developed in Euclidean optimization. In this work, we study the perturbation properties of (1.1). We locally transform (1.1) into an optimization problem on the tangent space $T_{\bar{x}}\mathcal{M}$ at a point $\bar{x}$:

$$\begin{array}{ll} \min & f \circ R_{\bar{x}}(\xi) + \theta(g \circ R_{\bar{x}}(\xi)) \\ \text{s.t.} & \xi \in T_{\bar{x}}\mathcal{M}. \end{array} \tag{1.2}$$

This problem is locally equivalent to (1.1) via the transformation $x = R_{\bar{x}}(\xi)$. Compared with (1.1), problem (1.2) is more approachable because $T_{\bar{x}}\mathcal{M}$ is a Euclidean space. This reformulation allows us to study the stability of the KKT system of the original Riemannian problem (1.1) by examining the corresponding stability of the KKT system of (1.2).

Beyond the stability of the KKT solution mapping, variational sufficiency (and its strong form) provides a complementary lens on perturbation behavior. It is especially valuable for analyzing ALM when multipliers may be non-unique or the problem is non-convex. The recent notion proposed by Rockafellar [52] of variational (strong) convexity lies at the heart of this approach: a function is variationally (strongly) convex around a reference point if, in that neighborhood, both its values and subdifferentials coincide with those of some (strongly) convex function. In the Euclidean setting, the (strong) variational sufficient condition demands that the perturbed augmented objective exhibit this property at a first-order stationary point. Rockafellar showed that strong variational sufficiency guarantees local (strong) convexity of the augmented Lagrangian and thereby underpins fast local convergence of ALM. Furthermore, it is demonstrated that strong variational sufficiency is equivalent to the SSOSC if the function θ in (1.1) (when the manifold $\mathcal{M}$ is taken as a Euclidean space) is a polyhedral function [52] or an indicator function of a second-order cone or semidefinite cone [62]. This condition also coincides with augmented tilt stability, which characterizes Lipschitz continuity of the solution set under linear perturbations.

Motivated by these developments, we aim to extend the (strong) variational sufficient condition to nonsmooth optimization problems on Riemannian manifolds. One challenge is formulating convexity on manifolds. Unlike in Euclidean space, a 'line segment' between two points on a manifold cannot be expressed as a convex combination, which complicates directly extending convexity. To avoid defining convexity on the manifold itself, we examine the variational convexity of the tangent-space problem (1.2) at a stationary point $\bar{x}$, which is locally equivalent to the manifold problem (1.1). Since the tangent space $T_{\bar{x}}\mathcal{M}$ is a Euclidean space, this approach allows us to utilize tools from classical convex analysis. We define the manifold (strong) variational sufficient condition by requiring that the locally equivalent tangent problem satisfies the (strong) variational sufficient condition. Under appropriate assumptions, we show this condition is equivalent to the manifold strong second-order sufficient condition (M-SSOSC). Notably, this equivalence indicates that the concept is independent of the specific choice of retraction. Additionally, we demonstrate that the manifold strong variational sufficient condition guarantees local strong convexity of the augmented Lagrangian for the locally equivalent problem and implies augmented tilt stability. From the perspective of local convexity, variational sufficiency can be viewed as a specialized form of convexity for optimization problems. Since it is equivalent to augmented tilt stability, it maintains a close link to perturbation analysis. However, while the regularity properties of the KKT system discussed earlier involve both primal and dual variables, variational sufficiency concentrates solely on the behavior of the primal variables. Furthermore, this local convexity enables us to construct a local dual problem for manifold optimization, providing a new perspective on the convergence properties of manifold nonsmooth optimization algorithms.

After examining the perturbation characteristics of the KKT system and the manifold variational sufficient condition, we turn to their implications for the convergence analysis of manifold optimization algorithms. Specifically, we examine two recent methods: Riemannian sequential quadratic programming (RSQP) [43] and Riemannian Augmented Lagrangian Method (RALM) [20,64], investigating how the preceding theoretical results underpin their local convergence behavior. Assuming the robust isolated calmness of the KKT system, we demonstrate that RSQP achieves local superlinear (quadratic) convergence. For RALM, under the manifold strong variational sufficient condition, we establish R-linear convergence of the outer iterations and superlinear convergence of the semismooth Newton method used for the inner subproblems. These results not only extend classical Euclidean convergence rate analyses to the manifold setting but also highlight the pivotal role of our newly developed perturbation results in supporting algorithmic efficiency.

The rest of the paper is organized as follows. Section 2 reviews the requisite background on Riemannian geometry and nonsmooth analysis. Section 3 formulates the tangent-space surrogate problem of (1.1) and establishes perturbation properties of its KKT system. In Section 4, we define and study the manifold variational sufficient condition, including its relations to convexity, to M-SSOSC, and to perturbation properties. Sections 5 and 6 leverage these perturbation results for algorithmic analysis: we prove local superlinear convergence of RSQP under isolated calmness, and R-linear (outer) plus superlinear (inner) convergence of RALM under manifold strong variational sufficiency. Section 7 presents numerical experiments validating the theory, while Section 8 concludes and outlines directions for future research.

2 Preliminaries and notations

This section recalls several notions of manifolds that will be used in our discussion. Most of the properties referred to below can be found in the literature, specifically in [2,38].

Let $\mathcal{M}$ be an n-dimensional smooth manifold and $x \in \mathcal{M}$. $\mathfrak{F}_x(\mathcal{M})$ is defined as the set of all smooth real-valued functions on a neighborhood of x. The mapping ξ_x from $\mathfrak{F}_x(\mathcal{M})$ to $\mathbb{R}$ such that there exists a curve γ on $\mathcal{M}$ with $\gamma(0) = x$ satisfying $\xi_x f := \dot{\gamma}(0) f := \left.\frac{d(f(\gamma(t)))}{dt}\right|_{t=0}$ for all $f \in \mathfrak{F}_x(\mathcal{M})$ is called a tangent vector, and the tangent space $T_x\mathcal{M}$ is the set of all tangent vectors to $\mathcal{M}$ at x. If $\mathcal{M}$ is embedded in a Euclidean space $\mathbb{X}$, the normal space $N_x\mathcal{M}$ is defined as the orthogonal complement of $T_x\mathcal{M}$ in $\mathbb{X}$. The tangent bundle is defined as $T\mathcal{M} := \bigcup_x T_x\mathcal{M}$, which is the set of all tangent vectors to $\mathcal{M}$. A map $V : \mathcal{M} \to T\mathcal{M}$ is called a vector field on $\mathcal{M}$ if $V(x) \in T_x\mathcal{M}$ for all $x \in \mathcal{M}$.

Let $F : \mathcal{M} \to \mathbb{X}$ be a smooth mapping between a manifold $\mathcal{M}$ and a Euclidean space $\mathbb{X}$. The mapping $DF(x) : T_x\mathcal{M} \to T_{F(x)}\mathbb{X}$ which is defined by $(DF(x)\xi_x) f := \xi_x(f \circ F)$ for $\xi_x \in T_x\mathcal{M}$ and $f \in \mathfrak{F}_{F(x)}(\mathbb{X})$, is a linear mapping called the differential of F at x. If $\mathcal{M}$ is embedded in a Euclidean space, then $DF(x)$ reduces to the classical directional derivative, i.e., $DF(x)\xi_x = \lim_{t\to 0} \frac{F(x + t\xi_x) - F(x)}{t}$. To distinguish it from the Euclidean differential, we use $h'(x)\xi$ to represent the traditional directional derivative in the direction ξ and $\nabla h(x)$ to be the Euclidean gradient of h.

A Riemannian metric $\langle \cdot, \cdot \rangle_x$ on $\mathcal{M}$ is a smoothly varying inner product with respect to x defined for tangent vectors. A differentiable manifold whose tangent spaces are endowed with Riemannian metrics is called a Riemannian manifold. Whenever the point x is clear from the context, we suppress the subscript and simply write $\langle\cdot,\cdot\rangle$. The induced norm of this inner product is denoted by $\|\cdot\|$ with the subscript being omitted. Given $f \in \mathfrak{F}_x(\mathcal{M})$, the gradient of f at x, denoted by $\operatorname{grad} f(x)$, is defined as the unique tangent vector that satisfying $\langle \operatorname{grad} f(x), \xi\rangle := \xi_x f$ for all $\xi \in T_x\mathcal{M}$.

The length of a curve $\gamma : [a,b] \to \mathcal{M}$ on a Riemannian manifold $\mathcal{M}$ is defined by $L(\gamma) = \int_a^b \sqrt{\langle \dot\gamma(t), \dot\gamma(t)\rangle}\,\mathrm{d}t$, where $\langle\cdot,\cdot\rangle$ denotes the Riemannian metric on $\mathcal{M}$. The Riemannian distance between two points $y, z \in \mathcal{M}$ is given by $d(y,z) := \inf_{\gamma\in\Gamma} L(\gamma)$, where Γ denotes the set of all smooth curves in $\mathcal{M}$ joining y and z. The set $\{y \in \mathcal{M} \mid d(x,y) < \delta\}$ defines a neighborhood of x with radius $\delta > 0$. We use $\operatorname{dist}(x,\mathcal{D})$ to denote the distance from a point $x \in \mathcal{M}$ to a set $\mathcal{D} \subseteq \mathcal{M}$. A geodesic is a curve on $\mathcal{M}$ that locally minimizes arc length. For every tangent vector $\xi \in T_x\mathcal{M}$, there exists an interval $\mathcal{I} \subseteq \mathbb{R}$ containing 0 and a unique geodesic $\gamma(\cdot;x,\xi) : \mathcal{I} \to \mathcal{M}$ satisfying $\gamma(0) = x$ and $\dot\gamma(0) = \xi$. The exponential mapping at $x \in \mathcal{M}$ is defined by $\operatorname{Exp}_x : T_x\mathcal{M} \to \mathcal{M}, \quad \xi \mapsto \operatorname{Exp}_x(\xi) := \gamma(1;x,\xi)$. A vector field X along a smooth curve γ is said to be parallel if $\nabla_{\dot\gamma} X = 0$, where ∇ denotes the Riemannian connection on $\mathcal{M}$. Given a smooth curve γ and a vector $\eta \in T_{\gamma(0)}\mathcal{M}$, there exists a unique parallel vector field X_η along γ such that $X_\eta(0) = \eta$. The parallel transport of η along γ from 0 to t is then defined as $P_\gamma^{0\to t}\eta := X_\eta(t)$. When the geodesic from p to q is unique, denoted by γ_{pq}, we define $P_{pq} := P_{\gamma_{pq}}^{0\to 1}$.

Let $\mathrm{id}_{T_x\mathcal{M}}$ be the identity mapping on $T_x\mathcal{M}$. A retraction on a manifold $\mathcal{M}$ is a smooth mapping R from the tangent bundle $T\mathcal{M}$ onto $\mathcal{M}$ satisfying $R_x(0_x) = x$ and $DR_x(0_x) = \mathrm{id}_{T_x\mathcal{M}}$, where R_x denotes the restriction of R to $T_x\mathcal{M}$. The coordinate expression of $\operatorname{grad} f(x)$ is given by $DR_x(\xi)\operatorname{grad} f(x) = \nabla(f\circ R_x)(\xi)$. The Riemannian Hessian of $f \in \mathfrak{F}_x(\mathcal{M})$ at a point x in $\mathcal{M}$ is defined as the (symmetric) linear mapping $\operatorname{Hess} f(x)$ from $T_x\mathcal{M}$ into itself satisfying $\operatorname{Hess} f(x)\xi = \nabla_\xi \operatorname{grad} f(x)$ for all $\xi \in T_x\mathcal{M}$. By [2, Proposition 5.5.6], if $\operatorname{grad} f(x) = 0$, then

$$\operatorname{Hess} f(x) = \operatorname{Hess}(f\circ R_x)(0_x). \tag{2.1}$$

Consider a subset $\mathcal{A} \subseteq \mathcal{M}$. The epigraph of a function $f : \mathcal{A} \to (-\infty,\infty]$ is defined as $\operatorname{epi} f := \{(x,\alpha) \in \mathcal{A}\times\mathbb{R} \mid f(x) \le \alpha\}$. A proper function $f : \mathcal{A} \to (-\infty,\infty]$ is called lower semicontinuous (lsc) if $\operatorname{epi} f$ is closed. When considering the manifold $\mathcal{M}$ as a topological space, we can use the classical definition of lower semicontinuity of a function $f : \mathcal{M} \to (-\infty,\infty]$, that for any $x \in \mathcal{M}$, it holds that $\liminf_{y\to x} f(y) \ge f(x)$. It is not difficult to verify that this definition is equivalent to the epigraph definition if we consider $\mathcal{M}$ itself as a closed set of the topological space. The following definition of the Fréchet subdifferential is taken from [5, Corollary 4.5].

Definition 1 Let $f : \mathcal{M} \to (-\infty,\infty]$ be a function defined on a Riemannian manifold, and $(\mathcal{U},\varphi)$ is a chart of $\mathcal{M}$. The Fréchet subdifferential $\partial f(x)$ of f at a point $x \in \operatorname{dom} f = \{x \in \mathcal{M} \mid f(x) < \infty\}$ is defined as

$$\begin{aligned}\partial f(x) &= \Big\{D\varphi(x)\zeta \mid \zeta \in \mathbb{R}^n, \liminf_{v\to 0} \frac{f\circ\varphi^{-1}(\varphi(x)+v) - f(x) - \langle\zeta, v\rangle}{\|v\|} \ge 0\Big\} \\ &= \big\{D\varphi(x)\zeta \mid \zeta \in \partial\left(f\circ\varphi^{-1}\right)(\varphi(x))\big\}.\end{aligned}$$

From this definition, one may deduce that $\partial f(x) = \partial(f\circ R_x)(0_x)$ for any given retraction R_x.

With the Riemannian distance $d(\cdot,\cdot)$ defined above, the Lipschitz continuous property of functions can be extended to manifolds. A function $f : \mathcal{M} \to \mathbb{R}$ is Lipschitz continuous in a set $\mathcal{U}$ if for any $y, z \in \mathcal{U}$, $|f(y) - f(z)| \le L d(y,z)$. If there exists a neighborhood $\mathcal{U}$ of $x \in \mathcal{M}$ such that f is Lipschitz continuous on $\mathcal{U}$, we say that f is locally Lipschitz continuous at x. The generalized directional derivative of a locally Lipschitz continuous function f at $x \in \mathcal{M}$ in the direction $v \in T_x\mathcal{M}$ is defined in [29] as

$$f^\circ(x;v) := \limsup_{y\to x, t\downarrow 0} \frac{f\circ\varphi^{-1}(\varphi(y) + tD\varphi(x)v) - f\circ\varphi^{-1}(\varphi(y))}{t},$$

where $(\mathcal{U}, \varphi)$ is a chart containing x. The definition of the generalized directional derivative implies that $f^{\circ}(x, v) = (f \circ R_x)^{\circ}(0_x, v)$ for any retraction. The Clarke subdifferential of a locally Lipschitz continuous function f at $x \in \mathcal{M}$, denoted by $\partial_C f(x)$, is defined as

$$\partial_C f(x) = \{\xi \in T_x\mathcal{M} \mid \langle \xi, v \rangle \leq f^{\circ}(x; v) \text{ for all } v \in T_x\mathcal{M}\}.$$

According to [29, Proposition 2.5], an equivalence can be established between the Clarke subdifferential of f and the Clarke subdifferential of $f \circ R$ for any retraction R, as described below.

Proposition 1 *Let $\mathcal{M}$ be a Riemannian manifold and $x \in \mathcal{M}$. Suppose that $f : \mathcal{M} \to \mathbb{R}$ is Lipschitz continuous near x and $(\mathcal{U}, \varphi)$ is a chart at x. Then*

$$\partial_C f(x) = D\varphi(x)\left[\partial_C\left(f \circ \varphi^{-1}\right)(\varphi(x))\right].$$

Therefore, we have $\partial_C f(x) = \partial_C (f \circ R_x)(0_x)$ for any retraction R_x.

For a given function $\theta : \mathbb{Y} \to (-\infty, +\infty]$, the lower and upper directional epiderivatives (cf. e.g., [10, (2.68) and (2.69)]) of θ at $y \in \operatorname{dom}\theta$ in the direction $h \in \mathbb{Y}$ are defined as

$$\theta_{-}^{\downarrow}(y; h) := \liminf_{t \downarrow 0,\, h' \to h} \frac{\theta(y + th') - \theta(y)}{t}, \quad \theta_{+}^{\downarrow}(y; h) := \sup_{\{t_n\} \in \Sigma} \left(\liminf_{n \to \infty,\, h' \to h} \frac{\theta(y + t_n h') - \theta(y)}{t_n} \right),$$

respectively, where Σ is the set of all positive sequences $\{t_n\}$ converging to zero. Since θ is a proper closed convex function, it follows that $\theta_{-}^{\downarrow}(y; \cdot) = \theta_{+}^{\downarrow}(y; \cdot)$. If $\theta_{-}^{\downarrow}(y; h)$ is finite for $x \in \operatorname{dom}\theta$ and $h \in \mathbb{Y}$, the lower second order epiderivatives [10, (2.76)] for $w \in \mathbb{Y}$ is defined as:

$$\theta_{-}^{\downarrow\downarrow}(y; h, w) := \liminf_{t \downarrow 0,\, w' \to w} \frac{\theta\left(y + th + \frac{1}{2}t^2 w'\right) - \theta(y) - t\theta_{-}^{\downarrow}(y; h)}{\frac{1}{2}t^2}.$$

3 Regularity properties of the KKT solution mapping

In this section, we study the regularity properties of the KKT solution mapping associated with problem (1.1), to formulate intrinsic criteria for stability on Riemannian nonsmooth optimization problems. In the following two subsections, we first reformulate the manifold problem on the tangent space via a suitable retraction and develop the associated first- and second-order calculus, then employ this framework to characterize the desired regularity properties and error bounds for the KKT mapping. These results will lay the theoretical foundations required for the convergence results established in later sections.

3.1 Perturbed problem and variational properties

To analyze the regularity of the KKT solution mapping, we first establish a perturbation scheme for problem (1.1) based on a retraction. Unlike the chart-based construction in [6], our retraction-centered framework is aligned directly with the operations used by manifold algorithms and thus offers both theoretical insights and algorithmic foundations.

For a given $x \in \mathcal{M}$, the inverse function theorem [38, Theorem 4.5] guarantees that any retraction R_x acts as a diffeomorphism in a neighborhood of the origin $0_x \in T_x\mathcal{M}$. The following definition given in [31] defines the injectivity radius of a Riemannian manifold with respect to retraction, which aligns with the definition of the classical injectivity radius [12, Definition 10.19], when the retraction is chosen as the exponential mapping.

Definition 2 The injectivity radius of the manifold $\mathcal{M}$ at a point x with respect to a retraction R_x, denoted by r_{R_x}, is defined as the supremum over all radii $r > 0$ for which R_x is well defined and acts as a diffeomorphism on the open ball $B_x(r) = \{v \in T_x\mathcal{M} \mid \|v\| < r\}$.

By the inverse function theorem, we have $r_{R_x} > 0$ for each given R_x. Additionally, as outlined in [12, Proposition 10.22], we have $d(x, y) = \|v\|$ within the ball $B_x(r_{\mathrm{Exp}_x})$ if $y = \mathrm{Exp}_x(v)$.

To build up our local perturbation framework of (1.1), we first introduce the following function defined on the tangent space at a given point $x \in \mathcal{M}$. Let $x \in \mathcal{M}$ and r_{R_x} be the injectivity radius of $\mathcal{M}$ at x with respect to R_x. For a given function $f : \mathcal{M} \to (-\infty, \infty]$, we denote $f_{R_x} : T_x\mathcal{M} \to (-\infty, \infty]$ by

$$f_{R_x}(\xi) = \begin{cases} f\left(R_x(\xi)\right), & \xi \in B_x(r_{R_x}), \\ +\infty, & \xi \notin B_x(r_{R_x}). \end{cases}$$

This function is the restriction of the pullback of f at a given point x, as defined in [2, Section 4], where the general pullback function refers to $f \circ R$ at each point on the manifold. It is worth noting that f_{R_x} inherit the lower semicontinuity of f within the ball $B_x(r_{R_x})$, although such continuity may fail on its boundary. Fortunately, this limitation does not affect our analysis, as we are concerned only with the behavior of f_{R_x} inside $B_x(r_{R_x})$.

