STRONG VARIATIONAL SUFFICIENCY OF NONSMOOTH OPTIMIZATION PROBLEMS ON RIEMANNIAN MANIFOLDS[*]

YUEXIN ZHOU[†], CHAO DING[‡], AND YANGJING ZHANG[§]

Abstract. The Riemannian Augmented Lagrangian Method (RALM), a recently proposed algorithm for nonsmooth optimization problems on Riemannian manifolds, has consistently exhibited high efficiency as evidenced in prior studies [42, 43]. It often demonstrates a rapid local linear convergence rate. However, a comprehensive local convergence analysis of the RALM under more realistic assumptions, notably without the imposition of the uniqueness assumption on the multiplier, remains an uncharted territory. In this paper, we introduce the manifold variational sufficient condition and demonstrate that its strong version is equivalent to the manifold strong second-order sufficient condition (M-SSOSC) in certain circumstances. Critically, we construct a local dual problem based on this condition and implement the Euclidean proximal point algorithm, which leads to the establishment of the linear convergence rate of the RALM. Moreover, we illustrate that under suitable assumptions, the M-SSOSC is equivalent to the nonsingularity of the generalized Hessian of the augmented Lagrangian function, which is an essential attribute for the semismooth Newton-type methods.

Key words. nonsmooth optimizations on Riemannian manifold, strong variational sufficiency, augmented Lagrangian method, rate of convergence

MSC codes. 90C30, 90C46, 49J52, 65K05

1. Introduction. This paper is concerned with the nonsmooth optimization problems on Riemannian manifolds in the following form:

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

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 [44], low-rank matrix completion problems [7], orthogonal dictionary learning problems [38, 12] and compressed modes problems [29]. For more detailed information on these applications, see [1, 20].

Different algorithms have been designed for solving the manifold nonsmooth optimization problem (1.1), such as subgradient methods [14, 16], proximal gradient methods [9, 21, 22], alternating direction methods of multipliers [24, 25] and proximal point methods (PPA) [8, 15]. In recent papers [13, 42], the augmented Lagrangian method (ALM) has also been extended to solve the nonsmooth

[*]This version: October 31, 2023

[†]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. (zhouyuexin19@mails.ucas.ac.cn).

[‡]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. (dingchao@amss.ac.cn). The work of this author was supported in part by National Key R&D Program of China 2021YFA1000300, 2021YFA1000301, National Natural Science Foundation of China under project No. 12071464 and CAS Project for Young Scientists in Basic Research No. YSBR-034.

[§]Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, P.R. China. (yangjing.zhang@amss.ac.cn).The work of this author was supported by the National Natural Science Foundation of China under project No. 12201617.

manifold optimization problem (1.1). By adding a perturbation parameter u, we can obtain the perturbed problem for (1.1):

$$\begin{array}{ll} \min & \varphi(x,u) := f(x) + \theta(g(x)+u) \\ \text{s.t.} & x \in \mathcal{M}. \end{array} \tag{1.2}$$

Problem (1.2) is reduced to problem (1.1) when $u = 0$. The Lagrangian function for (1.1) is

$$l(x,y) = \inf_u \{\varphi(x,u) - \langle y, u\rangle\} = L(x,y) - \theta^*(y),$$

where $L(x,y) = f(x) + \langle y, g(x)\rangle$ and θ^* is the conjugate function of θ. Moreover, for $\rho > 0$, the augmented Lagrangian function for (1.1) is defined by

$$l^\rho(x,y) = \inf_u \left\{\varphi(x,u) - \langle y, u\rangle + \frac{\rho}{2}\|u\|^2\right\}. \tag{1.3}$$

The inexact Riemannian augmented Lagrangian method (RALM) proposed in [13, 42] for solving (1.1) then takes the form of

$$\begin{cases} x^{k+1} \approx \bar{x}^{k+1} := \underset{x\in\mathcal{U}}{\operatorname{argmin}}\, l^{\rho_k}\left(x, y^k\right), \\ y^{k+1} = y^k + \tilde{\rho}_k \nabla_y l^{\rho_k}\left(x^{k+1}, y^k\right), \end{cases} \tag{1.4}$$

where $\mathcal{U}$ is a subset of $\mathcal{M}$, $\tilde{\rho}_k > 0$ and $\rho_k > 0$ are nondecreasing sequences satisfying $\tilde{\rho}_k \le \rho_k$. As evidenced in [42, 43], the RALM consistently displays high efficiency, often demonstrating a fast local linear convergence rate. Furthermore, when the semismooth Newton method is utilized to solve the augmented Lagrangian subproblems in (1.4), it has frequently been observed that for certain problem classes, the associated generalized Hessian of the augmented Lagrangian function remains consistently positive definite. Nevertheless, the fast local convergence rate the algorithm under more realistic assumptions (specifically, without imposing the singletonness assumption on the multiplier) and the characterization of the non-singularity of the generalized Hessian of the augmented Lagrangian function have yet to be fully explored.

The classical Euclidean ALM, originally proposed by Hestenes [17] and Powell [30] for equality constrained problems, was later expanded to nonlinear programming (NLP) by Rockafellar [32]. Over the decades, the convergence analysis of ALM under the Euclidean setting has been extensively studied. In literature, it is posited that the local linear convergence rate of ALM for NLP requires both the linear independence constraint qualification (LICQ) and the second-order sufficient condition (SOSC) (see e.g., [5, 10, 28]). For non-convex non-polyhedral problems, such as nonlinear second-order cone programs and semidefinite programs, Liu and Zhang [27] and Sun et al. [39] obtain the local linear convergence rate of ALM under the constraint non-degeneracy condition and the strong SOSC at Karush-Kuhn-Tucker (KKT) points, respectively. More recently, for $\mathcal{C}^2$-cone reducible constraints, Kanzow and Steck [23] obtained the primal-dual linear convergence rate under the strong Robinson constraint qualification (SRCQ) and SOSC. These analyses for the Euclidean ALM all hinge on the uniqueness of the multipliers. Turning to the RALM, the first convergence result is established in [42] under the constant positive linear dependence constraint qualification and LICQ. Subsequently, the local linear convergence rate for the RALM was provided in [43] under the manifold strict Robinson constraint qualification (M-SRCQ) and the manifold second-order sufficient condition (M-SOSC). Nevertheless, existing convergence results for the RALM still necessitate a unique multiplier. Interestingly, we have observed that the RALM can still perform effectively in certain scenarios even when the multiplier set is not a singleton. This observation has driven us to explore alternative conditions that could ensure the local linear convergence rate of the RALM without requiring the uniqueness of the multipliers.

The recent studies [36, 41] shed light on the possibility of relaxing the uniqueness assumption of the multipliers under the so-called (strong) variational sufficient condition. In [35], Rockafellar introduces the concept of variational (strong) convexity, which requires the function value and the subdifferential of the non-convex function to be locally identical to a (strongly) convex function. By utilizing this property, the (strong) variational sufficient condition for optimizations in Euclidean setting is built up by requiring the perturbed augmented objective function to be variationally (strongly) convex with respect to a first-order stationary point. It is proven in [35] that the strong variational sufficient condition implies the local strong convexity of the augmented Lagrangian function and the augmented tilt stability. Furthermore, it is demonstrated that strong variational sufficiency is equivalent to the strong SOSC if the function θ in (1.1) (when the manifold $\mathcal{M}$ is taken as a Euclidean space) is a polyhedral function ([35]) or an indicator function of a second-order cone or semidefinite cone ([41]). By leveraging the strong convexity of the augmented Lagrangian function, one may be able to construct an augmented dual problem locally around a KKT pair and apply the PPA to this dual problem to achieve a local linear rate. Rockafellar [36] then extends the classical results in [33], which posit the equivalence of the dual PPA and primal ALM for convex problems, to obtain the local linear convergence rate of ALM under the strong variational sufficient condition. Inspired by this approach, our objective is to extend the (strong) variational sufficient condition to manifold optimization. The first challenge here is defining convexity on a manifold. It is widely understood that translating the concept of convexity from Euclidean settings to functions and sets on a Riemannian manifold is not straightforward, primarily because a standard 'line segment' connecting two points x and y on a manifold cannot be represented by a convex combination of x and y. To address this issue, we propose an alternative problem that is locally equivalent to (1.1). Suppose that $(\bar{x}, \bar{y})$ is a stationary point of problem (1.2), as defined in (3.5). Let $R_{\bar{x}}$ denote the retraction and $T_{\bar{x}}\mathcal{M}$ denote the tangent space of $\mathcal{M}$ at $\bar{x}$ (see Section 2 for definitions). We locally transform the manifold problem (1.1) into an optimization problem on the tangent space $T_{\bar{x}}\mathcal{M}$ at $\bar{x}$ as follows:

$$\begin{array}{ll} \min & f \circ R_{\bar{x}}(\xi) + \theta(g \circ R_{\bar{x}}(\xi)) \\ \text{s.t.} & \xi \in T_{\bar{x}}\mathcal{M}. \end{array} \tag{1.5}$$

This problem is locally equivalent to (1.1) if we set $x = R_{\bar{x}}(\xi)$. Compared with (1.1), problem (1.5) is more approachable due to $T_{\bar{x}}\mathcal{M}$ being a Euclidean space. Consequently, we can now explore properties related to the convexity of (1.5).

In this paper, we establish the ALM for solving (1.5), which is locally equivalent to RALM (1.4) for solving (1.1). By assuming the variational sufficient condition for problem (1.5) (a property we will refer to as the manifold variational sufficient condition), we are able to construct a local augmented dual problem in Euclidean space for (1.1). Furthermore, we discover that the manifold strong variational sufficient condition is equivalent to the manifold strong second-order sufficient condition (M-SSOSC) under certain circumstances. This also implies that the manifold strong variational sufficient condition is independent of the choice of retractions. Consequently, by applying PPA to the local dual problem, we ultimately obtain R-linear convergence rate of the primal iterations of RALM and Q-linear rate of the multipliers sequence. Moreover, we show that the M-SSOSC is also equivalent to the non-singularity of the generalized Hessian of the augmented Lagrangian function, which ensures the efficiency of the semismooth Newton method when solving the subproblem of RALM.

The rest of the paper is organized as follows. In Section 2, we revisit some fundamentals of smooth manifolds and nonsmooth analysis. In Section 3, we define the local equivalent problem for (1.1) in the tangent space and explore the relation of Lagrangian functions and first-order conditions between these two problems. The variational sufficient condition is discussed in Section 4. Moreover, the local duality under variational sufficient condition is given in this section. Section 5 establishes the local convergence analysis of RALM. The semismooth Newton method for solving RALM subproblem and its local convergence rate are discussed in Section 6. The applications and numerical results are presented in Section 7. Finally, we give our conclusion in Section 8.

2. Preliminaries and notations. We begin by introducing some basic concepts of manifolds that will be used throughout the discussion. Most of the following properties can be found in the literature, specifically in [2, 26].