We say that a point $\bar{x} \in \mathcal{M}$ is a KKT point of problem (1.1) if there exists a multiplier $\bar{y} \in \mathbb{Y}$ such that the pair $(\bar{x}, \bar{y})$ satisfies the following KKT conditions:

$$\mathrm{grad}_x\, L(x, y) = \mathrm{grad}\, f(x) + Dg(x)^* y = 0, \quad y \in \partial\theta(g(x)), \tag{3.1}$$

where

$$L(x, y) = f(x) + \langle g(x), y\rangle. \tag{3.2}$$

In what follows, we analyze the problem locally around the KKT point $\bar{x}$. Let $R_{\bar{x}}$ be a retraction at $\bar{x}$ acting as a diffeomorphism on some neighborhood $\mathcal{U} \subseteq T_{\bar{x}}\mathcal{M}$. We define the perturbed problem on $T_{\bar{x}}\mathcal{M}$ in a neighborhood of $\bar{x}$ by

$$\begin{array}{ll} \min & f_{R_{\bar{x}}}(\xi) + \theta(g_{R_{\bar{x}}}(\xi) + b) - \langle \hat{a}, \xi\rangle \\ \text{s.t.} & \xi \in \mathcal{U}, \end{array} \tag{3.3}$$

where $(\hat{a}, b) \in \mathbb{R}^n \times \mathbb{Y}$ is the perturbation parameter. Since $R_{\bar{x}}$ is a diffeomorphism, (3.3) is locally equivalent to the following perturbed problem of (1.1) on $\mathcal{M}$:

$$\begin{array}{ll} \min & f(x) + \theta(g(x) + b) - \langle \hat{a}, R_{\bar{x}}^{-1}(x)\rangle \\ \text{s.t.} & x \in R_{\bar{x}}(\mathcal{U}). \end{array} \tag{3.4}$$

For any given $(\hat{a}, b) \in T_{\bar{x}}\mathcal{M} \times \mathbb{Y}$, we define the KKT solution mapping of (3.3) as

$$\widehat{\mathcal{S}}_{\mathrm{KKT}}(\hat{a}, b) = \left\{(\xi, y) \mid \hat{a} = \nabla f_{R_{\bar{x}}}(\xi) + g'_{R_{\bar{x}}}(\xi)^* y,\ y \in \partial\theta(g_{R_{\bar{x}}}(\xi) + b)\right\},$$

and we define the KKT solution mapping of (3.4) as

$$\mathcal{S}_{\mathrm{KKT}}(\hat{a}, b) = \left\{(x, y) \mid (DR_{\bar{x}}^{-1}(x))^* \hat{a} = \mathrm{grad}\, f(x) + Dg(x)^* y,\ y \in \partial\theta(g(x) + b)\right\}. \tag{3.5}$$

Let $x = R_{\bar{x}}(\xi)$. By the chain rule, we have $\nabla f_{R_{\bar{x}}}(\xi) + g'_{R_{\bar{x}}}(\xi)^* y = DR_{\bar{x}}(\xi)^*(\mathrm{grad}\, f(x) + Dg(x)^* y)$, and $(DR_{\bar{x}}(\xi)^*)^{-1}\hat{a} = (DR_{\bar{x}}^{-1}(x))^*\hat{a}$. It then follows that the KKT solution mappings satisfy

$$\widehat{\mathcal{S}}_{\mathrm{KKT}}(\hat{a}, b) = \mu_{\bar{x}}(\mathcal{S}_{\mathrm{KKT}}(\hat{a}, b) \cap (R_{\bar{x}}(\mathcal{U}) \times \mathbb{Y})), \tag{3.6}$$

where $\mu_{\bar{x}} := (R_{\bar{x}}^{-1}, \mathrm{id}_{\mathbb{Y}})$, and $\mathrm{id}_{\mathbb{Y}}$ denotes the identity mapping on $\mathbb{Y}$. The set of Lagrangian multipliers associated with (3.3) at $(\xi, \hat{a}, b)$ and with (3.4) at $(x, \hat{a}, b)$ are defined by

$$\widehat{M}(\xi, \hat{a}, b) := \left\{ y \in \mathbb{Y} \mid (\xi, y) \in \widehat{\mathcal{S}}_{\mathrm{KKT}}(\hat{a}, b) \right\}, \quad M(x, \hat{a}, b) := \left\{ y \in \mathbb{Y} \mid (x, y) \in \mathcal{S}_{\mathrm{KKT}}(\hat{a}, b) \right\},$$

respectively. It follows from (3.6) that $\widehat{M}(\xi, \hat{a}, b) = M(x, \hat{a}, b)$.

Let $(\hat{a}, b) = (0, 0)$. For problem (3.3) we denote

$$L_{R_{\bar{x}}}(\xi, y) = f_{R_{\bar{x}}}(\xi) + \langle g_{R_{\bar{x}}}(\xi), y \rangle. \tag{3.7}$$

The manifold Robinson constraint qualification (M-RCQ) is said to hold at a feasible solution x of problem (3.4) if

$$Dg(x) T_x \mathcal{M} + \mathcal{T}_{\mathrm{dom}\,\theta}(g(x)) = \mathbb{Y}, \tag{3.8}$$

where $\mathcal{T}_{\mathrm{dom}\,\theta}(g(x))$ is the tangent cone for $g(x)$ and $\mathrm{dom}(\theta)$. Define the critical cone of functions θ and g at a KKT point x and its corresponding multiplier $y \in M(x, 0, 0)$ by

$$\mathcal{C}_{\theta, g}(x, y) := \left\{ d \in \mathbb{Y} \mid \theta^{\downarrow}(g(x); d) = \langle d, y \rangle \right\}.$$

The M-SRCQ is said to hold at a KKT point x with respect to $y \in M(x, 0, 0)$ if

$$Dg(x) T_x \mathcal{M} + \mathcal{C}_{\theta, g}(x, y) = \mathbb{Y}. \tag{3.9}$$

The manifold constraint nondegeneracy is said to hold at x if

$$Dg(x) T_x \mathcal{M} + \mathrm{lin}(\mathcal{T}_{\mathrm{dom}\,\theta}(g(x))) = \mathbb{Y}, \tag{3.10}$$

where $\mathrm{lin}(\mathcal{T}_{\mathrm{dom}\,\theta}(g(x)))$ is the lineality space of $\mathcal{T}_{\mathrm{dom}\,\theta}(g(x))$. If $\mathcal{M}$ is a Euclidean space $\mathbb{E}$, then the tangent space $T_x\mathcal{M}$ coincides with $\mathbb{E}$ itself (see [2, Section 3.5.2]). In this case, the M-RCQ, M-SRCQ, and manifold constraint nondegeneracy conditions reduce to the standard constraint qualifications defined in Euclidean spaces (cf. e.g., [10]).

The critical cone at a stationary point x of problem (3.4) is

$$\mathcal{C}(x) = \left\{ \xi \in T_x\mathcal{M} \mid Df(x)\xi + (\theta \circ g)^{\circ}(x; \xi) = 0 \right\}.$$

By the $\mathcal{C}^2$-cone reducibility of θ, the M-SOSC at x is said to hold if for any $\xi \in \mathcal{C}(x) \backslash \{0\}$,

$$\sup_{y \in M(x,0,0)} \left\{ \langle \xi, \mathrm{Hess}_x\, L(x, y)\xi \rangle - \psi^*_{(g(x), Dg(x)\xi)}(y) \right\} > 0, \tag{3.11}$$

where $\psi^*_{(g(x), Dg(x)\xi)}(\cdot)$ is the conjugate function of $\psi_{(g(x), Dg(x)\xi)}(\cdot) = \theta_-^{\downarrow\downarrow}(g(x); Dg(x)\xi, \cdot)$ for any $g(x) \in \mathrm{dom}\,\theta$ and any $\xi \in T_x\mathcal{M}$. The M-SSOSC at (x, y) is defined as

$$\langle \xi, \mathrm{Hess}_x\, L(x, y)\xi \rangle - \psi^*_{(g(x), Dg(x)\xi)}(y) > 0 \quad \forall\, Dg(x)\xi \in \mathrm{aff}\, \mathcal{C}_{\theta, g}(x, y) \backslash \{0\},$$

where $\mathrm{aff}\, \mathcal{C}_{\theta, g}(x, y)$ is the affine hull of the critical cone $\mathcal{C}_{\theta, g}(x, y)$. Moreover, if θ is the indicator function of a $\mathcal{C}^2$-cone reducible set $\mathcal{K}$, then the M-SSOSC can be equivalently written as

$$\langle \xi, \mathrm{Hess}_x\, L(x, y)\xi \rangle - \sigma\left(\bar{y}, \mathcal{T}^2_{\mathcal{K}}\left(g\left(x\right), Dg\left(x\right)\xi\right)\right) > 0 \quad \forall\, Dg(x)\xi \in \mathrm{aff}\, \mathcal{C}_{\theta, g}(x, y) \backslash \{0\}, \tag{3.12}$$

where $\sigma(y, \mathcal{D})$ is the support function of the set $\mathcal{D}$ at y and $\mathcal{T}^2_{\mathcal{K}}(y, h)$ is the second-order tangent set of $\mathcal{K}$ at y in direction h.

We now establish a precise correspondence between the Riemannian constraint qualifications and second-order optimality conditions of the original problem (3.4) and those of the locally equivalent problem (3.3) in the tangent space. The following proposition shows that these conditions are in fact equivalent in the two settings.

Proposition 2 *Let $(\hat{a}, b) = (0,0)$. The M-RCQ (respectively, M-SRCQ, manifold constraint nondegeneracy, M-SOSC, M-SSOSC) holds at $\bar{x}$ (with respect to $\bar{y}$) for problem (3.4) if and only if the RCQ (respectively, SRCQ, constraint nondegeneracy, SOSC, SSOSC) holds at $0_{\bar{x}}$ (with respect to $\bar{y}$) for problem (3.3).*

Proof The RCQ holds at $0_{\bar{x}}$ for problem (3.3) if

$$(g \circ R_{\bar{x}})'(0_{\bar{x}}) T_{\bar{x}}\mathcal{M} + \mathcal{T}_{\mathrm{dom}\,\theta}(g \circ R_{\bar{x}}(0_{\bar{x}})) = \mathbb{Y},$$

where for any $\xi \in T_{\bar{x}}\mathcal{M}$, $(g \circ R_{\bar{x}})'(0_{\bar{x}})\xi = DR_{\bar{x}}(0_{\bar{x}})[Dg(\bar{x})\xi] = Dg(\bar{x})\xi$, and $g \circ R_{\bar{x}}(0_{\bar{x}}) = g(\bar{x})$. Therefore, M-RCQ and manifold constraint nondegeneracy hold at $\bar{x}$ for (3.4) if and only if RCQ and constraint nondegeneracy hold at $0_{\bar{x}}$ for (3.3). For the M-SRCQ, it suffices to show that $\mathcal{C}_{\theta,g}(\bar{x}, \bar{y}) = \mathcal{C}_{\theta, g\circ R_{\bar{x}}}(0_{\bar{x}}, \bar{y})$, which follows immediately from the definition. By (2.1), we have $\mathrm{Hess}_x\, L(\bar{x}, \bar{y}) = \mathrm{Hess}_\xi\, L_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y})$ at the stationary point $\bar{x}$, which shows the equivalence between SOSC (resp. SSOSC) for problem (3.3) and M-SOSC (resp. M-SSOSC) for problem (3.4). The proof is completed. □

Remark 1 The definitions of M-RCQ, M-SRCQ, manifold constraint nondegeneracy, M-SOSC, and M-SSOSC for problem (3.4) are independent of the choice of retraction at $\bar{x}$. Although the formulation of the locally equivalent problem (3.3) in the tangent space involves on a specific retraction $R_{\bar{x}}$, Proposition 2 shows that the constraint qualifications and second-order optimality conditions admit the Riemannian counterparts regardless of the retraction used. This confirms that these properties are intrinsically independent of $R_{\bar{x}}$.

3.2 Characterizations of the regularity properties

In this subsection, we shall investigate the characterizations of regularity properties for set-valued mappings on Riemannian manifolds. Specifically, we extend the Euclidean concepts of the Aubin property, calmness, isolated calmness, and strong regularity, which are central tools for assessing solution sensitivity in optimization [23, Chapter 3], to mappings whose domain and range lie on smooth manifolds. These properties are essential for understanding the stability of solutions. We first extend the notion of excess, which measures the distance between two sets in Euclidean space and is defined in [23, Section 3A], to the Riemannian settings.

Definition 3 For two given sets C and D in $\mathcal{M}$, the excess of C beyond D is defined by

$$e(C, D) = \sup_{x \in C} \mathrm{dist}(x, D).$$

Building on the definition of distance between two sets on a Riemannian manifold, we are able to extend the notions of regularity for set-valued mappings, including the Aubin property, calmness, isolated calmness, and strong regularity, within the context of Riemannian geometry.

Definition 4 Let $\mathbb{E}$ be a Euclidean space. A set-valued mapping $\Psi : \mathbb{E} \rightrightarrows \mathcal{M}$ is said to have the Aubin property at $\bar{p}$ for $\bar{x}$ if there exist a constant $\kappa > 0$ and open neighborhoods $\mathcal{U}$ of $\bar{p}$ and $\mathcal{V}$ of $\bar{x}$ such that

$$e(\Psi(p') \cap \mathcal{V}, \Psi(p)) \le \kappa \|p' - p\| \quad \forall\, p, p' \in \mathcal{U}.$$

Definition 5 Let $\mathbb{E}$ be a Euclidean space. A set-valued mapping $\Psi : \mathbb{E} \rightrightarrows \mathcal{M}$ is said to be calm at $\bar{p}$ for $\bar{x}$ if there exist a constant $\kappa > 0$ and open neighborhoods $\mathcal{U}$ of $\bar{p}$ and $\mathcal{V}$ of $\bar{x}$ such that

$$e(\Psi(p) \cap \mathcal{V}, \Psi(\bar{p})) \le \kappa \|p - \bar{p}\| \quad \forall\, p \in \mathcal{U}.$$

Definition 6 Let $\mathbb{E}$ be a Euclidean space. A set-valued mapping $\Psi : \mathbb{E} \rightrightarrows \mathcal{M}$ is said to be isolated calm at $\bar{p}$ for $\bar{x}$ if there exist a constant $\kappa > 0$ and open neighborhoods $\mathcal{U}$ of $\bar{p}$ and $\mathcal{V}$ of $\bar{x}$ such that

$$d(x, \bar{x}) \leq \kappa \|p - \bar{p}\| \quad \forall\, x \in \Psi(p) \cap \mathcal{V} \quad \text{and} \quad p \in \mathcal{U}. \tag{3.13}$$

Moreover, Ψ is said to be robustly isolated calm at $\bar{p}$ for $\bar{x}$ if (3.13) holds and, for each $p \in \mathcal{U}$, $\Psi(p) \cap \mathcal{V} \neq \emptyset$.

Definition 7 Let $\mathbb{E}$ be a Euclidean space. A set-valued mapping $\Psi : \mathbb{E} \rightrightarrows \mathcal{M}$ is said to be strongly regular at $\bar{p}$ for $\bar{x}$ if there exist open neighborhoods $\mathcal{U}$ of $\bar{p}$ and $\mathcal{V}$ of $\bar{x}$ such that for any $p \in \mathcal{U}$, the set $\Psi(p) \cap \mathcal{V}$ is a singleton, and the mapping $\Psi(\cdot) \cap \mathcal{V}$ is Lipschitz continuous on $\mathcal{U}$. That is, there exists $L > 0$ such that for any $p_1, p_2 \in \mathcal{U}$, one has $d(x_1, x_2) \leq L\|p_1 - p_2\|$ for $x_1 := \Psi(p_1) \cap \mathcal{V}$ and $x_2 := \Psi(p_2) \cap \mathcal{V}$.

Notably, when the underlying manifold is structured as a Euclidean space, these definitions naturally reduce to their classical counterparts within the established framework of variational analysis [23], thus maintaining consistency with established regularity theory in the Euclidean setting. Let $\bar{x} \in \mathcal{M}$ be a given point on the manifold. For any retraction $R_{\bar{x}}$ around $\bar{x}$, there exists an open subset $\mathcal{U} \subseteq T_{\bar{x}}\mathcal{M}$ on which $R_{\bar{x}}$ is a diffeomorphism. Given a set-valued mapping $\Psi : \mathbb{E} \rightrightarrows \mathcal{M}$ from a Euclidean space $\mathbb{E}$ to the manifold $\mathcal{M}$, we define its Euclidean counterpart $\widetilde{\Psi}(\cdot) : \mathbb{E} \rightrightarrows T_{\bar{x}}\mathcal{M}$ under $R_{\bar{x}}$ by $\widetilde{\Psi}(\cdot) := R_{\bar{x}}^{-1}(\Psi(\cdot)) \cap \mathcal{U}$. The next lemma establishes the relationship between the manifold mapping Ψ and its Euclidean counterpart $\widetilde{\Psi}$.

Lemma 1 *Let $\Psi : \mathbb{E} \rightrightarrows \mathcal{M}$ be a set-valued mapping, and let $R_{\bar{x}}$ be a retraction around $\bar{x} \in \Psi(\bar{p})$. Assume that $R_{\bar{x}}$ is a diffeomorphism on $\mathcal{U} \subseteq T_{\bar{x}}\mathcal{M}$. If the Euclidean counterpart $\widetilde{\Psi} : \mathbb{E} \rightrightarrows T_{\bar{x}}\mathcal{M}$ defined by $\widetilde{\Psi}(\cdot) := R_{\bar{x}}^{-1}(\Psi(\cdot)) \cap \mathcal{U}$ is Aubin (calm, isolated calm, strongly regular) at $\bar{p}$ with respect to $0_{\bar{x}}$, then Ψ is Aubin (calm, isolated calm, strongly regular) at $\bar{p}$ with respect to $\bar{x}$.*

Proof If $\widetilde{\Psi}$ is Aubin at $\bar{p}$ with respect to $0_{\bar{x}}$, then there exist a constant $\kappa > 0$ and open neighborhoods $\mathcal{P}$ of $\bar{p}$ and $\mathcal{V}$ of $\bar{x}$ such that for any $p, p' \in \mathcal{P}$, $e(\widetilde{\Psi}(p') \cap \mathcal{V}, \widetilde{\Psi}(p)) \leq \kappa\|p' - p\|$, where

$$e(\widetilde{\Psi}(p') \cap \mathcal{V}, \widetilde{\Psi}(p)) = \sup_{\xi' \in \widetilde{\Psi}(p') \cap \mathcal{V}} \inf_{\xi \in \widetilde{\Psi}(p)} \|\xi' - \xi\| = \sup_{\xi' \in R_{\bar{x}}^{-1}(\Psi(p') \cap R_{\bar{x}}(\mathcal{V}))} \inf_{\xi \in R_{\bar{x}}^{-1}(\Psi(p)) \cap \mathcal{U}} \|\xi' - \xi\|.$$

Shrinking $\mathcal{U}$ if necessary, it follows from [46, Corollary 5.7.11] that the Riemannian distance function $d : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ is smooth on $R_{\bar{x}}(\mathcal{U})$. Hence, there exists $\alpha > 0$ such that for any $x, x' \in R_{\bar{x}}(\mathcal{U})$, $d(x', x) \leq \alpha\|R_{\bar{x}}^{-1}(x') - R_{\bar{x}}^{-1}(x)\|$. Therefore,

$$\begin{aligned} e(\Psi(p') \cap R_{\bar{x}}(\mathcal{V}), \Psi(p)) &= \sup_{x' \in \Psi(p') \cap R_{\bar{x}}(\mathcal{V})} \inf_{x \in \Psi(p)} d(x', x) \\ &\leq \sup_{x' \in \Psi(p') \cap R_{\bar{x}}(\mathcal{V})} \inf_{x \in \Psi(p) \cap R_{\bar{x}}(\mathcal{U})} d(x', x) \\ &\leq \sup_{\xi' \in R_{\bar{x}}^{-1}(\Psi(p') \cap R_{\bar{x}}(\mathcal{V}))} \inf_{\xi \in R_{\bar{x}}^{-1}(\Psi(p)) \cap \mathcal{U}} \alpha\|\xi' - \xi\| \\ &\leq \alpha\kappa\|p' - p\|. \end{aligned}$$

This shows that Ψ is Aubin at $\bar{p}$ with respect to $\bar{x}$. Similarly, if $\widetilde{\Psi}$ is calm or isolated calm at $\bar{p}$ with respect to $0_{\bar{x}}$, we can prove that Ψ is calm or isolated calm at $\bar{p}$ with respect to $\bar{x}$.

Finally, if $\widetilde{\Psi}$ is strongly regular at $\bar{p}$ with respect to $0_{\bar{x}}$, then there exists a neighborhood $\mathcal{V}$ of $0_{\bar{x}}$ such that the Aubin property of $\widetilde{\Psi}$ about $0_{\bar{x}}$ holds, which means that Ψ has Aubin property at $\bar{p}$ with $\bar{x}$. Since $\widetilde{\Psi}(p) \cap \mathcal{V}$ is a singleton for any $p \in \mathcal{P}$, where $\mathcal{P}$ is a neighborhood of $\bar{p}$, and $R_{\bar{x}}$ is a diffeomorphism over $\mathcal{U}$, it can be proved that $\Psi(p) \cap R_{\bar{x}}(\mathcal{V})$ is also a singleton. This means that Ψ is strongly regular at $\bar{p}$ with respect to $\bar{x}$. $\square$

We now present the main results of this section, namely, the characterization of the regularity properties of the KKT solution mapping (3.5). Most of the proofs are omitted, as many of the results follow directly from Proposition 2, Lemma 1, and the corresponding conclusions established in the Euclidean setting. Define the natural mapping $F : \mathcal{M} \times \mathbb{Y} \to T\mathcal{M} \times \mathbb{Y}$ for problem (3.4) by

$$F(x, y) := \begin{bmatrix} \operatorname{grad} f(x) + Dg(x)^* y \\ g(x) - \operatorname{prox}_\theta (g(x) + y) \end{bmatrix}, \quad (x, y) \in \mathcal{M} \times \mathbb{Y},$$

where $\operatorname{prox}_\theta(u) := \operatorname{argmin}_{y \in \mathbb{Y}} \left\{ \theta(y) + \frac{1}{2}\|u - y\|^2 \right\}$ denotes the proximal mapping of θ. It is straightforward to verify that $(0, 0, \bar{x}, \bar{y}) \in \operatorname{gph} \mathcal{S}_{\mathrm{KKT}}$ if and only if $(0, 0, \bar{x}, \bar{y}) \in \operatorname{gph} F^{-1}$. Let $\partial_C F(\bar{x}, \bar{y})$ denote Clarke's generalized Jacobian of F at $(\bar{x}, \bar{y})$. With this setup, we are now ready to establish the following proposition on the nonsingularity of $\partial_C F(\bar{x}, \bar{y})$ for nonlinear programming (NLP), NLSOC, and NLSDP.