Let $\mathcal{M}$ be an n-dimensional smooth manifold, and consider any point $x \in \mathcal{M}$. We define $\mathfrak{F}_x(\mathcal{M})$ as the set of all smooth real-valued functions on a neighborhood of x. The tangent vector at x, denoted ξ_x, is a mapping from $\mathfrak{F}_x(\mathcal{M})$ to $\mathbb{R}$. This mapping can be characterized by 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 $f \in \mathfrak{F}_x(\mathcal{M})$. The tangent space at x, denoted $T_x\mathcal{M}$, 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 $T\mathcal{M} := \bigcup_{x\in\mathcal{M}} T_x\mathcal{M}$, 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. 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 $DF(x)$ is reduced to the classical definition of 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 Riemannian 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 a manifold $\mathcal{M}$ is a smoothly varying inner product defined on the tangent space $T_x\mathcal{M}$ at each point $x \in \mathcal{M}$. A Riemannian manifold is a differentiable manifold whose tangent spaces are equipped with these smoothly varying inner products. For simplicity, we often omit the subscript x and write $\langle\cdot,\cdot\rangle$ for the Riemannian metric. The induced norm of a tangent vector $\xi \in T_x\mathcal{M}$ is given by $\|\xi\| = \sqrt{\langle\xi,\xi\rangle}$. Given a smooth function $f \in \mathfrak{F}_x(\mathcal{M})$, the Riemannian gradient of f at x, denoted by $\operatorname{grad} f(x)$, is the unique tangent vector in $T_x\mathcal{M}$ that satisfies $\langle \operatorname{grad} f(x), \xi\rangle = \xi_x f$ for all $\xi \in T_x\mathcal{M}$.

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

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 $DR_x(0_x) = \mathrm{id}_{T_x\mathcal{M}}$. Let R_x denote the restriction of R to $T_x\mathcal{M}$. 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

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

Suppose that a subset $\mathcal{A} \subset \mathcal{M}$. 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 semi-continuous (lsc) if $\operatorname{epi} F$ is closed. When considering the manifold $\mathcal{M}$ as a topology 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 topology space. The following definition of the Fréchet subdifferential is taken from [4, Corollary 4.5].

Definition 2.1. *Let $f : \mathcal{M} \to \mathbb{R}$ 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 $x \in \operatorname{dom} f = \{x \in M : f(x) < \infty\}$ is defined as*

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

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

With the distance function $d(\cdot,\cdot)$ defined above, the Lipschitz property of functions can be extended to manifold. A function $f : \mathcal{M} \to \mathbb{R}$ is Lipschitz of rank $L > 0$ 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 of rank L on $\mathcal{U}$, we say that f is Lipschitz of rank L at x; if for every $x \in \mathcal{M}$, f is Lipschitz of rank L at x for some $L > 0$, then f is said to be locally Lipschitz on $\mathcal{M}$. The generalized directional derivative of a locally Lipschitz function f at $x \in \mathcal{M}$ in the direction $v \in T_x\mathcal{M}$, is defined in [18] 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 function f at $x \in \mathcal{M}$, denoted by $\partial_C f(x)$, is defined as

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

According to [18, 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 2.2. *Let $\mathcal{M}$ be a Riemannian manifold and $x \in \mathcal{M}$. Suppose that $f : \mathcal{M} \to \mathbb{R}$ is Lipschitz 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.

As mentioned in the introduction, a useful tool for analyzing the local convergence of ALM is the variational sufficient condition. We first introduce the variational convexity proposed in [35].

DEFINITION 2.3. *Given a lsc function $F : \mathbb{R}^n \to (-\infty, \infty]$, the variational convexity of F with respect to a pair $(\bar{w}, \bar{z}) \in \operatorname{gph} \partial F$ is said to hold if there exist open convex neighborhoods $\mathcal{W}$ of $\bar{w}$ and $\mathcal{Z}$ of $\bar{z}$ such that there exists a proper lsc convex function $h \leq F$ 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, for (w, z) belonging to this common set, $h(w) = F(w)$. If h is strongly convex on $\mathcal{W}$, we say that F is variationally strongly convex with respect to $(\bar{w}, \bar{z})$.

For a Euclidean optimization problem, the variational sufficient condition is said to hold at a first-order stationary point if the augmented perturbed objective function φ^ρ is variationally convex, with $\varphi^\rho(x, u) = \varphi(x, u) + \frac{\rho}{2}\|u\|^2$. The strong variational sufficient condition holds if φ^ρ is variationally strongly convex.

3. The localization problems and the Lagrangian functions. At a given point $x \in \mathcal{M}$, the optimization problem (1.1) can be locally transformed into an equivalent optimization problem on the tangent space $T_x\mathcal{M}$ by employing a retraction $R_x : T_x\mathcal{M} \to \mathcal{M}$. This local transformation relies on the inverse function theorem [26, Theorem 4.5], which ensures that any retraction R_x is a diffeomorphism within a neighborhood of the zero vector $0_x \in T_x\mathcal{M}$. Thus, within this neighborhood, the optimization problem on the manifold can be treated as an optimization problem in the tangent space $T_x\mathcal{M}$. The injectivity radius of a Riemannian manifold with respect to retraction, as defined in [19], refers to the largest radius within which the retraction map is a diffeomorphism. This concept aligns with the classical definition of injectivity radius [6, Definition 10.19], when the retraction is chosen as the exponential mapping.

DEFINITION 3.1. *The injectivity radius of manifold $\mathcal{M}$ at a point x with respect to retraction R_x, denoted by r_{R_x}, is the supremum over radii $r > 0$ such that R_x is defined and is a diffeomorphism on the open ball $B_x(r) = \{v \in \mathrm{T}_x\mathcal{M} : \|v\| < r\}$. By the inverse function theorem, $r_{R_x} > 0$.*

Additionally, as outlined in [6, 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 construct a problem that is locally equivalent to (1.1), we begin by introducing the following function defined on the tangent space at a given point $x \in \mathcal{M}$. Definition 3.2 on tangent space is different from the pullback function defined in [2, Section 4], where the latter is formulated as $f \circ R$ at each point on the manifold.

DEFINITION 3.2. *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 $F : \mathcal{M} \to \mathbb{R}$, we define $F_{R_x} : T_x\mathcal{M} \to (-\infty, \infty]$ by*

$$F_{R_x}(\xi) = \begin{cases} F(R_x\xi), & \xi \in B_x(r_{R_x}), \\ +\infty, & \xi \notin B_x(r_{R_x}). \end{cases} \tag{3.1}$$

It is worth noting that the lower semicontinuity of F is inherited by F_{R_x} within the ball $B_x(r_{R_x})$. However, this continuity may not hold on its boundary. Fortunately, this limitation does not affect our discussions since our focus is solely on the properties of F_{R_x} within the ball $B_x(r_{R_x})$.

For a given point $\bar{x}$ and a retraction $R_{\bar{x}}$, applying (3.1) to f and g, we obtain the following problem on tangent space $T_{\bar{x}}\mathcal{M}$ as

$$\begin{array}{ll} \min & f_{R_{\bar{x}}}(\xi) + \theta(g_{R_{\bar{x}}}(\xi)) \\ \text{s.t.} & \xi \in T_{\bar{x}}\mathcal{M}, \end{array} \tag{3.2}$$

where $f_{R_{\bar{x}}}$ and $g_{R_{\bar{x}}}$ are defined by (3.1). By Definitions 3.1 and 3.2, problem (3.2) is locally equivalent to problem (1.1) in the injectivity ball $B_{\bar{x}}(r_{R_{\bar{x}}})$ as we can build the one-to-one relationship $x = R_{\bar{x}}(\xi)$ for any $\xi \in B_{\bar{x}}(r_{R_{\bar{x}}})$. The perturbed problem for (3.2) can be written as

$$\begin{array}{ll} \min & \varphi_{R_{\bar{x}}}(\xi, u) := f_{R_{\bar{x}}}(\xi) + \theta(g_{R_{\bar{x}}}(\xi) + u) \\ \text{s.t.} & \xi \in T_{\bar{x}}\mathcal{M}, \end{array} \tag{3.3}$$

when $u = 0$ this is problem (3.2). This problem is also locally equivalent to the manifold perturbed problem (1.2) insider the ball $B_{\bar{x}}(r_{R_{\bar{x}}})$. The Lagrangian function for (3.2) is defined by