Proposition 3 *Suppose that* $\operatorname{epi} \theta$ *is* $\mathcal{C}^2$*-cone reducible in the sense of [10, Definition 3.135]. Let* $(\bar{x}, \bar{y}) \in \mathcal{M} \times \mathbb{Y}$ *be a KKT pair of problem* (1.1) *satisfying* (3.1)*. Assume that the M-RCQ* (3.8) *holds at* $\bar{x}$*. Then the following statement holds:*

- (i) *The M-SOSC* (3.11) *holds at* $\bar{x}$ *and the M-SRCQ* (3.9) *holds at* $\bar{x}$ *with respect to* $\bar{y}$ *if and only if the KKT solution mapping* $\mathcal{S}_{\mathrm{KKT}}$ *in* (3.5) *is robustly isolated calm at the origin for* $(\bar{x}, \bar{y})$*.*

Furthermore, if $\operatorname{epi} \theta$ *is a convex polyhedral set, a second-order cone, or a semidefinite cone, then the following statements are equivalent:*

- (ii) *The M-SSOSC holds at* $(\bar{x}, \bar{y})$ *and the manifold constraint nondegeneracy holds at* $\bar{x}$*;*
- (iii) *The KKT solution mapping* $\mathcal{S}_{\mathrm{KKT}}$ *is strongly regular at the origin for* $(\bar{x}, \bar{y})$*;*
- (iv) *All elements in* $\partial_C F(\bar{x}, \bar{y})$ *are nonsingular.*

Proof Statement (i) has already been established in [6, Proposition 5]. The equivalence between statements (ii) and (iii) follows from Proposition 2, Lemma 1, and the results in [9, 10, 58]. Hence, it remains to prove the equivalence between (iii) and (iv). Define the natural mapping $\widehat{F} : T_{\bar{x}}\mathcal{M} \times \mathbb{Y} \to T_{\bar{x}}\mathcal{M} \times \mathbb{Y}$ for problem (3.3) by

$$\widehat{F}(\xi, y) := \begin{bmatrix} \nabla f_{R_{\bar{x}}}(\xi) + g'_{R_{\bar{x}}}(\xi)^* y \\ g_{R_{\bar{x}}}(\xi) - \operatorname{prox}_\theta (g_{R_{\bar{x}}}(\xi) + y) \end{bmatrix}, \quad (\xi, y) \in T_{\bar{x}}\mathcal{M} \times \mathbb{Y}.$$

For the cases of NLP, NLSOC, and NLSDP, it follows from [10, Proposition 5.38], [9, Theorem 30], and [58, Theorem 4.1] that every element in $\partial_C \widehat{F}(0_{\bar{x}}, \bar{y})$ is nonsingular if and only if (iii) holds. It remains to show $\partial_C F(\bar{x}, \bar{y}) = \partial_C \widehat{F}(0_{\bar{x}}, \bar{y})$. For $x = R_{\bar{x}}(\xi)$, by the chain rule, we have

$$\widehat{F}(\xi, y) = \begin{bmatrix} DR_{\bar{x}}(\xi)^* (\operatorname{grad} f(x) + Dg(x)^* y) \\ g_{R_{\bar{x}}}(\xi) - \operatorname{prox}_\theta (g_{R_{\bar{x}}}(\xi) + y) \end{bmatrix}.$$

The first components of F and $\widehat{F}$ are equal by applying [2, Proposition 5.5.6] at the KKT pair $(\bar{x}, \bar{y})$. Define $h(x, y) := g(x) - \operatorname{prox}_\theta (g(x) + y)$ and $\widehat{h}(\xi, y) := g_{R_{\bar{x}}}(\xi) - \operatorname{prox}_\theta (g_{R_{\bar{x}}}(\xi) + y)$. Clearly, $h(x, y) = \widehat{h}(R_{\bar{x}}^{-1}(x), y)$. It follows from [6, Proposition 1] that

$$\partial_C h(x, y) \subseteq \{(DR_{\bar{x}}^{-1}(x)^*[u], w) \mid (u, w) \in \partial_C \widehat{h}(R_{\bar{x}}^{-1}(x), y)\}.$$

Moreover, since $DR_{\bar{x}}^{-1}(x)^*[u] = u$ at $\bar{x}$ for any u, an extension of [18, Theorem 2.3.10] to the Riemannian setting gives $\partial_C h(\bar{x}, \bar{y}) = \partial_C \widehat{h}(0_{\bar{x}}, \bar{y})$. Combining the above, we obtain $\partial_C F(\bar{x}, \bar{y}) = \partial_C \widehat{F}(0_{\bar{x}}, \bar{y})$. This completes the proof. □

Recently, new results concerning the equivalence between the Aubin property and strong regularity for the KKT solution mappings of NLSOC and NLSDP have been established in [14, 15]. By combining these findings with the classical equivalence results for convex polyhedral programming from [22], we obtain the following proposition.

Proposition 4 *Suppose that* $\operatorname{epi}\theta$ *is a convex polyhedral set, a second-order cone, or a semidefinite cone. Let* $(\bar{x}, \bar{y}) \in \mathcal{M} \times \mathbb{Y}$ *be a KKT pair of problem* (1.1) *satisfying* (3.1)*. Assume that the M-RCQ (3.8) holds at* $\bar{x}$*. Then the KKT solution mapping* $\mathcal{S}_{\mathrm{KKT}}$ *in* (3.5) *has the Aubin property at the origin for* $(\bar{x}, \bar{y})$ *if and only if it is strongly regular at the origin for* $(\bar{x}, \bar{y})$*.*

Beyond the above equivalence results, the robust isolated calmness or strong regularity of the KKT solution mapping guarantees a two-sided error bound in a neighborhood of the KKT pair $(\bar{x}, \bar{y})$. This two-sided error bound plays a crucial role in establishing the local linear convergence of the Riemannian augmented Lagrangian method (RALM), as we will show in Section 6. To formalize this, define the KKT residual mapping $E : \mathcal{M} \times \mathbb{Y} \to \mathbb{R}$ by

$$E(x, y) := \|\operatorname{grad} f(x) + Dg(x)^* y\| + \|g(x) - \operatorname{prox}_\theta (g(x) + y)\|, \quad (x, y) \in \mathcal{M} \times \mathbb{Y}. \tag{3.14}$$

Proposition 5 *Let* $(\bar{x}, \bar{y}) \in \mathcal{M} \times \mathbb{Y}$ *be a KKT pair of problem* (1.1) *satisfying* (3.1)*. Suppose that the KKT solution mapping* $\mathcal{S}_{\mathrm{KKT}}$ *in* (3.5) *is robustly isolated calm (or strongly regular) at* $(\bar{x}, \bar{y})$*. Then* $M(\bar{x}, 0, 0) = \{\bar{y}\}$*. Furthermore, there exist constants* $c_1, c_2 > 0$ *such that, for any* $(x, y) \in \mathcal{M} \times \mathbb{Y}$ *sufficiently close to* $(\bar{x}, \bar{y})$ *with* $E(x, y)$ *sufficiently small, it holds that*

$$c_1 E(x, y) \leq d(x, \bar{x}) + \|y - \bar{y}\| \leq c_2 E(x, y).$$

Proof The conclusion follows by combining Proposition 3, [6, Proposition 5], [6, Theorem 7], and the fact that strong regularity implies isolated calmness. □

Remark 2 In [6], the results in Proposition 5, as well as the equivalence between M-SOSC with M-SRCQ and robust isolated calmness, were established using the normal coordinate chart around the KKT point.In contrast, this paper derives these results via the retraction mapping at the KKT point $\bar{x}$. These two approaches are essentially equivalent, since there exists a local diffeomorphism between the normal coordinate chart and the exponential map around $\bar{x}$, and the retraction serves as a first-order approximation of the exponential map.

4 Manifold variational sufficiency

In the previous section, we study the stability of the KKT system of Riemannian nonsmooth optimization problem (1.1), which involves both the primal variable and its multipliers. These properties offer insights into local convergence, in particular, the local convergence of Newton-type methods (e.g., RSQP), and will be analyzed in detail in subsequent sections. In this section, we focus on another concept related to perturbation analysis, (strong) variational sufficiency, which is recently introduced in [52]. In Euclidean setting, the strong variational sufficiency is equivalent to the augmented tilt stability [52], and serves as a key tool for analyzing the local convergence rate of the ALM in situations where standard assumptions like local convexity or uniqueness of Lagrange multipliers do not hold [53]. In this section, we shall extend this concept to the Riemannian setting and examine how manifold (strong) variational sufficiency relates to optimality conditions and perturbation analysis. Its significance in the convergence analysis of RALM will be demonstrated in a later section.

4.1 Augmented problem and its first-order optimality conditions

As in Section 3.1, given a KKT point $\bar{x} \in \mathcal{M}$, let $R_{\bar{x}}$ be a retraction at $\bar{x}$ and a diffeomorphism in $\mathcal{U} \subseteq T_{\bar{x}}\mathcal{M}$. Let $\hat{a} = 0$ in this section. Consider the following augmented problem for (3.4):

$$\begin{array}{rl} \min & f(x) + \theta(g(x) + b) + \frac{\rho}{2}\|b\|^2 \\ \text{s.t.} & x \in R_{\bar{x}}(\mathcal{U}). \end{array} \tag{4.1}$$

Locally, (4.1) is equivalent to the following augmented problem for (3.3):

$$\begin{array}{rl} \min & f_{R_{\bar{x}}}(\xi) + \theta(g_{R_{\bar{x}}}(\xi) + b) + \frac{\rho}{2}\|b\|^2 \\ \text{s.t.} & \xi \in \mathcal{U}. \end{array} \tag{4.2}$$

Denote the objective functions by

$$\varphi(x, b) = f(x) + \theta(g(x) + b), \quad \varphi_{R_{\bar{x}}}(\xi, b) = f_{R_{\bar{x}}}(\xi) + \theta(g_{R_{\bar{x}}}(\xi) + b). \tag{4.3}$$

Then the Lagrangian functions for problems (3.4) and (3.3) are given, respectively, by

$$\begin{aligned} l(x, y) &= \inf_b \{\varphi(x, b) - \langle y, b\rangle\} = L(x, y) - \theta^*(y), \\ l_{R_{\bar{x}}}(\xi, y) &= \inf_b \{\varphi_{R_{\bar{x}}}(\xi, b) - \langle y, b\rangle\} = L_{R_{\bar{x}}}(\xi, y) - \theta^*(y), \end{aligned}$$

where L and $L_{R_{\bar{x}}}$ are defined in (3.2) and (3.7), respectively. For $\rho > 0$, the augmented Lagrangian function of problems (3.4) and (3.3) are given, respectively, by

$$l^\rho(x, y) = \inf_b \left\{\varphi(x, b) - \langle y, b\rangle + \frac{\rho}{2}\|b\|^2\right\}, \quad l^\rho_{R_{\bar{x}}}(\xi, y) = \inf_b \left\{\varphi_{R_{\bar{x}}}(\xi, b) - \langle y, b\rangle + \frac{\rho}{2}\|b\|^2\right\}. \tag{4.4}$$

The augmented Lagrangian functions can be regarded as the Lagrangian functions of the augmented objective functions

$$\varphi^\rho(x, b) = \varphi(x, b) + \frac{\rho}{2}\|b\|^2, \quad \varphi^\rho_{R_{\bar{x}}}(\xi, b) = \varphi_{R_{\bar{x}}}(\xi, b) + \frac{\rho}{2}\|b\|^2, \tag{4.5}$$

respectively. Thus, by definition, $-l(x, \cdot)$ and $-l_{R_{\bar{x}}}(\xi, \cdot)$ are the conjugate functions of $\varphi(x, \cdot)$ and $\varphi_{R_{\bar{x}}}(\xi, \cdot)$, respectively, and similarly, $-l^\rho(x, \cdot)$ and $-l^\rho_{R_{\bar{x}}}(\xi, \cdot)$ are the conjugate functions of $\varphi^\rho(x, \cdot)$ and $\varphi^\rho_{R_{\bar{x}}}(\xi, \cdot)$, respectively. Since θ is closed proper convex, by [49, Theorem 12.2],

$$\begin{array}{ll} \varphi(x, b) = \sup_y \{l(x, y) + \langle y, b\rangle\}, & \varphi^\rho(x, b) = \sup_y \left\{l^\rho(x, y) + \langle y, b\rangle\right\}, \\ \varphi_{R_{\bar{x}}}(\xi, b) = \sup_y \{l_{R_{\bar{x}}}(\xi, y) + \langle y, b\rangle\}, & \varphi^\rho_{R_{\bar{x}}}(\xi, b) = \sup_y \{l^\rho_{R_{\bar{x}}}(\xi, y) + \langle y, b\rangle\}. \end{array} \tag{4.6}$$

We now examine the relationships between the first-order conditions of problems (3.4) and (3.3), together with those of their augmented counterparts (4.1) and (4.2). In fact, the first-order condition in Proposition 6 is also equivalent to the KKT conditions (3.1).

Proposition 6 *Let $\rho > 0$, $(\bar{x}, \bar{y})$ be given, and let $R_{\bar{x}}$ be a retraction at $\bar{x}$. The following statements are equivalent:*

(i) $(\bar{x}, \bar{y})$ *satisfies the first-order condition of problem* (3.4), *i.e.,* $(0, \bar{y}) \in \partial\varphi(\bar{x}, 0)$*;*

(ii) $(0_{\bar{x}}, \bar{y})$ *satisfies the first-order condition of problem* (3.3), *i.e.,* $(0, \bar{y}) \in \partial\varphi_{R_{\bar{x}}}(0_{\bar{x}}, 0)$*;*

(iii) $(\bar{x}, \bar{y})$ *satisfies the first-order condition of problem* (4.1), *i.e.,* $(0, \bar{y}) \in \partial\varphi^\rho(\bar{x}, 0)$*;*

(iv) $(0_{\bar{x}}, \bar{y})$ *satisfies the first-order condition of problem* (4.2), *i.e.,* $(0_{\bar{x}}, \bar{y}) \in \partial\varphi^\rho_{R_{\bar{x}}}(0_{\bar{x}}, 0)$*;*

(v) $\operatorname{grad}_x l(\bar{x},\bar{y}) = 0,\ 0 \in \partial_y[-l](\bar{x},\bar{y})$, *or* $\operatorname{grad}_x L(\bar{x},\bar{y}) = 0,\ \bar{y} \in \partial\theta(g(\bar{x}))$;

(vi) $\nabla_\xi l_{R_{\bar{x}}}(0_{\bar{x}},\bar{y}) = 0,\ 0 \in \partial_y[-l_{R_{\bar{x}}}](0_{\bar{x}},\bar{y})$, *or* $\nabla_\xi L_{R_{\bar{x}}}(0_{\bar{x}},\bar{y}) = 0,\ \bar{y} \in \partial\theta(g_{R_{\bar{x}}}(0_{\bar{x}}))$;

(vii) $\operatorname{grad}_x l^\rho(\bar{x},\bar{y}) = 0,\ 0 \in \nabla_y l^\rho(\bar{x},\bar{y})$, *or* $\operatorname{grad}_x L(\bar{x},\bar{y}) = 0,\ \nabla \operatorname{env}_\rho \theta(g(\bar{x}) + \rho^{-1}\bar{y}) = \bar{y}$, *where* $\operatorname{env}_\rho \theta$ *is the Moreau envelope of* θ *defined by* $\operatorname{env}_\rho \theta(p) := \min_{y\in\mathbb{Y}} \theta(y) + \dfrac{\rho}{2}\|p-y\|^2$;

(viii) $\nabla_\xi l^\rho_{R_{\bar{x}}}(0_{\bar{x}},\bar{y}) = 0,\ 0 \in \nabla_y l^\rho_{R_{\bar{x}}}(0_{\bar{x}},\bar{y})$, *or* $\nabla_\xi L_{R_{\bar{x}}}(0_{\bar{x}},\bar{y}) = 0,\ \nabla \operatorname{env}_\rho \theta(g_{R_{\bar{x}}}(0_{\bar{x}}) + \rho^{-1}\bar{y}) = \bar{y}$.

Proof It follows from Proposition 1 that $(0,\bar{y}) \in \partial\varphi(\bar{x},0) \iff (0,\bar{y}) \in \partial\varphi_{R_{\bar{x}}}(0_{\bar{x}},0)$. Moreover, when $b=0$, it is obvious that $\partial\varphi(\bar{x},0) = \partial\varphi^\rho(\bar{x},0)$ and $\partial\varphi_{R_{\bar{x}}}(0_{\bar{x}},0) = \partial\varphi^\rho_{R_{\bar{x}}}(0_{\bar{x}},0)$, which implies the equivalence of (i), (ii), (iii), and (iv). By (4.6) and the chain rules of the subdifferential of manifold functions [6, Proposition 1], we have

$$(0_{\bar{x}},\bar{y}) \in \partial\varphi(\bar{x},0) \iff \bar{y} \in \partial\theta(g(\bar{x})),\ 0_{\bar{x}} = \operatorname{grad}_x L(\bar{x},\bar{y}) \iff 0_{\bar{x}} \in \partial_x l(\bar{x},\bar{y}),\ 0 \in \partial_y[-l](\bar{x},\bar{y}).$$

Moreover, Proposition 1 yields

$$\begin{aligned}(0_{\bar{x}},\bar{y}) \in \partial\varphi(\bar{x},0) \iff (0_{\bar{x}},\bar{y}) \in \partial\varphi_{R_{\bar{x}}}(0_{\bar{x}},0) &\iff \bar{y} \in \partial\theta(g_{R_{\bar{x}}}(0_{\bar{x}})),\ 0_{\bar{x}} = \nabla_\xi l_{R_{\bar{x}}}(0_{\bar{x}},\bar{y}) \\ &\iff 0_{\bar{x}} \in \partial_\xi l_{R_{\bar{x}}}(0_{\bar{x}},\bar{y}),\ 0 \in \partial_y[-l_{R_{\bar{x}}}](0_{\bar{x}},\bar{y}).\end{aligned}$$

Therefore, statements (i), (ii), (v), and (vi) are equivalent. Similarly, we can prove the equivalence between (i), (ii), (vii) and (viii). □

4.2 Characterizations of manifold variational sufficiency

We now turn to the notion of manifold variational sufficiency. Before that, let us first recall the concept of variational convexity in the Euclidean case, as introduced in [52]. Let $f : \mathbb{R}^n \to (-\infty,\infty]$ be a lsc function. We say that f is variationally convex with respect to $(\bar{w},\bar{z}) \in \operatorname{gph}\partial f$ if there exist convex open neighborhoods $\mathcal{W}$ of $\bar{w}$ and $\mathcal{Z}$ of $\bar{z}$, together with a proper lsc convex function h satisfying $h \le f$ on $\mathcal{W}$, such that

$$[\mathcal{W}\times\mathcal{Z}] \cap \operatorname{gph}\partial h = [\mathcal{W}\times\mathcal{Z}] \cap \operatorname{gph}\partial f,$$

and $h(w) = f(w)$ holds for all (w,z) in this common set. If, in addition, h is strongly convex on $\mathcal{W}$, then we say that f is variationally strongly convex with respect to $(\bar{w},\bar{z})$. In the Euclidean setting, the (strong) variational sufficient condition for problem $\min\{\varphi(x,b) \mid b=0\}$ is said to hold with respect to $(\bar{x},\bar{y})$ satisfying the first-order condition $(0,\bar{y}) \in \partial\varphi(\bar{x},0)$ at level $\rho > 0$ if the augmented objective φ^ρ is variationally (strongly) convex with respect to $((\bar{x},0),(0,\bar{y})) \in \operatorname{gph}\partial\varphi^\rho$. Here, φ and φ^ρ are defined in (4.3) and (4.5), respectively.

We define the manifold (strong) variational sufficiency for problem (1.1) through the (strong) variational sufficiency for its locally equivalent formulation (1.2) on the tangent space, which is problem (4.2) with $b=0$. This definition is then based on the variational (strong) convexity of the objective $\varphi^\rho_{R_{\bar{x}}}$ in problem (4.2), and therefore depends on the choice of the retraction $R_{\bar{x}}$. We will later establish, however, that the notion of manifold (strong) variational sufficiency is in fact independent of the particular retraction selected.

Definition 8 Let $(\bar{x},\bar{y}) \in \mathcal{M}\times\mathbb{Y}$ be a KKT pair of problem (1.1) satisfying (3.1). With respect to $(\bar{x},\bar{y})$ at level $\rho > 0$, the manifold (strong) variational sufficient condition for problem (1.1) under the retraction $R_{\bar{x}}$ is said to hold if the (strong) variational sufficient condition holds for problem (1.2).

Motivated by the recent developments in [52,53], we next discuss the local augmented dual property for problem (1.1). In particular, the following proposition establishes the equivalence between the manifold (strong) variational sufficiency and the (strong) convexity of the augmented Lagrangian function, which follows directly from [52, Theorem 1] and Proposition 1.

Proposition 7 *Let $(\bar{x}, \bar{y}) \in \mathcal{M} \times \mathbb{Y}$ be a KKT pair of problem* (1.1) *satisfying* (3.1)*. The manifold (strong) variational sufficient condition for problem* (1.1) *under the retraction $R_{\bar{x}}$ holds with respect to $(\bar{x}, \bar{y})$ at level $\rho > 0$ if and only if there exists a closed convex neighborhood $\mathcal{W} \times \mathcal{Y}$ of $(0_{\bar{x}}, \bar{y})$ such that*

(i) *$l^{\rho}_{R_{\bar{x}}}(\xi, y)$ is locally (strongly) convex at $0_{\bar{x}}$ for every $y \in \mathcal{Y}$;*
(ii) *$l^{\rho}_{R_{\bar{x}}}(\xi, y)$ is concave in y for every $\xi \in \mathcal{W} \subseteq B_{\bar{x}}(r_{R_{\bar{x}}})$.*

Under these conditions, $(\bar{x}, \bar{y})$ is a saddle point of $l^{\rho}(x, y)$ with respect to minimization over $x \in R_{\bar{x}}(\mathcal{W})$ and maximization over $y \in \mathcal{Y}$. Moreover, these properties hold for any $\rho' \geq \rho$.

Proof By applying [52, Theorem 1] and Proposition 1 to (1.2) at $(0_{\bar{x}}, \bar{y})$, we obtain that $l^{\rho}_{R_{\bar{x}}}(\xi, y)$ is convex in $\xi \in \mathcal{W}$ for every $y \in \mathcal{Y}$, and concave in $y \in \mathcal{Y}$ for every $\xi \in \mathcal{W} \subseteq B_{\bar{x}}(r_{R_{\bar{x}}})$. It remains to show that $(\bar{x}, \bar{y})$ is a saddle point of $l^{\rho}(x, y)$ in $R_{\bar{x}}(\mathcal{W}) \times \mathcal{Y}$, or that $l^{\rho}(x, y)$ attains its minimum at $\bar{x}$ in $R_{\bar{x}}(\mathcal{W})$. This is shown by $l^{\rho}(\bar{x}, \bar{y}) = l^{\rho}_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y}) \leq l^{\rho}_{R_{\bar{x}}}\left(R_{\bar{x}}^{-1}(x), \bar{y}\right) = l^{\rho}(x, \bar{y})$ for any $x \in R_{\bar{x}}(\mathcal{W})$. Hence we complete the proof. □

In the Euclidean setting, a key feature of the strong variational sufficiency is its equivalence to the SSOSC. In our work, we further reveal the intrinsic relationship between the manifold strong variational sufficiency and the M-SSOSC; see Theorem 1.

For a differentiable function $f : \mathcal{M} \to \mathbb{R}$ with a locally Lipschitz continuous gradient and a given retraction $R_{\bar{x}}$, the Hessian bundle of $f_{R_{\bar{x}}}$ at $0_{\bar{x}}$ is defined as

$$\overline{\nabla}^2 f_{R_{\bar{x}}}(0_{\bar{x}}) := \left\{ H \;\middle|\; \exists\, \xi_k \to 0_{\bar{x}} \text{ with } \nabla^2 f_{R_{\bar{x}}}(\xi_k) \to H,\ \xi_k \in T_{\bar{x}}\mathcal{M} \right\},$$

where $\nabla^2 f_{R_{\bar{x}}}(\cdot)$ denotes the Euclidean Hessian of $f_{R_{\bar{x}}}(\cdot)$. For the augmented Lagrangian function $l^{\rho}_{R_{\bar{x}}}(\xi, y)$ in (4.4), consider any matrix $H \in \overline{\nabla}^2 l^{\rho}_{R_{\bar{x}}}(\xi, y)$. we partition H into four blocks $H = \begin{bmatrix} H_{\xi\xi} & H_{\xi y} \\ H_{y\xi} & H_{yy} \end{bmatrix}$, and define $\overline{\nabla}^2_{\xi\xi} l^{\rho}_{R_{\bar{x}}}(0_{\bar{x}}, y) := \left\{ H_{\xi\xi} \mid H \in \overline{\nabla}^2 l^{\rho}_{R_{\bar{x}}}(\xi, y) \right\}$. Moreover, $\mathcal{C}_{\theta,g}(x, y) = \mathcal{C}_{\theta, g_{R_x}}(0_x, y)$ holds if $\mathcal{C}_{\theta, g_{R_x}}(0_x, y) = \left\{ d \in \mathbb{Y} \mid \theta^{\downarrow}(g_{R_x}(0_x); d) = \langle d, y \rangle \right\}$. By leveraging [52, Theorem 3], we can now establish the connection between the manifold strong variational sufficiency and the M-SSOSC for problem (1.1), as stated in the following theorem.

Theorem 1 *Let $(\bar{x}, \bar{y}) \in \mathcal{M} \times \mathbb{Y}$ be a KKT pair of problem* (1.1) *satisfying* (3.1)*. Then, the manifold strong variational sufficient condition for problem* (1.1) *under the retraction $R_{\bar{x}}$ holds with respect to $(\bar{x}, \bar{y})$ at level $\rho > 0$ if and only if every matrix in $\overline{\nabla}^2_{\xi\xi} l^{\rho}_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y})$ is positive definite. Moreover, any $H_{\xi\xi} \in \overline{\nabla}^2_{\xi\xi} l^{\rho}_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y})$ can be written in the form*

$$H_{\xi\xi} = \operatorname{Hess}_x L(\bar{x}, \bar{y}) + Dg(\bar{x})^* G \operatorname{grad} g(\bar{x}) \quad \textit{for some} \quad G \in \overline{\nabla}^2 \operatorname{env}_{\rho} \theta \left(g(\bar{x}) + \rho^{-1} \bar{y} \right).$$

If θ is a polyhedral convex function, then the manifold strong variational sufficient condition is equivalent to the following M-SSOSC at $(\bar{x}, \bar{y})$:

$$\langle \xi, \operatorname{Hess}_x L(\bar{x}, \bar{y}) \xi \rangle > 0 \quad \forall\, Dg(\bar{x})\xi \in \operatorname{aff} \mathcal{C}_{\theta,g}(\bar{x}, \bar{y}) \setminus \{0\}.$$

Furthermore, if θ is the indicator function of a second-order cone or a positive semidefinite cone, then the manifold strong variational sufficient condition is equivalent to the M-SSOSC at $(\bar{x}, \bar{y})$ given in (3.12)*, where $\mathcal{K}$ denotes the respective cone.*

Proof By [52, Theorem 3], every $H_{\xi\xi} \in \overline{\nabla}^2_{\xi\xi} l^\rho_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y})$ is positive definite and of the form

$$H_{\xi\xi} = \nabla^2_{\xi\xi} l_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y}) + \nabla g_{R_{\bar{x}}}(0_{\bar{x}})^* G \nabla g_{R_{\bar{x}}}(0_{\bar{x}}) \text{ for some } G \in \overline{\nabla}^2 \operatorname{env}_\rho \theta \left(g_{R_{\bar{x}}}(0_{\bar{x}}) + \rho^{-1}\bar{y} \right).$$

Since $\operatorname{grad}_x L(\bar{x}, \bar{y}) = 0$, the definition of the retraction together with (2.1) shows that this representation of $H_{\xi\xi}$ coincides with the expression stated in the theorem.

If θ is polyhedral convex, [52, Theorem 4] shows that strong variational sufficiency holds if and only if

$$\left\langle \xi, \nabla^2_{\xi\xi} l_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y})\, \xi \right\rangle > 0 \quad \forall\ g'_{R_{\bar{x}}}(0_{\bar{x}})\xi \in \operatorname{aff} \mathcal{C}_{\theta, g_{R_{\bar{x}}}}(0_{\bar{x}}, \bar{y}) \setminus \{0\}.$$

This condition is equivalent to the M-SSOSC stated in the theorem, since

$$\nabla^2_{\xi\xi} l_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y}) = \operatorname{Hess}_x L(\bar{x}, \bar{y}), \quad g'_{R_{\bar{x}}}(0_{\bar{x}})\xi = Dg(\bar{x})\xi, \quad \mathcal{C}_{\theta, g_{R_{\bar{x}}}}(0_{\bar{x}}, \bar{y}) = \mathcal{C}_{\theta, g}(\bar{x}, \bar{y}).$$

Finally, if θ is the indicator function of a second-order cone or a positive semidefinite cone, the equivalence follows from [62] and the above analysis. Therefore the proof is completed. □

Remark 3 Theorem 1 shows that the matrices in the Hessian bundle of $l^\rho_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y})$ are independent of the particular choice of retraction $R_{\bar{x}}$. This fact can also be established directly. Given two retractions R_1 and R_2 at $\bar{x}$, it suffices to prove that the Hessian bundles $\overline{\nabla}^2_{\xi\xi} l^\rho_{R_1}(0_{\bar{x}}, \bar{y})$ and $\overline{\nabla}^2_{\xi\xi} l^\rho_{R_2}(0_{\bar{x}}, \bar{y})$ are identical. Specifically, for any $H \in \overline{\nabla}^2_{\xi\xi} l^\rho_{R_2}(0_{\bar{x}}, \bar{y})$, there exists a sequence $\{\xi_k\}$ converging to $0_{\bar{x}}$ such that $\nabla^2_{\xi\xi} l^\rho_{R_2}(\xi_k, \bar{y}) \to H$. Let $W_{\bar{x}} := R_1^{-1} R_2$. Then we have

$$\nabla^2_{\xi\xi} l^\rho_{R_2}(\xi_k, \bar{y}) = W'_{\bar{x}}(\xi_k)^* \nabla^2_{\xi\xi} l^\rho_{R_1}(W_{\bar{x}}(\xi_k), \bar{y}) \nabla W_{\bar{x}}(\xi_k) + \nabla(W'_{\bar{x}}(\xi_k)^*) \nabla^2_{\xi\xi} l^\rho_{R_1}(W_{\bar{x}}(\xi_k), \bar{y}).$$

It follows from the definition of $W_{\bar{x}}$ that when $\xi_k \to 0_{\bar{x}}$, $\nabla W_{\bar{x}}(\xi_k) \to \mathrm{id}$ and $\nabla(W'_{\bar{x}}(\xi_k)^*) \to 0_{\bar{x}}$. Therefore, if $\nabla^2_{\xi\xi} l^\rho_{R_2}(\xi_k, \bar{y}) \to H$, then there exists a sequence $\{W_{\bar{x}}(\xi_k)\}$ converge to $0_{\bar{x}}$ such that $\nabla^2_{\xi\xi} l^\rho_{R_1}(W_{\bar{x}}(\xi_k), \bar{y}) \to H$. By symmetry, exchanging R_1 and R_2 yields $\overline{\nabla}^2_{\xi\xi} l^\rho_{R_1}(0_{\bar{x}}, \bar{y}) = \overline{\nabla}^2_{\xi\xi} l^\rho_{R_2}(0_{\bar{x}}, \bar{y})$. Therefore, the Hessian bundle does not depend on the choice of retraction. Consequently, the manifold strong variational sufficient condition is an intrinsic property of manifold optimization problems, inherently independent of the choice of retraction.

Remark 4 Building on the above construction, we can introduce a pointwise retraction-based local convexity on manifolds. A function $f : \mathcal{M} \to (-\infty, \infty]$ is said to be (strongly) locally retraction-convex at $x \in \mathcal{M}$ with respect to a retraction R_x if f_{R_x} is locally (strongly) convex on the tangent space $T_x\mathcal{M}$. Likewise, for a pair $(x, z) \in \operatorname{gph} \partial f$, we say that f is variationally (strongly) convex at (x, z) under the retraction R_x when f_{R_x} enjoys Euclidean variational (strong) convexity at $(0_x, z) \in \operatorname{gph} \partial f_{R_x}$ on $T_x\mathcal{M} \times T_x\mathcal{M}$. This concept is closely related to the retraction-convexity introduced in [34, Definition 3.2], yet differs in the following respect: [34] imposes the property on a neighborhood of x, whereas we investigate it at the point x itself. Moreover, we establish that if f is continuously differentiable and f_{R_x} is prox-regular ([54, 13.27]), then the local strong retraction-convexity is invariant under the choice of retraction. Extending this invariance to weaker smoothness assumptions or to functions that are convex but not strongly convex remains an open question for future study.

Another important perturbation property that plays a crucial role in the convergence analysis of the ALM is augmented tilt stability. Tilt stability characterizes the sensitivity of the optimal solution set to small linear perturbations of the objective or the constraints, and thus serves as a fundamental measure of the robustness of an optimization problem. Recently, Rockafellar extended this concept to augmented tilt stability and employed it to derive local convergence rates of ALM for non-convex and nonsmooth problems in the Euclidean setting [53]. Inspired by this, we investigate augmented tilt stability for optimization on Riemannian manifolds, and in Section 6, we show how it underpins the convergence analysis of manifold algorithms. We now formalize the notion of augmented tilt stability on a manifold with respect to a chosen retraction and subsequently characterize it in terms of manifold strong variational sufficiency.

Definition 9 Within the framework of Proposition 7, which establishes the local convexity-concavity of $l^\rho(x,y)$ on the neighborhoods $\mathcal{W}\times\mathcal{Y}$, the manifold augmented tilt stability is said to hold at $\bar{x}\in\mathcal{M}$ with respect to the retraction $R_{\bar{x}}$ if there exists a neighborhood $\mathcal{V}$ of $0_{\bar{x}}$ such that the following mapping is single-valued and Lipschitz continuous:

$$(v,y)\mapsto \underset{x\in R_{\bar{x}}(\mathcal{W})}{\operatorname{argmin}}\left\{l^\rho(x,y)-\left\langle v,R_{\bar{x}}^{-1}(x)\right\rangle\right\}\quad\text{for}\quad (v,y)\in\mathcal{V}\times\mathcal{Y}.$$

Proposition 8 *The manifold strong variational sufficient condition holds at* $(\bar{x},\bar{y})$ *with respect to* $R_{\bar{x}}$ *for local optimality of problem* (1.1) *at level* ρ *if and only if the manifold augmented tilt stability holds at* $\bar{x}$ *with respect to* $R_{\bar{x}}$*.*

Proof For $(v,y)\in\mathcal{V}\times\mathcal{Y}$, denote

$$h(v,y):=\underset{x\in R_{\bar{x}}(\mathcal{W})}{\operatorname{argmin}}\left\{l^\rho(x,y)-\left\langle v,R_{\bar{x}}^{-1}(x)\right\rangle\right\}\quad\text{and}\quad h_{R_{\bar{x}}}(v,y):=\underset{\xi\in\mathcal{W}}{\operatorname{argmin}}\left\{l^\rho_{R_{\bar{x}}}(\xi,y)-\langle v,\xi\rangle\right\}.$$

If $x\in h(v,y)$, then $R_{\bar{x}}^{-1}(x)$ is a minimizer of $l^\rho_{R_{\bar{x}}}(\xi,y)-\langle v,\xi\rangle$, implying $h(v,y)\subseteq R_{\bar{x}}(h_{R_{\bar{x}}}(v,y))$. The converse inclusion holds as well, and thus $h(v,y)=R_{\bar{x}}(h_{R_{\bar{x}}}(v,y))$. Therefore, $h(v,y)$ is single-valued and Lipschitz continuous if and only if $h_{R_{\bar{x}}}(v,y)$ is single-valued and Lipschitz continuous. Thus, the augmented tilt stability holds at both $\bar{x}$ and $0_{\bar{x}}$. Finally, since $l^\rho_{R_{\bar{x}}}(\xi,\cdot)$ is concave in $\mathcal{Y}$ for each $\xi\in\mathcal{W}$, the desired result follows directly from [52, Theorem 2]. □

5 Application I: Riemannian sequential quadratic programming

Riemannian sequential quadratic programming (RSQP) extends the classical Euclidean sequential quadratic programming (SQP) to updates performed in the tangent spaces of a manifold. Classical Euclidean SQP generally enjoys strong theoretical guarantees and numerical efficiency. The foundational papers of Han [27] and Powell [48] proved local convergence under strict complementarity and the SOSC. For more details of Euclidean SQP methods, we refer the reader to [7]. In contrast, the theoretical development of RSQP is less mature. Schiela and Ortiz [55] analyzed RSQP for equality constrained problems on manifolds and proved local convergence. More recently, Obara, Okuno, and Takeda [43] extended the method to accommodate both equality and inequality constraints, demonstrating global convergence and, under manifold linear independence constraint qualification (M-LICQ), M-SOSC, and strict complementarity, attaining local quadratic convergence.

In this section, we investigate the local convergence rate of the RSQP using the perturbation results obtained in Section 3. In particular, we show that without requiring strict complementarity, RSQP attains superlinear local convergence under the M-SRCQ and M-SOSC conditions. We begin by outlining the RSQP framework for problem (1.1) in Algorithm 1.

Since M^k is symmetric positive semidefinite, the subproblem (5.1) is convex. And ξ solves (5.1) if there exists a multiplier η such that the pair (ξ,η) satisfies the KKT conditions

$$\operatorname{grad} f(x^k)+M^k\xi+Dg(x^k)^*\eta=0,\quad \eta\in\partial\theta(g(x^k)+Dg(x^k)\xi).\tag{5.3}$$

Algorithm 1 Riemannian sequential quadratic programming (RSQP) for solving (1.1)

Input: $(x^0, y^0) \in \mathcal{M} \times \mathbb{Y}$, $v \in (0, 1]$, a nonnegative sequence $\{\delta^k\}$ converging to zero. Set $k = 0$.

1: If (x^k, y^k) satisfies a suitable termination criterion: STOP.

2: Choose a symmetric positive semidefinite vector field M^k.

3: Find $\xi^k \in T_{x^k}\mathcal{M}$ as an approximate solution to the subproblem

$$\min \ Df(x^k)\xi + \frac{1}{2}\langle \xi, M^k \xi\rangle + \theta(g(x^k) + Dg(x^k)\xi) \quad \text{s.t.} \quad \xi \in T_{x^k}\mathcal{M}, \tag{5.1}$$

equivalently via finding (ξ^k, γ^k) satisfying the following system

$$\tilde{\delta}^k \in \begin{bmatrix} \operatorname{grad}_x L(x^k, y^k) \\ -g(x^k) \end{bmatrix} + \begin{bmatrix} M^k & Dg(x^k)^* \\ -Dg(x^k) & 0 \end{bmatrix} \begin{bmatrix} \xi \\ \gamma \end{bmatrix} + \begin{bmatrix} 0 \\ \partial\theta^*(y^k + \gamma^k) \end{bmatrix}, \tag{5.2}$$

where $\tilde{\delta}^k := \min \left\{ \delta^k, \left\| \begin{bmatrix} \operatorname{grad}_x L(x^k, y^k) \\ g(x^k) - \operatorname{prox}_\theta(g(x^k) + y^k) \end{bmatrix} \right\|^{1+v} \right\}$ controls the inexactness.

4: Compute $x^{k+1} = R_{x^k}(\xi^k)$ and $y^{k+1} = y^k + \gamma^k$.

5: Set $k = k + 1$ and go to step 1.

We now interpret the RSQP (Algorithm 1) as a Newton-type method [8] for solving the KKT system (3.1) of problem (1.1). Since θ is closed convex, by [49, Theorem 23.5], the KKT system (3.1) is equivalent to

$$0 \in \begin{bmatrix} \operatorname{grad}_x L(x, y) \\ -g(x) \end{bmatrix} + \begin{bmatrix} 0 \\ \partial\theta^*(y) \end{bmatrix}.$$

Applying the Newton-type methods in [8] to this system yields the iteration (5.2) with $\tilde{\delta}^k = 0$. By direct computation, one verifies that (ξ^k, γ^k) solves (5.2) with $\tilde{\delta}^k = 0$ if and only if $(\xi^k, y^k + \gamma^k)$ solve (5.3), i.e., ξ^k solves (5.1). Hence, the Newton iteration and the RSQP iteration are equivalent. Moreover, if $\theta = 0$ and M^k is chosen as the Hessian of f, Algorithm 1 reduces to a Riemannian Newton method. If M^k is taken as the generalized Jacobian of $\operatorname{grad} f$ and certain semismoothness conditions are met, Algorithm 1 becomes a semismooth Newton algorithm. This close connection between RSQP and Newton-type methods motivates the establishment of local superlinear convergence, as demonstrated in the following proposition.

Proposition 9 *Let f and g be twice continuously differentiable. Let $(\bar{x}, \bar{y}) \in \mathcal{M} \times \mathbb{Y}$ be a KKT pair of problem (1.1) satisfying (3.1) and suppose that the KKT solution mapping $\mathcal{S}_{\text{KKT}}$ in (3.5) is robust isolated calm (or strongly regular) at $(\bar{x}, \bar{y})$. Let $\{(\xi^k, \gamma^k, x^k, y^k)\}$ be the sequence generated by Algorithm 1, and define*

$$\overline{\mathcal{J}}^k := \begin{bmatrix} P_{\bar{x}x^k} \operatorname{Hess}_x L(\bar{x}, \bar{y}) & Dg(\bar{x})^* \\ -P_{\bar{x}x^k} Dg(\bar{x}) & 0 \end{bmatrix} \quad \text{and} \quad \mathcal{J}^k := \begin{bmatrix} M^k & Dg(x^k)^* \\ -Dg(x^k) & 0 \end{bmatrix}.$$

Assume $(x^k, y^k) \to (\bar{x}, \bar{y})$. If

$$(\overline{\mathcal{J}}^k - \mathcal{J}^k)[\xi^k; \gamma^k] = o(\|[\xi^k; \gamma^k]\|),$$

then $(x^k, y^k) \to (\bar{x}, \bar{y})$ converges superlinearly. Moreover, if

$$(\overline{\mathcal{J}}^k - \mathcal{J}^k)[\xi^k; \gamma^k] = O(\|[\xi^k; \gamma^k]\|^2),$$

and the parameter v in Algorithm 1 is set to $v = 1$, then $(x^k, y^k) \to (\bar{x}, \bar{y})$ converges quadratically.

Proof Let $\bar{z} = (\bar{x}, \bar{y})$, $z^k = (x^k, y^k)$, $\zeta^k = [\xi^k; \gamma^k]$, $\mu^k = (\overline{\mathcal{J}}^k - \mathcal{J}^k)\zeta^k$, and

$$r^k = \mu^k + \tilde{\delta}^k + \begin{bmatrix} P_{x^{k+1}x^k} \operatorname{grad}_x L(x^{k+1}, y^{k+1}) \\ -g(x^{k+1}) \end{bmatrix} - \begin{bmatrix} \operatorname{grad}_x L(x^k, y^k) \\ -g(x^k) \end{bmatrix} - \overline{\mathcal{J}}^k \zeta^k.$$

Then (5.2) can be equivalently written as

$$r^k \in \begin{bmatrix} P_{x^{k+1}x^k} \operatorname{grad}_x L(x^{k+1}, y^{k+1}) \\ -g(x^{k+1}) \end{bmatrix} + \begin{bmatrix} 0 \\ \partial\theta^*(y^{k+1}) \end{bmatrix}. \tag{5.4}$$

Since $L(x, y) = f(x) + \langle g(x), y\rangle$ is twice continuously differentiable in x, and $\tilde{\delta}^k = o(d(z^k, \bar{z}))$ due to the local Lipschitz property of both $\operatorname{grad}_x L(x, y)$ and $\operatorname{prox}_\theta(g(x) + y)$ with respect to (x, y) at $\bar{z}$, it follows that, for sufficiently large k, $r^k = \mu^k + o(\|\zeta^k\|) + o(d(z^k, \bar{z}))$. By the condition that $\mu^k = o(\|\zeta^k\|)$, we have

$$\begin{aligned} & d(z^{k+1}, \bar{z}) = O(r^k) = o(\|\zeta^k\|) + o(d(z^k, \bar{z})) \\ =& o(d(z^{k+1}, z^k)) + o(d(z^k, \bar{z})) = o(d(z^{k+1}, \bar{z}) + d(z^k, \bar{z})) + o(d(z^k, \bar{z})). \end{aligned}$$

The first equality follows from the isolated calmness of $\bar{z}$ to the perturbed KKT system (5.4), and the third equality uses the fact that R_{x^k} is a local homomorphism. Hence $d(z^{k+1}, \bar{z}) = o(d(z^k, \bar{z}))$, which establishes superlinear convergence of $\{z^k\}$.

Now suppose $(\overline{\mathcal{J}}^k - \mathcal{J}^k)[\xi^k; \gamma^k] = O(\|[\xi^k; \gamma^k]\|^2)$. First, $z^k \to \bar{z}$ converges superlinearly implies that $d(z^{k+1}, z^k)/d(z^k, \bar{z}) \to 1$. By the assumption that f and g are twice continuously differentiable, the mapping $(x, y) \mapsto \begin{bmatrix} \operatorname{Hess}_x L(x, y) & Dg(x)^* \\ -Dg(x) & 0 \end{bmatrix}$ is locally Lipschitz at $\bar{z}$, with some Lipschitz constant $c > 0$. For k sufficiently large, we have

$$\begin{aligned} & \left\| \begin{bmatrix} P_{x^{k+1}x^k} \operatorname{grad}_x L(z^{k+1}) \\ -g(x^{k+1}) \end{bmatrix} - \begin{bmatrix} \operatorname{grad}_x L(z^k) \\ -g(x^k) \end{bmatrix} - \overline{\mathcal{J}}^k \zeta^k \right\| \\ \le & \left\| \begin{bmatrix} P_{x^{k+1}x^k} \operatorname{grad}_x L(z^{k+1}) \\ -g(x^{k+1}) \end{bmatrix} - \begin{bmatrix} \operatorname{grad}_x L(z^k) \\ -g(x^k) \end{bmatrix} - \begin{bmatrix} P_{x^{k+1}x^k} \operatorname{Hess}_x L(z^{k+1}) & Dg(x^{k+1})^* \\ -Dg(x^{k+1}) & 0 \end{bmatrix} \zeta^k \right\| \\ & + \left\| \begin{bmatrix} P_{x^{k+1}x^k} \operatorname{Hess}_x L(z^{k+1}) & Dg(x^{k+1})^* \\ -Dg(x^{k+1}) & 0 \end{bmatrix} - \overline{\mathcal{J}}^k \right\| \|\zeta^k\| \\ \le & O(\|\zeta^k\|^2) + c\|\zeta^k\| d(z^k, \bar{z}). \end{aligned}$$

Again, applying the isolated calmness of $\bar{z}$ to the perturbed KKT system (5.4), and noting that $v = 1$ implies that $\tilde{\delta}^k = O(d(z^k, \bar{z})^2)$, we then obtain that

$$d(z^{k+1}, \bar{z}) = O(r^k) = O(\|\zeta^k\|^2) + L\|\zeta^k\| d(z^k, \bar{z}) + O(d(z^k, \bar{z})^2).$$

Since $d(z^{k+1}, z^k)/d(z^k, \bar{z}) \to 1$ and R_{x^k} is a local homomorphism, we conclude that $d(z^{k+1}, \bar{z}) = O(d(z^{k+1}, z^k)^2) = O(d(z^k, \bar{z})^2)$. The proof is then completed. $\square$

To satisfy the hypotheses of Proposition 9, the following notion of hemistability was introduced by Bonnans [8, Definition 2.2] under the Euclidean setting. We generalize this concept to Riemannian manifolds and, to accommodate inexact computations, formulate a perturbed hemistability condition that covers inexact subproblem solves and data perturbations in the following definition.

Definition 10 A point $\bar{z} = (\bar{x}, \bar{y})$ is called a hemistable solution of the KKT system (3.1) if the following holds: for any $\alpha > 0$, there exist constants $\bar{\delta} > 0$ and $\varepsilon > 0$ such that, for every $\hat{z} = (\hat{x}, \hat{y}) \in \mathcal{M} \times \mathbb{Y}$ and every symmetric positive semidefinite tangent vector $M \in T_{\hat{x}}\mathcal{M}$ satisfying

$$d(\hat{z}, \bar{z}) + \left\| \begin{bmatrix} P_{\bar{x}\hat{x}} \operatorname{Hess}_x L(\bar{x}, \bar{y}) & Dg(\bar{x})^* \\ -P_{\bar{x}\hat{x}} Dg(\bar{x}) & 0 \end{bmatrix} - \begin{bmatrix} M & Dg(\hat{x})^* \\ -Dg(\hat{x}) & 0 \end{bmatrix} \right\| < \varepsilon,$$

the following system

$$\delta \in \begin{bmatrix} \operatorname{grad}_x L(\hat{x}, \hat{y}) \\ -g(\hat{x}) \end{bmatrix} + \begin{bmatrix} M & Dg(\hat{x})^* \\ -Dg(\hat{x}) & 0 \end{bmatrix} \begin{bmatrix} \xi \\ \gamma \end{bmatrix} + \begin{bmatrix} 0 \\ \partial\theta^*(\hat{y} + \gamma) \end{bmatrix}, \quad \text{for any } \delta \in [0, \bar{\delta}],$$

admits a solution (ξ, γ) such that $z = \big(R_{\hat{x}}(\xi),\, \hat{y} + \gamma\big)$ satisfies $d(z, \bar{z}) \le \alpha$.

Now we are able to present the following local convergence results of Algorithm 1, when M^k is chosen as the Riemannian Hessian of the Lagrangian function $L(x^k, y^k)$ at each step.

Theorem 2 (Local convergence of RSQP) *Let f and g be twice continuously differentiable. Let $\bar{z} = (\bar{x}, \bar{y}) \in \mathcal{M} \times \mathbb{Y}$ be a KKT pair of problem* (1.1) *satisfying* (3.1)*. Assume that $\bar{z}$ is a hemistable solution of* (3.1)*, the M-SOSC* (3.11) *holds at $\bar{x}$, and the M-SRCQ* (3.9) *holds at $\bar{x}$ with respect to $\bar{y}$. Choose $M^k = \operatorname{Hess}_x L(x^k, y^k)$ in Algorithm 1. Let $\{(x^k, y^k)\}$ be the sequence generated by Algorithm 1, then $\{(x^k, y^k)\}$ converges superlinearly to $(\bar{x}, \bar{y})$. Moreover, if the parameter v in Algorithm 1 is set to $v = 1$, then the convergence is quadratic.*

Proof By Proposition 3, M-SOSC and M-SRCQ hold at $\bar{z}$ implies that $\mathcal{S}_{\mathrm{KKT}}$ is robust isolated calm at $\bar{z}$. Therefore, given $\delta = (\hat{a}, b)$ in a neighborhood of $(0_{\bar{x}}, 0)$, there exist $\varepsilon_1, \kappa > 0$ such that for any $z \in \mathcal{S}_{\mathrm{KKT}}(\delta)$ and $d(z, \bar{z}) \le \varepsilon_1$, we have $d(z, \bar{z}) \le \kappa\|\delta\|$. Now choose $\alpha \le \min(\varepsilon_1, 1/3\kappa)$. Taking $M^k = \operatorname{Hess}_x L(x^k, y^k)$, the hemistability of $\bar{z}$ yields that for the given α, there exist $\varepsilon_2 \ge 0$ and $c > 0$ such that for any $\tilde{\delta}^k \le c$ and for z^k satisfying

$$d(z^k, \bar{z}) + \|\overline{\mathcal{J}}^k - \mathcal{J}^k\| < \varepsilon_2,$$

the system

$$\tilde{\delta}^k \in \begin{bmatrix} \operatorname{grad}_x L(x^k, y^k) \\ -g(x^k) \end{bmatrix} + \mathcal{J}^k \begin{bmatrix} \xi \\ \gamma \end{bmatrix} + \begin{bmatrix} 0 \\ \partial\theta^*(y^k + \gamma^k) \end{bmatrix}$$

has a solution (ξ^k, γ^k) such that $z^{k+1} = (R_{x^k}(\xi^k), y^k + \gamma^k)$ and $d(z^{k+1}, \bar{z}) \le \alpha$. Now let $\zeta^k = [\xi^k; \gamma^k]$, and denote

$$\eta^k := \tilde{\delta}^k + \begin{bmatrix} P_{x^{k+1}x^k} \operatorname{grad}_x L(x^{k+1}, y^{k+1}) \\ -g(x^{k+1}) \end{bmatrix} - \begin{bmatrix} \operatorname{grad}_x L(x^k, y^k) \\ -g(x^k) \end{bmatrix} - \mathcal{J}^k \zeta^k,$$

then we have

$$\eta^k \in \begin{bmatrix} P_{x^{k+1}x^k} \operatorname{grad}_x L(x^{k+1}, y^{k+1}) \\ -g(x^{k+1}) \end{bmatrix} + \begin{bmatrix} 0 \\ \partial\theta^*(y^{k+1}) \end{bmatrix}.$$

By the homomorphism of R_{x^k} and its local Lipschitz continuity, there exists $\mu > 0$, such that $\|\xi^k\| \le \mu d(R_{x^k}(\xi^k), x^k)$. Since $\tilde{\delta}^k = o(d(z^k, \bar{z}))$, there exists $\varepsilon_3 > 0$ such that $\|\tilde{\delta}^k\| \le \frac{1}{6\kappa} d(z^k, \bar{z})$ for $d(z^k, \bar{z}) \le \varepsilon_3$. Shrinking α, ε_2 and ε_3 if necessary, the twice continuous differentiability of $L(x, y)$ ensures that $\|\eta^k\| \le \frac{1}{3\kappa\mu} \|\zeta^k\| + \|\tilde{\delta}^k\|$. Therefore, by the isolated calmness, we have

$$d(z^{k+1}, \bar{z}) \le \kappa \|\eta^k\| \le \frac{1}{3\mu} \|\zeta^k\| + \kappa\|\tilde{\delta}^k\| \le \frac{1}{3} d(z^{k+1}, z^k) + \frac{1}{6} d(z^k, \bar{z}) \le \frac{1}{3} d(z^{k+1}, \bar{z}) + \frac{1}{2} d(z^k, \bar{z}),$$

which implies $d(z^{k+1}, \bar{z}) \le \frac{3}{4} d(z^k, \bar{z})$ and thus $\{z^k\} \to \bar{z}$. Now the local convergence rate can be directly obtained by using Proposition 9. ☐

Theorem 2 establishes convergence results when $M^k = \operatorname{Hess}_x L(x^k, y^k)$. In fact, the result extends to broader choices of the approximate Hessians M^k. For example, M^k can be generated by quasi-Newton updates, provided the sequence remains uniformly positive definite and satisfies standard secant conditions. The Euclidean analysis is presented in [8], and the same arguments extend to the Riemannian setting by working on tangent spaces via retraction. For brevity, we omit the routine details.

In [8], it was shown that the robust isolated calmness of the KKT solution mapping implies hemistability under the polyhedral assumption of θ with $\delta = 0$. This result was later extended in [41] to the case where θ is parabolically regular, a property introduced in [54, Definition 13.65] and shown in [42] to include the $\mathcal{C}^2$-cone reducible case. In the Riemannian setting, we establish that under the parabolic regularity assumption of θ, the robust isolated calmness of the KKT solution mapping still implies hemistability. Furthermore, our result strengthens [41, Theorem 5.1] by allowing for inexact cases.

Proposition 10 *Let f and g be twice continuously differentiable and θ be parabolically regular. Let $(\bar{x}, \bar{y}) \in \mathcal{M} \times \mathbb{Y}$ be a KKT pair of problem* (1.1) *satisfying* (3.1). *Assume that the M-SOSC* (3.11) *holds at $\bar{x}$ and the M-SRCQ* (3.9) *holds at $\bar{x}$ with respect to $\bar{y}$. Then $(\bar{x}, \bar{y})$ is a hemistable solution of* (3.1).

Proof It can be verified that $(0_{\bar{x}}, \bar{y})$ is a solution of the KKT system (5.3) associated with the RSQP subproblem (5.1) at $\bar{x}$. Moreover, the SRCQ condition for (5.3) at this point takes the form

$$Dg(\bar{x})T_{\bar{x}}\mathcal{M} + \mathcal{T}_{\mathrm{dom}(\theta)}(g(\bar{x})) \cap \bar{y}^{\perp} = \mathbb{Y},$$

which coincides with the M-SRCQ condition for problem (1.1) at $(\bar{x}, \bar{y})$. Hence, the SRCQ condition holds under the M-SRCQ assumption. Likewise, the SOSC condition for (5.3) is also satisfied at $(0_{\bar{x}}, \bar{y})$, as it has the same structure as the M-SOSC condition for (1.1) at $(\bar{x}, \bar{y})$. Given that θ is parabolically regular, we can apply [41, Theorem 4.2] and obtain that for any neighborhood $\mathcal{U}_1$ of $\bar{z}$, there exists a neighborhood $\mathcal{U}_2$ of $\bar{z}$ such that for any $\hat{z} \in \mathcal{U}_2$, the perturbed KKT system (5.4) has a solution within $\mathcal{U}_1$ when δ is sufficiently small. This confirms the hemistability of the system at $\bar{z}$. □

By combining Theorem 2 and Proposition 10, we are now ready to present the main result of this section.

Proposition 11 *Let f and g be twice continuously differentiable and θ be parabolically regular. Let $(\bar{x}, \bar{y}) \in \mathcal{M} \times \mathbb{Y}$ be a KKT pair of problem* (1.1) *satisfying* (3.1). *Assume that the M-SOSC* (3.11) *holds at $\bar{x}$ and the M-SRCQ* (3.9) *holds at $\bar{x}$ with respect to $\bar{y}$. Choose $M^k = \operatorname{Hess}_x L(x^k, y^k)$ in Algorithm 1. Let $\{(x^k, y^k)\}$ be the sequence generated by Algorithm 1, then (x^k, y^k) converges superlinearly (quadratically if $v = 1$) to $(\bar{x}, \bar{y})$.*

Remark 5 This result relaxes the assumptions in [43], where quadratic convergence of RSQP was established under M-LICQ (equivalent to manifold constraint nondegeneracy in NLP on Riemannian manifolds) together with M-SOSC. In contrast, our analysis requires the M-SRCQ, which is a weaker requirement compared with M-LICQ, together with M-SOSC. To illustrate this, we provide a toy example in which M-SRCQ holds but manifold constraint nondegeneracy (and thus M-LICQ) fails. Consider the optimization on the sphere $\mathcal{S}^1$:

$$\begin{array}{ll} \min & \|x\|_2^2 \\ \text{s.t.} & x \in \mathcal{P} = \{x_1 + x_2 \geq 1,\ x_1 - x_2 \leq 1,\ x_2 \geq 0\}, \\ & x_1^2 + x_2^2 = 1. \end{array}$$

It can be verified that $x = (1,0)^\top$, $y = (0,0)^\top$ is a KKT pair. Now let $g(x) := x$. By definitions (3.9) and (3.10), the M-SRCQ is said to hold at (x, y) if

$$Dg(x)T_x\mathcal{S}^1 + \mathcal{T}_{\mathcal{P}}(g(x)) \cap y^\perp = \mathbb{R}^2,$$

and the manifold constraint nondegeneracy holds at x if

$$Dg(x)T_x\mathcal{S}^1 + \mathrm{lin}(\mathcal{T}_{\mathcal{P}}(g(x))) = \mathbb{R}^2.$$

Since $\mathcal{T}_{\mathcal{P}}(g(x)) = \{\xi \mid \xi_2 \geq |\xi_1|\}$, we have $\mathrm{lin}(\mathcal{T}_{\mathcal{P}}(g(x))) = 0$. Combining that $Dg(x)T_x\mathcal{S}^1 = T_x\mathcal{S}^1 = \{\xi \mid x^\top\xi = 0\} = \{\xi \mid \xi_1 = 0,\ \xi_2 \in \mathbb{R}\}$, it is easy to see that the manifold constraint nondegeneracy does not hold at $x = (1,0)^\top$. However, since $y = (0,0)^\top$ and $y^\perp = \mathbb{R}^2$, this implies that

$$Dg(x)T_x\mathcal{S}^1 + \mathcal{T}_{\mathcal{P}}(g(x)) \cap y^\perp = \{\xi \mid \xi_1 = 0,\ \xi_2 \in \mathbb{R}\} + \{\xi \mid \xi_2 \geq |\xi_1|\} = \mathbb{R}^2,$$

which means that the M-SRCQ holds at this point.

We now discuss the RSQP algorithm and compare it with the ManPG [17] and AManPG [33] methods. Consider the case where $\mathcal{M}$ is an embedded submanifold and g is the identity mapping, i.e., $g(x) = x$ for all $x \in \mathcal{M}$. In this case, the RSQP subproblem (5.1) simplifies to

$$\min\ Df(x^k)\xi + \frac{1}{2}\langle \xi, M^k\xi\rangle + \theta(x^k + \xi) \quad \text{s.t.} \quad \xi \in T_{x^k}\mathcal{M}.$$

This formulation closely resembles that in [33, Algorithm 1]. However, the ManPG and AManPG guarantee convergence without convergence rates. In contrast, our RSQP (Algorithm 1) achieves at least local superlinear convergence rate under mild assumptions (M-SRCQ and M-SOSC), as established in Theorem 2 and Proposition 11, without requiring acceleration techniques. Furthermore, the RSQP bears structural resemblance to the ManPQN algorithm [61], but with a critical distinction: our update matrix M^k approximates the Hessian of the Lagrangian function, rather than merely $\mathrm{Hess}\, f(x)$. This adjustment enables RSQP to achieve superlinear convergence rate, outperforming the linear convergence rate of ManPQN. Additionally, we note that the recently proposed Riemannian proximal Newton method (RPN) [56] also attains superlinear convergence. However, RPN follows a two-stage procedure: it first computes a proximal gradient direction via ManPG, followed by a Riemannian Newton update. In contrast, RSQP achieves superlinear convergence in a single stage, offering a simpler and more direct approach.

6 Application II: Riemannian augmented Lagrangian method

Building on the stability results obtained in previous sections, especially the manifold (strong) variational sufficiency developed in Section 4, we will analyze the local convergence of the Riemannian Augmented Lagrangian Method (RALM) in this section as an application.

The classical Euclidean ALM was initially proposed by Hestenes [28] and Powell [47] for equality constrained problems and later extended to NLP by Rockafellar [50]. Over the decades, the convergence analysis of ALM in the Euclidean setting has been extensively studied. The RALM for nonsmooth manifold optimization was recently introduced in [64] and becomes a powerful tool for solving constrained optimization problems on manifolds (see also [20] and [39]). In [39], the convergence of the ALM to a KKT point is established under the linear independence constraint qualification. Similarly, in [64], the convergence of the RALM to a KKT point is established under suitable conditions, together with numerical experiments. Recent advancements, such as those in [63] and [4], have introduced approximate KKT conditions and reformulated constraint qualifications to enhance convergence guarantees. Regarding the local convergence analysis of the RALM, [6] established the Q-linear convergence rate of RALM under the M-SOSC and M-SRCQ conditions, which implies the uniqueness of the multiplier.

Prior analyses often relied on the uniqueness of the multiplier, which simplifies the analysis but limits applicability when multiple multipliers exist. Inspired by recent Euclidean advances [53], where a local dual problem under variational sufficiency and a proximal point framework yield Q-linear rates for PPA and, in turn, R-linear rates for ALM, we may remove the assumption of unique multiplier in the manifold setting by employing the manifold strong variational sufficiency for local optimality proposed in Section 4.

6.1 Convergence analysis under manifold variational sufficiency

In this subsection, we will discuss the convergence of the RALM for solving problem (1.1) when the manifold variational sufficiency holds, see Definition 8 and Proposition 7. We begin by presenting the RALM framework for problem (1.1) in Algorithm 2.

Algorithm 2 Riemannian augmented Lagrangian method (RALM) for solving (1.1)

Input: $(x^0, y^0) \in \mathcal{M} \times \mathbb{Y}$, $\rho_0 > \rho$. Set $k = 0$.

1: If (x^k, y^k) satisfies a suitable termination criterion: STOP.

2: Find an approximate solution x^{k+1}

$$x^{k+1} \approx \bar{x}^{k+1} = \operatorname{argmin}\, l^{\rho_k}(x, y^k) \quad \text{s.t.} \quad x \in R_{\bar{x}}(\mathcal{W}). \tag{6.1}$$

3: Update the multiplier $y^{k+1} = y^k + \tilde{\rho}_k \nabla_y l^{\rho_k}(x^{k+1}, y^k)$, with $\tilde{\rho}_k = \rho_k - \rho$.

4: Update $\rho_{k+1} \geq \rho_k$.

5: Set $k = k + 1$ and go to step 1.

When using the tangent-space formulation, the RALM can also be written as

$$\begin{cases} \xi^{k+1} \approx \bar{\xi}^{k+1} := \operatorname{argmin}\, l^{\rho_k}_{R_{\bar{x}}}(\xi, y^k) \quad \text{s.t.} \quad \xi \in \mathcal{W}, \\ x^{k+1} = R_{\bar{x}}(\xi^{k+1}), \\ y^{k+1} = y^k + \tilde{\rho}_k \nabla_y l^{\rho_k}_{R_{\bar{x}}}(\xi^{k+1}, y^k). \end{cases} \tag{6.2}$$

The first iteration of (6.2) for finding ξ^{k+1} can be considered as a traditional inexact ALM step in the tangent space, and the second iteration pulls ξ^{k+1} back to the manifold using the retraction $R_{\bar{x}}$. Now we begin to analyze the convergence of the RALM (Algorithm 2).

For subproblem (6.1), three conditions for measuring the inexactness of x^{k+1} are given in [53, (1.15)]

$$\left(2\tilde{\rho}_k \left[l^{\rho_k}(x^{k+1}, y^k) - \inf_{x \in R_{\bar{x}}(\mathcal{W})} l^{\rho_k}(x, y^k)\right]\right)^{\frac{1}{2}} \leq \begin{cases} \text{(a)} \quad \varepsilon_k, \\ \text{(b)} \quad \varepsilon_k \min\{1, \|\tilde{\rho}_k \nabla_y l^{\rho_k}(x^{k+1}, y^k)\|\}, \\ \text{(c)} \quad \varepsilon_k \min\left\{1, \|\tilde{\rho}_k \nabla_y l^{\rho_k}(x^{k+1}, y^k)\|^2\right\}. \end{cases} \tag{6.3}$$

Here, $\varepsilon_k \in (0,1)$ for all $k \geq 0$, and $\sum_{k=0}^{\infty} \varepsilon_k = \sigma < \infty$. As a counterpart, for (6.2), three conditions for measuring the inexactness of ξ^{k+1} are given by

$$\left(2\tilde{\rho}_k \left[l_{R_{\bar{x}}}^{\rho_k}(\xi^{k+1}, y^k) - \inf_{\xi \in \mathcal{W}} l_{R_{\bar{x}}}^{\rho_k}(\xi, y^k) \right] \right)^{\frac{1}{2}} \leq \begin{cases} (\mathrm{a}') & \varepsilon_k, \\ (\mathrm{b}') & \varepsilon_k \min\{1, \|\tilde{\rho}_k \nabla_y l_{R_{\bar{x}}}^{\rho_k}(\xi^{k+1}, y^k)\|\}, \\ (\mathrm{c}') & \varepsilon_k \min\left\{1, \|\tilde{\rho}_k \nabla_y l_{R_{\bar{x}}}^{\rho_k}(\xi^{k+1}, y^k)\|^2\right\}. \end{cases} \tag{6.4}$$

Algorithm 2 under the stopping criteria (6.3) is equivalent to the iteration (6.2) under the stopping criteria (6.4).

We now present the convergence analysis of Algorithm 2 under the manifold variational sufficiency. A point $\hat{x} \in \mathcal{M}$ is called a local optimal solution of (1.1) over $\mathcal{U} \subseteq \mathcal{M}$ if $f(\hat{x}) + \theta(g(\hat{x})) \leq f(x) + \theta(g(x))$ for any $x \in \mathcal{U}$. Following the approach of [53], the main idea is to first establish the local augmented duality for problem (1.1), and then apply the PPA to the local dual problem to derive the convergence of RALM.

Proposition 12 *Let $(\bar{x}, \bar{y}) \in \mathcal{M} \times \mathbb{Y}$ be a KKT pair of problem* (1.1) *satisfying* (3.1)*. Assume that the manifold variational sufficient condition for problem* (1.1) *under the retraction $R_{\bar{x}}$ holds with respect to $(\bar{x}, \bar{y})$ at level $\rho > 0$. Let $\mathcal{W}$ and $\mathcal{Y}$ be the closed convex neighborhoods in Proposition 7. Suppose that the set* $\operatorname{argmin}_{\xi \in \mathcal{W}} l_{R_{\bar{x}}}^{\rho}(\xi, y)$ *is nonempty and bounded for any $y \in \operatorname{int} \mathcal{Y}$. Denote $Z := \arg\max_{y \in \mathcal{Y}} \inf_{x \in R_{\bar{x}}(\mathcal{W})} l^{\rho}(x, y)$. Assume there exists $p > \operatorname{dist}(y^0, Z) + \sigma$ such that $\{y \mid \|y - y^0\| < 3p\} \subset \mathcal{Y}$. Let $\{(x^k, y^k)\}$ be the sequence generated by Algorithm 2 under* (6.3)*. Then $\{x^k\} \subseteq R_{\bar{x}}(\mathcal{W})$ is bounded, and every accumulation point of $\{x^k\}$ is a local optimal solution of problem* (1.1) *over $R_{\bar{x}}(\mathcal{W})$.*

Proof In fact, the update of (ξ^k, y^k) in (6.2) corresponds to a standard Euclidean ALM iteration. This structure allows us to analyze its convergence using a dual approach. Define the function $H(y) := \inf_{x \in R_{\bar{x}}(\mathcal{W})} l^{\rho}(x, y)$ for $y \in \mathcal{Y}$ and $H(y) := -\infty$ otherwise. Then it holds that $H_{R_{\bar{x}}}(y) := \inf_{\xi \in \mathcal{W}} l_{R_{\bar{x}}}^{\rho}(\xi, y)$ for $y \in \mathcal{Y}$. The local primal and dual problems associated with (1.2) are given by $\min_{\xi \in \mathcal{W}} \ \sup_{y \in \mathcal{Y}} l_{R_{\bar{x}}}^{\rho}(\xi, y)$ and $\max_{y \in \mathcal{Y}} \ H_{R_{\bar{x}}}(y)$, respectively. By applying the proof of [53, Theorem 2.1] to these local problems at a saddle point $(0_{\bar{x}}, \bar{y})$ of $l_{R_{\bar{x}}}^{\rho}$ on $\mathcal{W} \times \mathcal{Y}$, we conclude that both optimal values are attained at $(0_{\bar{x}}, \bar{y})$, and that $0_{\bar{x}}$ is the solution to (1.2) over $\mathcal{W}$.

We now apply the PPA to the local dual problem. The solution set of the dual problem is denoted by $Z = \arg\max_{y \in \mathcal{Y}} H_{R_{\bar{x}}}(y) = \arg\max_{y \in \mathcal{Y}} H(y)$. Then for $c_k > 0$, the PPA iteration is

$$y^{k+1} \approx \bar{y}^{k+1} = \operatorname*{argmax}_{y} \left\{ H_{R_{\bar{x}}}^{k}(y) := H_{R_{\bar{x}}}(y) - \frac{1}{2c_k} \left\| y - y^k \right\|^2 \right\}.$$

The stopping criteria for PPA follow the standard approximate conditions as specified in [51, (1.5), (1.6)]. By applying [53, Theorem 2.3] to the update of (ξ^k, y^k) in (6.2), we conclude that $\{y^k\}$ corresponds to an inexact PPA sequence with $c_k = \tilde{\rho}_k$, and the stopping criteria for PPA are related to (6.4). Consequently, $\{y^k\}$ converges to a local maximizer of $H_{R_{\bar{x}}}(y)$ over $y \in \mathcal{Y}$. Moreover, the sequence $\{\xi^{k+1}\}$ is bounded, and its accumulation points form solutions of the local primal problem. Finally, by the relationships that $x^{k+1} = R_{\bar{x}}(\xi^{k+1})$ and $l_{R_{\bar{x}}}^{\rho_k}(\xi^{k+1}, y^k) = l^{\rho_k}(x^{k+1}, y^k)$, the desired result follows. □

Remark 6 We should emphasize that the PPA used in the proof of Proposition 12 differs from the one developed in [25]. In [25], the authors proposed a PPA on Hadamard manifolds and analyzed its convergence under geodesic convexity assumptions. In contrast, the PPA considered here is constructed for the local dual problem around the KKT point $(0_{\bar{x}}, \bar{y})$ and is solely used to analyze the convergence behavior of RALM. It is essentially a standard Euclidean algorithm and cannot be directly applied to solve the original Riemannian problem (1.1). Rather, it serves as a theoretical tool in our convergence analysis.

Corollary 1 *Let $(\bar{x},\bar{y}) \in \mathcal{M}\times\mathbb{Y}$ be a KKT pair of problem* (1.1) *satisfying* (3.1). *Assume that the manifold strong variational sufficient condition for problem* (1.1) *under the retraction $R_{\bar{x}}$ holds with respect to $(\bar{x},\bar{y})$ at level $\rho > 0$. Then the sequence $\{x^k\}$ generated by Algorithm 2 converges to $\bar{x}$.*

Proof The strong sufficiency refers to the isolated minimizing property of $\bar{x}$. By Proposition 12, x^k will converge to the unique solution $\bar{x}$. □

Now we assume that the manifold strong variational sufficiency holds at $(\bar{x},\bar{y})$ in the remaining part of this section. The locally strong convexity of $l^{\rho_k}_{R_{\bar{x}}}(\cdot,y^k)$ implies that each RALM subproblem (6.1) has a unique solution. Next we give the local convergence rate of $\{x^k\}$ generated by Algorithm 2.

Theorem 3 (Local convergence of RALM) *Let $(\bar{x},\bar{y}) \in \mathcal{M}\times\mathbb{Y}$ be a KKT pair of problem* (1.1) *satisfying* (3.1). *Assume that the manifold strong variational sufficient condition for problem* (1.1) *under the retraction $R_{\bar{x}}$ holds with respect to $(\bar{x},\bar{y})$ at level $\rho > 0$. Let $\{(x^k,y^k)\}$ be the sequence generated by Algorithm 2 under* (6.3). *Under the assumptions in Proposition 12, $\{y^k\}$ converges to $\bar{y}$, and consequently $x^k \to \bar{x}$. Moreover, if $y^k \to \bar{y}$ converges Q-linearly, and $\|\operatorname{grad}_x l^{\rho_k}(x^{k+1},y^k)\| \le c\|y^{k+1}-y^k\|$ for some $c > 0$, then $x^k \to \bar{x}$ also converges R-linearly at the same rate.*

Proof Similar to the proof of Proposition 12, by applying [53, Theorem 3.1] to the update of (ξ^k,y^k) in (6.2) and using $x^{k+1} = R_{\bar{x}}(\xi^{k+1})$, we obtain the convergence $x^k \to \bar{x}$. According to the proof of [52, Theorem 2], the strong convexity of the function $l^{\rho}_{R_{\bar{x}}}(\cdot,y)$ implies that the mapping $\lambda(v,y) := \operatorname{argmin}_{\xi\in\mathcal{W}}\{l^{\rho}_{R_{\bar{x}}}(\xi,y) - \langle v,\xi\rangle\}$ is Lipschitz continuous with modulus s^{-1}. Since $0_{\bar{x}} = \lambda(0_{\bar{x}},\bar{y})$ and $\xi^{k+1} = \lambda(\nabla_\xi l^{\rho_k}_{R_{\bar{x}}}(\xi^{k+1},y^k),y^k)$, the Lipschitz continuity of λ yields

$$d(x^{k+1},\bar{x}) = \|\xi^{k+1} - 0_{\bar{x}}\| \le (1/s)(\|\nabla_\xi l^{\rho_k}_{R_{\bar{x}}}(\xi^{k+1},y^k)\| + \|y^k - \bar{y}\|).$$

Note that $x^{k+1} = R_{\bar{x}}(\xi^{k+1})$ and the sequence $\{\xi^k\}$ remains in the closed set $\mathcal{W}$. There exists $L > 0$ such that

$$\|\nabla_\xi l^{\rho_k}_{R_{\bar{x}}}(\xi^{k+1},y^k)\| \le L\|\operatorname{grad}_x l^{\rho_k}(x^{k+1},y^k)\| \le Lc\|y^{k+1}-y^k\|. \tag{6.5}$$

Hence, we further obtain

$$s^2 d^2(x^{k+1},\bar{x}) \le L^2c^2\|y^{k+1}-y^k\|^2 + \|y^k-\bar{y}\|^2 \le L^2c^2\Big(\frac{\|y^{k+1}-\bar{y}\|}{\|y^k-\bar{y}\|}+1\Big)\|y^k-\bar{y}\|^2 + \|y^k-\bar{y}\|^2.$$

The Q-linear convergence of $\{y^k\}$ implies $\limsup_k \|y^{k+1}-\bar{y}\|/\|y^k-\bar{y}\| = r$ for some $0 < r < 1$. Therefore, x^k converges to $\bar{x}$ R-linearly at the same rate. □

In Theorem 3, the local R-linear convergence of $\{x^k\}$ is established given that the multiplier sequence $\{y^k\}$ converges Q-linearly. As shown in the proof of Proposition 12, the sequence $\{y^k\}$ can be interpreted as being generated by PPA in the Euclidean setting. Therefore, its convergence properties can be analyzed directly using [53, Theorem 3.2]. If the condition (6.3c) is satisfied in each iteration, then $\{y^k\}$ converges to $\bar{y}$ Q-linearly.

Remark 7 As shown in [6], the primal dual convergence rate of RALM can be established under perturbation properties of the KKT solution mappings, which require unique multipliers. In contrast, our analysis under manifold variational sufficiency yields only local primal R-linear convergence rate but removes the uniqueness requirement, extending applicability to broader settings.

Remark 8 In Section 5, we establish that the RSQP method achieves local linear convergence under robust isolated calmness and hemistability. A similar convergence behavior can also be shown for RALM through the perturbation properties of the KKT solution mapping, using the error bound condition in Proposition 5. As shown in [6], the local Q-linear convergence of RALM can be established under the robust isolated calmness of the KKT solution mapping. This calmness can be verified in practice via the M-SOSC and M-SRCQ conditions, by Proposition 3. Again by Proposition 3, M-SSOSC combined with constraint nondegeneracy is equivalent to the strong regularity of the KKT solution mapping when $\mathrm{epi}(\theta)$ is a polyhedral convex set, a second-order cone, or a semidefinite cone, which also guarantees linear convergence of RALM. Actually, the local linear convergence of ALM under the SSOSC and constraint nondegeneracy is classical in the Euclidean setting, as shown in [59] for NLSDP and in [40] for NLSOC.

Now we discuss the convergence rate of the KKT residual. It follows from [53, Theorem 3.1] that the conditions (6.4) can be achieved by

$$\sqrt{\tilde{\rho}_k}\|\nabla_\xi l^{\rho_k}_{R_{\bar{x}}}(\xi^{k+1}, y^k)\| \leq \begin{cases} \text{(a)} \quad \varepsilon'_k, \\ \text{(b)} \quad \varepsilon'_k \min\{1, \|\tilde{\rho}_k \nabla_y l^{\rho_k}_{R_{\bar{x}}}(\xi^{k+1}, y^k)\|\}, \\ \text{(c)} \quad \varepsilon'_k \min\left\{1, \|\tilde{\rho}_k \nabla_y l^{\rho_k}_{R_{\bar{x}}}(\xi^{k+1}, y^k)\|^2\right\}, \end{cases} \tag{6.6}$$

where $\varepsilon'_k = \varepsilon_k\sqrt{s}$ as s is the modulus of the strong convexity of $l^{\rho_k}_{R_{\bar{x}}}(\cdot, y^k)$. Moreover, combining (6.6) with (6.5), the conditions (6.3) can be achieved by

$$\sqrt{\tilde{\rho}_k}\|\operatorname{grad}_x l^{\rho_k}(x^{k+1}, y^k)\| \leq \begin{cases} \text{(a)} \quad \varepsilon''_k, \\ \text{(b)} \quad \varepsilon''_k \min\{1, \|\tilde{\rho}_k \nabla_y l^{\rho_k}(x^{k+1}, y^k)\|\}, \\ \text{(c)} \quad \varepsilon''_k \min\left\{1, \|\tilde{\rho}_k \nabla_y l^{\rho_k}(x^{k+1}, y^k)\|^2\right\}. \end{cases} \tag{6.7}$$

where $\varepsilon''_k = \varepsilon_k\sqrt{s}/L$ as L is the modulus given in (6.5). The following proposition establishes the R-linear convergence of the KKT residual $\{E(x^{k+1}, y^{k+1})\}$ defined in (3.14). Its proof follows the argument of [62, Proposition 4.7], with the projection step replaced by a proximal mapping in our setting. For brevity, we omit the details.

Proposition 13 *Assume that the manifold strong variational sufficient condition for problem* (1.1) *under the retraction $R_{\bar{x}}$ holds with respect to $(\bar{x}, \bar{y})$ at level $\rho > 0$. Let $\{(x^k, y^k)\}$ be the sequence generated by Algorithm 2 under (6.7b). If $\sqrt{\tilde{\rho}_k}\varepsilon_k < 1$ for sufficiently large k, then there exists $L_g > 0$ such that*

$$E(x^{k+1}, y^{k+1}) \leq w^k \operatorname{dist}(Y^k, Z), \ \textit{where } w^k = (\varepsilon''_k\sqrt{\tilde{\rho}_k} + (1 + \rho + L_g\rho)(\tilde{\rho}_k)^{-1})(1 - \sqrt{\tilde{\rho}_k}\varepsilon_k)^{-1}.$$

6.2 Semismooth Newton method for RALM subproblem

For solving the RALM subproblem (6.1) efficiently, we adopt the globalized semismooth Newton method on Riemannian manifolds [64]. We begin by recalling the definition of the generalized covariant derivative for vector fields on manifolds from [44].

Definition 11 Let X be a locally Lipschitz vector field on $\mathcal{M}$. The B-derivative of X is a set-valued map $\partial_B X : \mathcal{M} \rightrightarrows \mathcal{L}(T\mathcal{M})$:

$$\partial_B X(x) = \{H \in \mathcal{L}(T_x\mathcal{M}) \mid \exists\{x^k\} \subseteq \mathcal{D}_X, \ x^k \to x, \ \nabla X(x^k) \to H\}, \ x \in \mathcal{M},$$

where the convergence $\nabla X(x^k) \to H$ means that $\|\nabla X(x^k)[P_{xx^k}v] - P_{xx^k}Hv\| \to 0$ for all $v \in T_x\mathcal{M}$. The Clarke generalized covariant derivative of X is a set-valued map $\partial X : \mathcal{M} \rightrightarrows \mathcal{L}(T\mathcal{M})$ such that $\partial X(x)$ is the convex hull of $\partial_B X(x)$.

If a retraction R_x additionally satisfies $\frac{\mathrm{D}^2}{\mathrm{d}t^2}R(t\xi)|_{t=0} = 0$ for any $\xi \in T_x\mathcal{M}$, where $\frac{\mathrm{D}^2}{\mathrm{d}t^2}\gamma$ denotes the acceleration of a curve γ, then R_x is called a second-order retraction. Such retractions can be interpreted as approximations of exponential mappings and include, for example, the polar retraction on the Stiefel manifold and the projective retraction on fixed rank manifolds; see [3, Example 4.12] for more details. In this section, we assume the retraction is second-order.

The globalized semismooth Newton method for solving the RALM subproblem (6.1), given in [64, Algorithm 4.1], aims to solve the equation $\operatorname{grad}_x l^{\rho_k}(x, y^k) = 0$ for x in the k-th iteration. Let $R_{\bar{x}}$ be a second-order retraction. Suppose that θ is either a polyhedral convex function or is the indicator function of a second-order cone or a positive semidefinite cone. Then, by Theorem 1, the M-SSOSC is equivalent to the positive definiteness of the elements in the Hessian bundle of $l^{\rho}_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y})$ for sufficiently large ρ. Moreover, we shall show that this condition also suffices to guarantee the superlinear convergence of the semismooth Newton method.

Lemma 2 *Let $(\bar{x}, \bar{y}) \in \mathcal{M} \times \mathbb{Y}$ be a KKT pair of problem* (1.1) *satisfying* (3.1)*. Assume that the manifold $\mathcal{M}$ is an embedded submanifold, and the retraction $R_{\bar{x}}$ is second-order. For any $\rho > 0$, we have $\partial \nabla_\xi l^{\rho}_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y}) = \partial \operatorname{grad}_x l^{\rho}(\bar{x}, \bar{y})$.*

Proof For any $G \in \partial \operatorname{grad}_x l^{\rho}(\bar{x}, \bar{y})$, we claim that $G \in \partial \nabla_\xi l^{\rho}_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y})$, or equivalently,

$$\lim_{\xi \to 0_{\bar{x}}} \|\nabla^2_{\xi\xi} l^{\rho}_{R_{\bar{x}}}(\xi, \bar{y})v - Gv\| = 0, \quad \forall\, v \in T_{\bar{x}}\mathcal{M}.$$

Denote $l^{\rho}(\cdot, \bar{y})$ by $h(\cdot)$. Since $h_{R_{\bar{x}}}(\xi) = h \circ R_{\bar{x}}(\xi)$, for any $v \in T_{\bar{x}}\mathcal{M}$ we have

$$\begin{aligned}
&\langle \nabla^2 h_{R_{\bar{x}}}(\xi)v, v\rangle = \frac{d^2}{dt^2}(h \circ R_{\bar{x}})(\xi + tv)\Big|_{t=0} \\
=&\frac{d}{dt}\Big[\frac{d}{dt}h(R_{\bar{x}}(\xi + tv))\Big]\Big|_{t=0} = \frac{d}{dt}Dh(R_{\bar{x}}(\xi + tv))\Big[\frac{d}{dt}R_{\bar{x}}(\xi + tv)\Big]\Big|_{t=0} \\
=&\langle \frac{D}{dt}\operatorname{grad} h(R_{\bar{x}}(\xi + tv)), DR_{\bar{x}}(\xi)v\rangle + \langle \operatorname{grad} h(R_{\bar{x}}(\xi + tv)), \frac{D^2}{dt^2}R_{\bar{x}}(\xi + tv)\rangle\Big|_{t=0} \\
=&\underbrace{\langle \nabla \operatorname{grad} h(R_{\bar{x}}(\xi))[DR_{\bar{x}}(\xi)v], DR_{\bar{x}}(\xi)v\rangle}_{(*)} + \underbrace{\langle \operatorname{grad} h(R_{\bar{x}}(\xi + tv)), \frac{D^2}{dt^2}R_{\bar{x}}(\xi + tv)\rangle\Big|_{t=0}}_{(**)}.
\end{aligned}$$

By the definition of retraction and $\mathcal{M}$ is an embedded submanifold, for any $v \in T_{\bar{x}}\mathcal{M}$, we have $DR_{\bar{x}}(\xi)v = D\exp_{\bar{x}}(\xi)v + O(\|\xi\|)$ when $\xi \to 0_{\bar{x}}$. Therefore, the first term of the last equation is

$$\begin{aligned}
(*) =&\langle \nabla \operatorname{grad} h(R_{\bar{x}}(\xi))[D\exp_{\bar{x}}(\xi)v], D\exp_{\bar{x}}(\xi)v\rangle + O(\|\xi\|) \\
=&\langle \nabla \operatorname{grad} h(R_{\bar{x}}(\xi))[P_{\bar{x}R_{\bar{x}}(\xi)}v], P_{\bar{x}R_{\bar{x}}(\xi)}v\rangle + O(\|\xi\|).
\end{aligned}$$

And the second term is zero $(**) = 0$ if $\xi \to 0_{\bar{x}}$ and $t \to 0$ as $R_{\bar{x}}$ is a second-order retraction. Thus, note that the parallel transport is isometry, for any $v \in T_{\bar{x}}\mathcal{M}$,

$$\begin{aligned}
\big\langle \nabla^2 h_{R_{\bar{x}}}\big(\xi\big)v - Gv, v\big\rangle = &\big\langle P_{\bar{x}R_{\bar{x}}(\xi)}\nabla^2 h_{R_{\bar{x}}}\big(\xi\big)v - \operatorname{Hess} h\big(R_{\bar{x}}(\xi)\big)\big[P_{\bar{x}R_{\bar{x}}(\xi)}v\big], P_{\bar{x}R_{\bar{x}}(\xi)}v\big\rangle \\
&+ \big\langle \operatorname{Hess} h\big(R_{\bar{x}}(\xi)\big)\big[P_{\bar{x}R_{\bar{x}}(\xi)}v\big] - P_{\bar{x}R_{\bar{x}}(\xi)}Gv, P_{\bar{x}R_{\bar{x}}(\xi)}v\big\rangle,
\end{aligned}$$

which converges to 0 when $\xi \to 0_{\bar{x}}$. The arbitrary taken v implies that $\lim_{\xi \to 0_{\bar{x}}} \big\|\nabla^2 h_{R_{\bar{x}}}\big(\xi\big)v - Gv\big\| = 0$. Similarly, $\partial \nabla_\xi l^{\rho}_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y}) \subseteq \partial \operatorname{grad}_x l^{\rho}(\bar{x}, \bar{y})$. The equality is proved and the proof is completed. □

Now, combining Lemma 2 with [64, Theorem 4.3], we obtain the following local convergence result for the globalized semismooth Newton method.

Theorem 4 *Assume that the manifold strong variational sufficient condition for problem* (1.1) *under the retraction* $R_{\bar{x}}$ *holds with respect to* $(\bar{x}, \bar{y})$ *at level* $\rho > 0$. *Let* $R_{\bar{x}}$ *be a second-order retraction and* $\{x^j\}$ *be the sequence generated by the semismooth Newton method [64, Algorithm 4.1]. Suppose there exists* $\delta > 0$ *such that the level set* $\Omega := \{x \in \mathcal{M} \mid l^{\rho_k}(x, y^k) \leq l^{\rho_k}(x^0, y^k) + \delta\}$ *is compact. Denote* $\hat{x}$ *to be any accumulation point of* $\{x^j\}$. *If* $\operatorname{grad}_x l^{\rho_k}(\cdot, y^k)$ *is* ν*-th order semismooth at* $\hat{x}$ *with respect to* $\partial \operatorname{grad}_x l^{\rho_k}(\cdot, y^k)$, *then we have the convergence* $x^k \to \hat{x}$, *and* $\hat{x}$ *is optimal to subproblem (6.1). Moreover, for sufficiently large* j, *it holds*

$$d(x^{j+1}, \hat{x}) \leq O\big(d(x^j, \hat{x})^{1+\min\{\nu, \bar{\nu}\}}\big),$$

where $\bar{\nu} \in (0, 1]$ *is the parameter defined in [64, Algorithm 4.1].*

Proof By Theorem 1, the manifold strong sufficiency ensures the positive definiteness of all elements in $\partial \nabla_\xi l^{\rho_j}_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y})$. Then by Lemma 2, all elements in $\partial \operatorname{grad}_x l^{\rho_j}(\bar{x}, \bar{y})$ are also positive definite. Thus the conditions in [64, Theorem 4.3] are satisfied, implying the local superlinear convergence of the globalized semismooth Newton method. The proof is completed. □

Remark 9 Based on Theorem 1 and Lemma 2, the positive definiteness of the generalized Hessian of the augmented Lagrangian function is equivalent to the manifold strong variational sufficient condition of problem (1.1) at the KKT point. This equivalence underscores the pivotal role of manifold strong variational sufficiency in ensuring the efficiency of the semismooth Newton method for solving the RALM subproblem.

7 Numerical experiments

In this section, we present numerical experiments on the fixed-rank manifold under the M-SOSC and M-SRCQ conditions, and on the Stiefel manifold under the M-SSOSC condition. These experiments illustrate the local convergence rates established in Theorems 2 and 3. The codes were implemented in Matlab (R2021b), and all experiments were performed on a 64-bit MacOS system with an Intel Cores i5 2.4GHz CPU and 16GB of memory.

7.1 Solving robust matrix completion by RSQP

We apply RSQP (Algorithm 1) to the robust matrix completion (RMC) problem proposed in [13]. Given $A \in \mathbb{R}^{m\times n}$, let $g(X) = P_\Omega(X - A)$ and $\theta(\cdot) = \mu\|\cdot\|_1$, where Ω is an index set and P_Ω is the projection operator defined by $(P_\Omega(X))_{ij} = X_{ij}$ if $(i, j) \in \Omega$ and 0 otherwise. We let $\mathcal{M} = Fr(m, n, r) := \{X \in \mathbb{R}^{m\times n} \mid \operatorname{rank}(X) = r\}$, and the RMC problem can be formulated as

$$\min \ \|P_\Omega(X - A)\|_1 \quad \text{s.t.} \quad X \in Fr(m, n, r). \tag{7.1}$$

As shown in [60], the tangent and normal spaces of $\mathcal{M}$ at a point X with the singular value decomposition $X = USV^\top$ are, respectively, given by

$$T_X\mathcal{M} = \left\{ \begin{bmatrix} U \ U_\perp \end{bmatrix} \begin{bmatrix} \mathbb{R}^{r\times r} & \mathbb{R}^{r\times(n-r)} \\ \mathbb{R}^{(m-r)\times r} & 0^{(m-r)\times(n-r)} \end{bmatrix} \begin{bmatrix} V \ V_\perp \end{bmatrix}^\top \right\},$$

$$N_X\mathcal{M} = \left\{ \begin{bmatrix} U \ U_\perp \end{bmatrix} \begin{bmatrix} 0^{r\times r} & 0^{r\times(n-r)} \\ 0^{(m-r)\times r} & \mathbb{R}^{(m-r)\times(n-r)} \end{bmatrix} \begin{bmatrix} V \ V_\perp \end{bmatrix}^\top \right\},$$

and the projection of a matrix $Y \in \mathbb{R}^{m\times n}$ onto the tangent space $T_X\mathcal{M}$ is given by $\Pi_X(Y) = P_U Y P_V + P_U^{\perp} Y P_V + P_U Y P_V^{\perp}$. The Lagrangian of (7.1) is given by $L(X, y) = \langle P_\Omega(X - A), y\rangle$. And its KKT conditions are

$$\Pi_X P_\Omega(y) = 0, \quad y \in \partial\|P_\Omega(X - A)\|_1. \tag{7.2}$$

By [32, Section 3], the Hessian of a function f on $\mathcal{M}$ at $X = U\Sigma V^\top$ is given by

$$\operatorname{Hess} f(X)\xi = U\hat{M}V^\top + \hat{U}_p V^\top + U\hat{V}_p^\top, \quad \xi \in T_X\mathcal{M},$$

where

$$\begin{aligned}
&\hat{M} = M\left(\nabla^2 f(X)[\xi]; X\right), M(Z; X) = U^\top Z X,\\
&\hat{U}_p = U_p\left(\nabla^2 f(X)[\xi]; X\right) + P_U^{\perp}\nabla f(X) V_p(\xi; X)/\Sigma, \quad U_p(Z; X) = P_U^{\perp} Z V,\\
&\hat{V}_p = V_p\left(\nabla^2 f(X)[\xi]; X\right) + P_V^{\perp}\nabla f(X) U_p(\xi; X)/\Sigma, \quad V_p(Z; X) = P_V^{\perp} Z^\top U.
\end{aligned}$$

Let $(\bar{X}, \bar{y})$ be a KKT pair satisfying (7.2). Since $\nabla_X L(\bar{X}, \bar{y}) = P_\Omega(\bar{y})$ and $\nabla^2_{XX} L(\bar{X}, \bar{y}) = 0$, it holds that

$$\begin{aligned}
\langle \xi, \operatorname{Hess}_X L(\bar{X}, \bar{y})\xi\rangle &= \langle \xi, P_U^{\perp} P_\Omega(\bar{y}) P_V^{\perp} \xi^\top U\Sigma^{-1}V^\top + U\Sigma^{-1}V^\top\xi^\top P_U^{\perp} P_\Omega(\bar{y}) P_V^{\perp}\rangle\\
&= 2\operatorname{tr}(\xi^\top P_U^{\perp} P_\Omega(\bar{y}) P_V^{\perp} V^{\perp}\xi^\top U\Sigma^{-1}V^\top) = 2\operatorname{tr}(\xi^\top P_\Omega(\bar{y})\xi^\top U\Sigma^{-1}V^\top),
\end{aligned}$$

where the last equality is obtained by the KKT condition (7.2) that $\Pi_{\bar{X}} P_\Omega(\bar{y}) = 0$. Moreover, by the definition of G, $Dg(X)\xi = g'(X)\xi = P_\Omega(\xi)$, for any $\xi \in T_X\mathcal{M}$. We can further obtain that $\mathcal{C}_{\theta,g}(X, y) = \left\{d \in \mathbb{R}^{m\times n} \mid \theta^{\downarrow}(P_\Omega(X - A); d) = \langle d, y\rangle\right\}$, in which

$$\theta^{\downarrow}(P_\Omega(X - A); d) = \sum_{P_\Omega(X-A)_{ij}=0} |d_{ij}| + \sum_{P_\Omega(X-A)_{ij}>0} d_{ij} - \sum_{P_\Omega(X-A)_{ij}<0} d_{ij}.$$

Therefore, the M-SOSC for problem (7.1) holds at the KKT pair $(\bar{X}, \bar{y})$ if

$$\operatorname{tr}\left(\xi^\top P_\Omega(\bar{y})\xi^\top U\Sigma^{-1}V^\top\right) > 0, \text{ for any } \xi \in T_{\bar{X}}\mathcal{M} \text{ satisfying } P_\Omega(\xi) \in \mathcal{C}_{\theta,g}(\bar{X}, \bar{y})\backslash\{0\}.$$

Next we consider a toy example such that the M-SOSC and M-SRCQ conditions can be verified. We take $m = n = 5$, $r = 3$, and Ω as the full index set in problem (7.1). Let

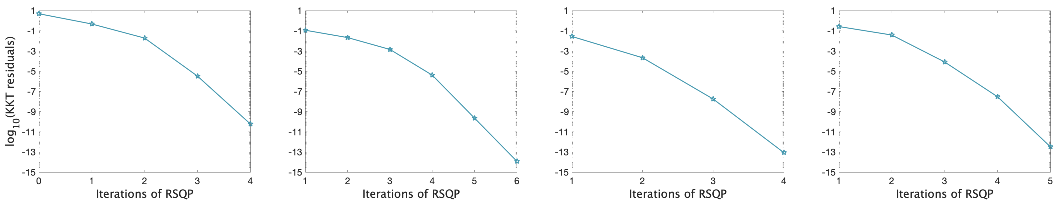
$$U = \begin{bmatrix} 1 & 0 & 0 & 0 & 0\\ 0 & -\sqrt{2}/2 & \sqrt{2}/2 & 0 & 0\\ 0 & \sqrt{2}/2 & \sqrt{2}/2 & 0 & 0\end{bmatrix}^\top, \quad V = \begin{bmatrix} 1 & 0 & 0 & 0 & 0\\ 0 & 0.6 & -0.8 & 0 & 0\\ 0 & 0.8 & 0.6 & 0 & 0\end{bmatrix}^\top, \quad \text{and} \quad S = \begin{bmatrix} 1 & & \\ & 2 & \\ & & 3\end{bmatrix}.$$

Set $A_{\mathrm{ex}} = USV^\top$. We then add noise by defining E_{out} as a matrix with Gaussian entries at positions $(i, j) \in \{4, 5\} \times \{4, 5\}$ and 0 elsewhere, and set $A = A_{\mathrm{ex}} + E_{\mathrm{out}}$.

We take $\bar{X} = A_{\mathrm{ex}}$ and $\bar{y}_{ij} = -\operatorname{sgn}((E_{\mathrm{out}})_{ij})$. It can be verified that $(\bar{X}, \bar{y})$ satisfies the KKT conditions (7.2). Next we show that the M-SOSC and M-SRCQ conditions hold at $(\bar{X}, \bar{y})$. First, we can compute the critical cone $\mathcal{C}_{\theta,g}(\bar{X}, \bar{y}) = \{d \in \mathbb{R}^{m\times n} \mid d_{ij} \in \mathbb{R} \text{ if } (E_{\mathrm{out}})_{ij} \neq 0,\ d_{ij} = 0 \text{ if } (E_{\mathrm{out}})_{ij} = 0\}$. The sparsity pattern of E_{out} implies that only the zero element $\xi = 0_{\bar{X}}$ satisfies $P_\Omega(\xi) \in \mathcal{C}_{\theta,g}(\bar{X}, y)$. Therefore, the M-SOSC holds at $(\bar{X}, \bar{y})$. Moreover, observe that $Dg(\bar{X})T_{\bar{X}}\mathcal{M} = T_{\bar{X}}\mathcal{M}$. Thus, the M-SRCQ condition holds if there exists a multiplier $\bar{y}$ such that $N_{\bar{X}}\mathcal{M} \subseteq \mathcal{C}_{\theta,g}(\bar{X}, \bar{y})$. For any multiplier y satisfying $Dg(\bar{X})^* y = 0$, we have $y \in N_{\bar{X}}\mathcal{M}$ and thus $y = P_U^\perp Y P_V^\perp$ for some $Y \in \mathbb{R}^{m\times n}$. For any $d = P_U^\perp \widetilde{Y} P_V^\perp \in N_{\overline{X}}\mathcal{M}$, it is easy to see that there exists a multiplier $y = P_U^\perp Y P_V^\perp$ satisfying

$$\sum_{i,j\in\{4,5\}} -\operatorname{sgn}((E_{\text{out}})_{ij})\widetilde{Y}_{ij} = \sum_{i,j\in\{4,5\}} -\operatorname{sgn}((E_{\text{out}})_{ij})d_{ij} = \theta^{\downarrow}(g(\overline{X}); d)$$
$$= \langle d, y\rangle = \langle P_U^\perp \widetilde{Y} P_V^\perp, P_U^\perp Y P_V^\perp\rangle = \sum_{i,j\in\{4,5\}} \widetilde{Y}_{ij} Y_{ij}.$$

Therefore, $d \in \mathcal{C}_{\theta,g}(\overline{X}, \bar{y})$, and the M-SRCQ condition holds at $\overline{X} = A_{\text{ex}}$ with respect to y.



(a) $m = n = 5$, $r = 3$ (b) $m = n = 50$, $r = 6$ (c) $m = n = 100$, $r = 8$ (d) $m = n = 200$, $r = 10$

Fig. 1: Superlinear decay of the KKT residuals (see (3.14)) across RSQP iterations (Algorithm 1) for the RMC problem (7.1): (a) toy instance; (b-d) random instances of varying dimensions.

Table 1: KKT residuals (see (3.14)) and prediction errors $\|X^k - A\|$ across RSQP iterations (Algorithm 1) for random instances of the RMC problem (7.1).

Problem size	Iteration k	KKT residuals	Errors $\|X^k - A\|$
$m = n = 50$, $r = 6$	1	1.14×10^{-1}	1.39×10^{-1}
	2	2.14×10^{-2}	1.49×10^{-2}
	3	1.42×10^{-3}	6.57×10^{-4}
	4	4.04×10^{-6}	1.78×10^{-6}
	5	2.23×10^{-10}	5.91×10^{-10}
	6	1.19×10^{-14}	6.18×10^{-14}
$m = n = 100$, $r = 8$	1	2.76×10^{-2}	7.79×10^{-2}
	2	2.15×10^{-4}	2.54×10^{-4}
	3	1.75×10^{-8}	1.24×10^{-8}
	4	8.62×10^{-14}	5.52×10^{-14}
$m = n = 200$, $r = 10$	1	2.65×10^{-1}	1.69×10^{-1}
	2	3.84×10^{-2}	5.01×10^{-2}
	3	8.23×10^{-5}	3.95×10^{-5}
	4	2.97×10^{-8}	1.06×10^{-8}
	5	3.26×10^{-13}	1.69×10^{-12}

We apply RSQP (Algorithm 1) for solving this toy instance of the RMC problem (7.1). Theoretically, by Proposition 11, the M-SOSC and M-SRCQ conditions ensure the superlinear convergence rate of RSQP (Algorithm 1). This is confirmed numerically in Figure 1(a), which displays the superlinear decay of the KKT residuals defined in (3.14).

We further test larger problem sizes with dimensions $(m, n, r) = (50, 50, 6)$, $(100, 100, 8)$, and $(200, 200, 10)$. The random matrix $A \in \mathbb{R}^{m\times n}$ is generated as $A = LR^\top$, where $L \in \mathbb{R}^{m\times r}$ and $R \in \mathbb{R}^{n\times r}$ have entries independently sampled from the uniform distribution on $[0, 1]$. The observation set Ω is formed by including each index independently with probability $p = 0.3$. Algorithm 1 is implemented based on a slightly modified version of the method in [43], where each subproblem is solved under an orthonormal basis of the tangent space. Since Algorithm 1 does not include a globalization strategy, we use a single iteration of RALM to produce the initial point before applying RSQP. We report the averaged performance over 10 independent replications of each experiment.

Figure 1(b-d) demonstrates the superlinear decay of the KKT residuals defined in (3.14). Correspondingly, Table 1 reports both the KKT residuals and the prediction errors of the iterates from the true matrix $\|X^k - A\|$. Although it is difficult to verify the M-SOSC and M-SRCQ conditions in these instances, the results in Figure 1 and Table 1 provide clear empirical evidence of superlinear convergence. Moreover, Table 1 shows that the iterates X^k recover the truth A almost exactly, with errors between 10^{-14} and 10^{-12}. This high accuracy is likely due to the fact that no noise is added to the low-rank true matrix A in these experiments.

7.2 Solving compressed modes by RALM

We apply RALM (Algorithm 2) to the compressed modes (CM) problem proposed in [45]. Let $f(X) = \mathrm{tr}(X^\top HX)$, $\theta(\cdot) = \mu\|\cdot\|_1$, $g(X) = X$, and $\mathcal{M} = \mathrm{St}(n, r) := \{X \in \mathbb{R}^{n\times r} \mid X^\top X = I_r\}$, Let H be a discretization of the Hamilton operator. Then the CM problem can be formulated as

$$\min\ \mathrm{tr}(X^\top HX) + \mu \|X\|_1 \quad \text{s.t.} \quad X \in \mathrm{St}(n, r). \tag{7.3}$$

The tangent and normal spaces of $\mathcal{M}$ at a point X are, respectively, given by

$$\begin{aligned} T_X\mathcal{M} &= \left\{\xi \in \mathbb{R}^{n\times r} \mid X^\top\xi + \xi^\top X = 0\right\}, \\ N_X\mathcal{M} &= \left\{XA \mid A = A^\top\right\}, \end{aligned} \tag{7.4}$$

and the the projection of a matrix $Y \in \mathbb{R}^{n\times r}$ onto the tangent space $T_X\mathcal{M}$ is given by $\Pi_X(Y) = Y - X\,\mathrm{sym}(X^\top Y)$, where $\mathrm{sym}(Z) := (Z + Z^\top)/2$ for any Z. The Lagrangian of (7.3) is given by $L(X, y) = \mathrm{tr}(X^\top HX) + \langle X, y\rangle$. And its KKT conditions are

$$\Pi_X(2HX + y) = 0, \quad y \in \mu\partial\|X\|_1. \tag{7.5}$$

The Euclidean gradient of $L(X, y)$ is $\nabla_X L(X, y) = 2HX + y$. And the Euclidean Hessian of $L(X, y)$ is $\nabla^2_{XX} L(X, y)\xi = 2H\xi$ for any $\xi \in T_X\mathcal{M}$. Therefore, by [32], the Riemannian gradient and Hessian of $L(X, y)$ can be computed as

$$\mathrm{grad}_X\, L(X, y) = \Pi_X(2HX + y), \quad \mathrm{Hess}_X\, L(X, y)\xi = \Pi_X(2H\xi - \xi\,\mathrm{sym}(X^\top\nabla_X L(X, y)).$$

By the KKT condition (7.5), we have $2HX + y \in N_X\mathcal{M}$, which implies that $2HX + y = XS$ for some symmetric matrix S. Therefore, for any $\xi \in T_X\mathcal{M}$,

$$\begin{aligned} \langle\xi, \mathrm{Hess}_X\, L(X, y)\xi\rangle &= \langle\xi, 2H\xi - \xi\,\mathrm{sym}(X^\top\nabla_X L(X, y))\rangle \\ &= \langle\xi, 2H\xi\rangle - \langle\xi, 2\xi X^\top HX\rangle - \langle\xi, \xi\,\mathrm{sym}(X^\top y)\rangle \\ &= \langle\xi, 2H\xi\rangle - \langle\xi, 2\xi X^\top HX\rangle - \langle\xi, \xi\,\mathrm{sym}(X^\top(XS - 2HX))\rangle \\ &= \mathrm{tr}(\xi^\top H\xi) - \mathrm{tr}(\xi^\top\xi S). \end{aligned}$$

We further obtain that $\mathcal{C}_{\theta,g}(X,y) = \{d \in \mathbb{R}^{n\times r} \mid \theta^{\downarrow}(X;d) = \langle d, y\rangle\}$, in which

$$\theta^{\downarrow}(X;d) = \sum_{X_{ij}=0} |d_{ij}| + \sum_{X_{ij}>0} d_{ij} - \sum_{X_{ij}<0} d_{ij}.$$

Its affine hull is

$$\operatorname{aff}\mathcal{C}_{\theta,g}(X,y) = \{d \in \mathbb{R}^{n\times r} \mid d_{ij} = 0 \text{ if } y_{ij} \neq \pm\mu,\ d_{ij} \in \mathbb{R} \text{ if } y_{ij} = \pm\mu\}. \tag{7.6}$$

Therefore, the M-SSOSC for problem (7.3) holds at the KKT pair $(\bar{X},\bar{y}) = (\bar{X}, -2H\bar{X} + \bar{X}\bar{S})$ if

$$\operatorname{tr}(\xi^\top H\xi) - \operatorname{tr}(\xi^\top \xi\bar{S}) > 0, \text{ for any } \xi \in T_{\bar{X}}\mathcal{M} \text{ satisfying } \xi \in \operatorname{aff}\mathcal{C}_{\theta,g}(\bar{X},\bar{y})\backslash\{0\}.$$

In [64], the authors formulate the CM problem in the context of solving the Schrödinger equation for a one-dimensional free-electron model with periodic boundary conditions:

$$-\tfrac{1}{2}\Delta\phi(x) = \lambda\phi(x), \quad x \in [0,50].$$

Their numerical results indicate that the smallest eigenvalue of $H_k \in \partial\,\mathrm{grad}_X\, L^{\rho_k}(X^k,y^k)$ remains strictly positive, suggesting that the M-SSOSC condition may hold in this setting. To further illustrate this conjecture, we present a simple example below. Consider the Schrödinger equation on the domain $[0,2]$ with periodic boundary conditions. We discretize $[0,2]$ into $n=4$ grid points and denote H by the discrete approximation of $-\frac{1}{2}\Delta$. In this case, H takes the form $H = -\begin{bmatrix} -4 & 2 & 0 & 2 \\ 2 & -4 & 2 & 0 \\ 0 & 2 & -4 & 2 \\ 2 & 0 & 2 & -4 \end{bmatrix}$. We take $\mu < 5\sqrt{2}$, $r=2$, $\bar{X} = \begin{bmatrix} 0 & 0 & \sqrt{2}/2 & \sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 & 0 & 0 \end{bmatrix}^\top$, and $\bar{y} = \mu\begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}^\top$. It can be verified that $(\bar{X},\bar{y})$ satisfies the KKT conditions (7.5). And for $\bar{S} = \begin{bmatrix} -4+\sqrt{2}\mu & 4 \\ 4 & -4+\sqrt{2}\mu \end{bmatrix}$, it holds that $2H\bar{X} + \bar{y} = \bar{X}\bar{S}$. Next we show that the M-SSOSC condition holds at $(\bar{X},\bar{y})$. By (7.6), if follows that

$$\operatorname{aff}\mathcal{C}_{\theta,g}(\bar{X},\bar{y}) = \left\{ \begin{bmatrix} 0 & 0 & c_1 & c_2 \\ c_3 & c_4 & 0 & 0 \end{bmatrix}^\top \;\middle|\; c_i \in \mathbb{R}, i = 1,2,3,4 \right\}.$$

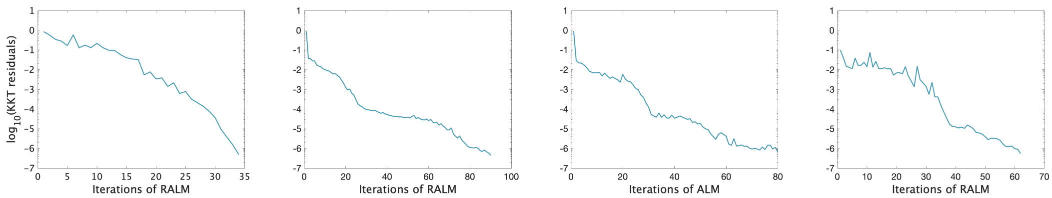
For any $\xi \in T_{\bar{X}}\mathcal{M}$ satisfying $\xi \in \operatorname{aff}\mathcal{C}_{\theta,g}(\bar{X},\bar{y})$, (7.4) implies that ξ takes the form

$$\xi = \begin{bmatrix} 0 & 0 & c_1 & -c_1 \\ c_2 & -c_2 & 0 & 0 \end{bmatrix}^\top, \quad c_1, c_2 \in \mathbb{R},$$

and thus for $\mu < 5\sqrt{2}$, $\operatorname{tr}(\xi^\top H\xi) - \operatorname{tr}(\xi^\top\xi\bar{S}) = 12(c_1^2+c_2^2) - 2(-4+\sqrt{2}\mu)(c_1^2+c_2^2) = (20 - 2\sqrt{2}\mu)(c_1^2+c_2^2) > 0$. Therefore the M-SSOSC holds at $(\bar{X},\bar{y})$ given $\mu < 5\sqrt{2}$.

We set $\mu = 0.8$ and apply RALM (Algorithm 2) for solving this toy instance of the CM problem (7.3). Figure 2(a) illustrates the linear convergence rate of the KKT residuals defined in (3.14), which confirms the theoretical results in Theorem 3.

We further evaluate larger problem sizes by discretizing the domain $[0,50]$ into n nodes with $n = 200$, 500, 1000, setting $r = 20$ and $\mu = 1$, and taking H as the discretized version of $-\frac{1}{2}\Delta$. Algorithm 2 is implemented following [64]. Figure 2(b-d) demonstrates the linear decay of the KKT residuals based on 10 independent replications. In parallel, Table 2 reports the smallest eigenvalues of all generalized Hessians arising in the algorithm, from 5 independent replications of each experiment. While direct verification of the M-SSOSC condition remains challenging in these instances and thus theoretical guarantees on eigenvalue positivity are absent, we still observe in Table 2 that the generalized Hessians arising in the semismooth Newton method consistently exhibit positive eigenvalues. This might suggest that the M-SSOSC condition may indeed hold in these instances.



(a) $n = 4, r = 2$ (b) $n = 200, r = 20$ (c) $n = 500, r = 20$ (d) $n = 1000, r = 20$

Fig. 2: Linear decay of the KKT residuals (see (3.14)) across RALM iterations (Algorithm 2) for the CM problem (7.3): (a) toy instance; (b-d) random instances of varying dimensions.

Table 2: Smallest eigenvalues of all generalized Hessians arising in the semismooth Newton based RALM (Algorithm 2) for random instances of the CM problem (7.3), based on 5 independent replications.

n	r	smallest eigenvalues				
200	20	0.30	0.32	0.19	0.28	0.14
500	20	0.28	0.11	0.12	0.09	0.11
1000	20	0.03	0.17	0.01	0.17	0.21

8 Conclusion

In this paper, we have developed a retraction-based perturbation framework for nonsmooth optimization on connected Riemannian manifolds. By reducing the analysis to locally equivalent tangent space models, we extend fundamental regularity concepts from Euclidean variational analysis to the manifold setting and provide explicit characterizations. In particular, we show that the robust isolated calmness of the KKT solution mapping is equivalent to the M-SOSC and M-SRCQ conditions. Moreover, for epi θ being a convex polyhedral set, a second-order cone, or a semidefinite cone, we establish that the strong regularity of the KKT solution mapping is equivalent to the M-SSOSC and manifold constraint nondegeneracy, which is also equivalent to the nonsingularity of every element in Clarke's generalized Jacobian of the natural mapping. We further introduce the notion of the manifold (strong) variational sufficiency, and clarify their role in perturbation analysis. In particular, we show for such θ the manifold strong variational sufficiency is equivalent to the M-SSOSC. These theoretical developments have facilitated the convergence analysis of nonsmooth optimization algorithms on manifolds. Specifically, under robust isolated calmness, the RSQP achieves local superlinear (quadratic) convergence. Under manifold strong variational sufficiency, the RALM attains R-linear convergence of its outer iterations. Moreover, for the associated inner subproblems, all generalized Hessians arising in the semismooth Newton method are positive definite, thereby guaranteeing superlinear local convergence of the inner iterations. Numerical experiments on the robust matrix completion and compressed modes problem confirm the predicted convergence rates.

Several challenges remain open. In the robust matrix completion problem, computing the Riemannian Hessian is computationally expensive in our current implementation. This motivates the development of Hessian-free or coordinate-free subproblem strategies that can more effectively exploit the underlying manifold structure. In addition, while the PPA for the primal problem and the ALM for the dual problem are known to be equivalent in the Euclidean setting, their precise relationship under convexity assumptions on Riemannian manifolds remains unclear and warrants further investigation.

References

1. P. A. Absil and S. Hosseini. A collection of nonsmooth Riemannian optimization problems. *Nonsmooth Optimization and Its Applications*, 170:1–15, 2019.
2. P. A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds.* Princeton University Press, 2008.
3. P. A. Absil and J. Malick. Projection-like retractions on matrix manifolds. *SIAM Journal on Optimization*, 22(1):135–158, 2012.
4. R. Andreani, K. R. Couto, O. P. Ferreira, and G. Haeser. Constraint qualifications and strong global convergence properties of an augmented Lagrangian method on Riemannian manifolds. *SIAM Journal on Optimization*, 34(2):1799–1825, 2024.
5. D. Azagra, J. Ferrera, and F. López-Mesas. Nonsmooth analysis and Hamilton–Jacobi equations on Riemannian manifolds. *Journal of Functional Analysis*, 220(2):304–361, 2005.
6. C. L. Bao, C. Ding, and Y. X. Zhou. On the robust isolated calmness of a class of nonsmooth optimizations on Riemannian manifolds and its applications. *Computational Optimization and Applications*, pages 1–46, 2025.
7. P. T. Boggs and J. W. Tolle. Sequential quadratic programming. *Acta Numerica*, 4:1–51, 1995.
8. J. F. Bonnans. Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Applied Mathematics and Optimization*, 29:161–186, 1994.
9. J. F. Bonnans and H. Ramírez C. Perturbation analysis of second-order cone programming problems. *Mathematical Programming*, 104(2):205–227, 2005.
10. J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems.* Springer Science & Business Media, 2013.
11. J. F. Bonnans and A. Sulem. Pseudopower expansion of solutions of generalized equations and constrained optimization problems. *Mathematical Programming*, 70:123–148, 1995.
12. N. Boumal. *An Introduction to Optimization on Smooth Manifolds.* Cambridge University Press, 2023.
13. L. Cambier and P. A. Absil. Robust low-rank matrix completion by Riemannian optimization. *SIAM Journal on Scientific Computing*, 38(5):S440–S460, 2016.
14. L. Chen, R. N. Chen, D. F. Sun, and Zhu J. Y. Aubin property and strong regularity are equivalent for nonlinear second-order cone programming. *SIAM Journal on Optimization*, 35(2):712–738, 2025.
15. L. Chen, R. N. Chen, D. F. Sun, and L. P. Zhang. Characterizations of the Aubin property of the solution mapping for nonlinear semidefinite programming. *Mathematical Programming*, pages 1–32, 2025.
16. S. X. Chen, Z. D. Deng, S. Q. Ma, and A. M. C. So. Manifold proximal point algorithms for dual principal component pursuit and orthogonal dictionary learning. *IEEE Transactions on Signal Processing*, 69:4759–4773, 2021.
17. S. X. Chen, S. Q. Ma, A. M. C. So, and T. Zhang. Proximal gradient method for nonsmooth optimization over the Stiefel manifold. *SIAM Journal on Optimization*, 30(1):210–239, 2020.
18. F. H. Clarke. *Optimization and Nonsmooth Analysis.* Society for Industrial and Applied Mathematics, 1990.
19. L. Demanet and P. Hand. Scaling law for recovering the sparsest element in a subspace. *Information and Inference: A Journal of the IMA*, 3(4):295–309, 2014.
20. K. K. Deng and Z. Peng. An inexact augmented Lagrangian method for nonsmooth optimization on Riemannian manifold. *arXiv preprint arXiv:1911.09900*, 2019.
21. C. Ding, D. F. Sun, and L. W. Zhang. Characterization of the robust isolated calmness for a class of conic programming problems. *SIAM Journal on Optimization*, 27(1):67–90, 2017.

22. A. L. Dontchev and R. T. Rockafellar. Characterizations of strong regularity for variational inequalities over polyhedral convex sets. *SIAM Journal on Optimization*, 6(4):1087–1105, 1996.
23. A. L. Dontchev and R. T. Rockafellar. *Implicit Functions and Solution Mappings*, volume 543. New York: Springer, 2009.
24. O. P. Ferreira and P. R. Oliveira. Subgradient algorithm on Riemannian manifolds. *Journal of Optimization Theory and Applications*, 97(1):93–104, 1998.
25. O. P. Ferreira and P. R. Oliveira. Proximal point algorithm on Riemannian manifolds. *Optimization*, 51(2):257–270, 2002.
26. P. Grohs and S. Hosseini. ε-subgradient algorithms for locally Lipschitz functions on Riemannian manifolds. *Advances in Computational Mathematics*, 42(2):333–360, 2016.
27. S. P. Han. A globally convergent method for nonlinear programming. *Journal of Optimization Theory and Applications*, 22(3):297–309, 1977.
28. M. R. Hestenes. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4(5):303–320, 1969.
29. S. Hosseini and M. R. Pouryayevali. Generalized gradients and characterization of epi-Lipschitz sets in Riemannian manifolds. *Nonlinear Analysis: Theory, Methods & Applications*, 74(12):3884–3895, 2011.
30. S. Hosseini and M. R. Pouryayevali. Nonsmooth optimization techniques on Riemannian manifolds. *Journal of Optimization Theory and Applications*, 158:328–342, 2013.
31. S. Hosseini and A. Uschmajew. A Riemannian gradient sampling algorithm for nonsmooth optimization on manifolds. *SIAM Journal on Optimization*, 27(1):173–189, 2017.
32. J. Hu, X. Liu, Z. Wen, and Y. X. Yuan. A brief introduction to manifold optimization. *Journal of the Operations Research Society of China*, 8:199–248, 2020.
33. W. Huang and K. Wei. An extension of fast iterative shrinkage-thresholding algorithm to Riemannian optimization for sparse principal component analysis. *Numerical Linear Algebra with Applications*, 29(1):e2409, 2022.
34. W. Huang and K. Wei. Riemannian proximal gradient methods. *Mathematical Programming*, 194(1-2):371–413, 2022.
35. C. Kanzow and D. Steck. Improved local convergence results for augmented Lagrangian methods in $\mathcal{C}^2$-cone reducible constrained optimization. *Mathematical Programming*, 177(1):425–438, 2019.
36. A. Kovnatsky, K. Glashoff, and M. M. Bronstein. MADMM: a generic algorithm for non-smooth optimization on manifolds. In *European Conference on Computer Vision*, pages 680–696. Springer, 2016.
37. R. J. Lai and S. Osher. A splitting method for orthogonality constrained problems. *Journal of Scientific Computing*, 58(2):431–449, 2014.
38. J. M. Lee. *Introduction to Smooth Manifolds*. New York, NY: Springer New York, 2003.
39. C. S. Liu and N. Boumal. Simple algorithms for optimization on Riemannian manifolds with constraints. *Applied Mathematics & Optimization*, 82(3):949–981, 2020.
40. Y. J. Liu and L. W. Zhang. Convergence of the augmented Lagrangian method for nonlinear optimization problems over second-order cones. *Journal of Optimization Theory and Applications*, 139(3):557–575, 2008.
41. A. Mohammadi, B. S. Mordukhovich, and M. E. Sarabi. Superlinear convergence of the sequential quadratic method in constrained optimization. *Journal of Optimization Theory and Applications*, 186(3):731–758, 2020.
42. A. Mohammadi, B. S. Mordukhovich, and M. E. Sarabi. Parabolic regularity in geometric variational analysis. *Transactions of the American Mathematical Society*, 374(3):1711–1763, 2021.
43. M. Obara, T. Okuno, and A. Takeda. Sequential quadratic optimization for nonlinear optimization problems on Riemannian manifolds. *SIAM Journal on Optimization*, 32(2):822–853, 2022.
44. F. R. Oliveira and O. P. Ferreira. Newton method for finding a singularity of a special class of locally Lipschitz continuous vector fields on Riemannian manifolds. *Journal of Optimization Theory and Applications*, 185(2):522–539, 2020.
45. V. Ozoliņš, R. J. Lai, R. Caflisch, and S. Osher. Compressed modes for variational problems in mathematics and physics. *Proceedings of the National Academy of Sciences*, 110(46):18368–18373, 2013.
46. Peter Petersen. *Riemannian Geometry*. New York, NY: Springer New York, 2006.
47. M. J. Powell. A method for nonlinear constraints in minimization problems. *Optimization*, pages 283–298, 1969.
48. M. J. Powell. A fast algorithm for nonlinearly constrained optimization calculations. In *Numerical Analysis: Proceedings of the Biennial Conference Held at Dundee*, pages 144–157. Springer, 2006.
49. R. T. Rockafellar. *Convex Analysis*, volume 18. Princeton University Press, 1970.
50. R. T. Rockafellar. A dual approach to solving nonlinear programming problems by unconstrained optimization. *Mathematical Programming*, 5(1):354–373, 1973.
51. R. T. Rockafellar. Advances in convergence and scope of the proximal point algorithm. *J. Nonlinear and Convex Analysis*, 22:2347–2375, 2021.
52. R. T. Rockafellar. Augmented Lagrangians and hidden convexity in sufficient conditions for local optimality. *Mathematical Programming*, 198(1):1–36, 2022.
53. R. T. Rockafellar. Convergence of augmented Lagrangian methods in extensions beyond nonlinear programming. *Mathematical Programming*, 199(1):1–46, 2022.

54. R. T. Rockafellar and R. J. Wets. *Variational Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998.
55. A. Schiela and J. Ortiz. An SQP method for equality constrained optimization on Hilbert manifolds. *SIAM Journal on Optimization*, 31(3):2255–2284, 2021.
56. W. T. Si, P. A. Absil, W. Huang, R. J. Jiang, and S. Vary. A Riemannian proximal Newton method. *SIAM Journal on Optimization*, 34(1):654–681, 2024.
57. D. A. Spielman, H. Wang, and J. Wright. Exact recovery of sparsely-used dictionaries. In *Conference on Learning Theory*, pages 37–1. JMLR Workshop and Conference Proceedings, 2012.
58. D. F. Sun. The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. *Mathematics of Operations Research*, 31(4):761–776, 2006.
59. D. F. Sun, J. Sun, and L. W. Zhang. The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming. *Mathematical Programming*, 114(2):349–391, 2008.
60. B. Vandereycken. Low-rank matrix completion by Riemannian optimization. *SIAM Journal on Optimization*, 23(2):1214–1236, 2013.
61. Q. S. Wang and W. H. Yang. Proximal quasi-Newton method for composite optimization over the Stiefel manifold. *Journal of Scientific Computing*, 95(2):39, 2023.
62. S. W Wang, C. Ding, Y. J. Zhang, and X. Y. Zhao. Strong variational sufficiency for nonlinear semidefinite programming and its implications. *SIAM Journal on Optimization*, 33(4):2988–3011, 2023.
63. Y. Yamakawa and H. Sato. Sequential optimality conditions for nonlinear optimization on Riemannian manifolds and a globally convergent augmented Lagrangian method. *Computational Optimization and Applications*, 81(2):397–421, 2022.
64. Y. H. Zhou, C. L. Bao, C. Ding, and J. Zhu. A semismooth Newton based augmented Lagrangian method for nonsmooth optimization on matrix manifolds. *Mathematical Programming*, 201(1-2):1–61, 2023.
65. H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15(2):265–286, 2006.