$$l_{R_{\bar{x}}}(\xi, y) = \inf_u \{\varphi_{R_{\bar{x}}}(\xi, u) - \langle y, u\rangle\} = L_{R_{\bar{x}}}(\xi, y) - \theta^*(y),$$

where $L_{R_{\bar{x}}}(\xi, y) := f_{R_{\bar{x}}}(\xi) + \langle y, g_{R_{\bar{x}}}(\xi)\rangle$. For $\rho > 0$, the augmented Lagrangian function of (3.2) is given by

$$l^{\rho}_{R_{\bar{x}}}(\xi, y) = \inf_u \left\{\varphi_{R_{\bar{x}}}(\xi, u) - \langle y, u\rangle + \frac{\rho}{2}\|u\|^2\right\}. \tag{3.4}$$

The augmented Lagrangian functions (1.3) and (3.4) can be regraded as the Lagrangian functions of the augmented objective functions $\varphi^{\rho}(x,u) = \varphi(x,u) + \frac{\rho}{2}\|u\|^2$ and $\varphi^{\rho}_{R_{\bar{x}}}(\xi, u) = \varphi_{R_{\bar{x}}}(\xi, u) + \frac{\rho}{2}\|u\|^2$, respectively. Thus by definition, $-l^{\rho}(x,\cdot)$ 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. Moreover, the lower semicontinuity of θ implies that $\varphi^{\rho}(x,\cdot)$ and $\varphi^{\rho}_{R_{\bar{x}}}(\xi,\cdot)$ are closed functions of u. Therefore, by [31, 12.2] we have

$$\begin{array}{rll} & \varphi(x,u) = \sup_y\{l(x,y) + \langle y,u\rangle\}, & \varphi^{\rho}(x,u) = \sup_y\{l^{\rho}(x,y) + \langle y,u\rangle\} \\ \text{and} & \varphi_{R_{\bar{x}}}(\xi,u) = \sup_y\{l_{R_{\bar{x}}}(\xi,y) + \langle y,u\rangle\}, & \varphi^{\rho}_{R_{\bar{x}}}(\xi,u) = \sup_y\{l^{\rho}_{R_{\bar{x}}}(\xi,y) + \langle y,u\rangle\}. \end{array}$$

For a given $\bar{x} \in \mathcal{M}$, we say $\bar{x}$ is a stationary point of the perturbed problem (1.2), if there exists $\bar{y} \in \mathbb{Y}$ such that

$$(0, \bar{y}) \in \partial\varphi(\bar{x}, 0). \tag{3.5}$$

The following proposition elucidates the relationships between the first-order conditions for problems (1.2) and (3.3).

Proposition 3.3. *Given $(\bar{x}, \bar{y})$ and a retraction $R_{\bar{x}}$, the following statements are equivalent:*

(i) $(\bar{x}, \bar{y})$ *satisfies condition* (3.5) *of problem (1.2);*

(ii) $(0_{\bar{x}}, \bar{y})$ *satisfies the following condition of problem (3.3), i.e.,*

$$(0, \bar{y}) \in \partial\varphi_{R_{\bar{x}}}(0_{\bar{x}}, 0); \tag{3.6}$$

(iii) For any $\rho > 0$, $(\bar{x}, \bar{y})$ *satisfies* $(0, \bar{y}) \in \partial\varphi^{\rho}(\bar{x}, 0)$*;*

(iv) For any $\rho > 0$, $(0_{\bar{x}}, \bar{y})$ *satisfies* $(0_{\bar{x}}, \bar{y}) \in \partial\varphi^{\rho}_{R_{\bar{x}}}(0_{\bar{x}}, 0)$*;*

(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-Yosida regularization of* θ *defined by* $\operatorname{env}_{\rho}\theta(p) := \min_{y\in\mathbb{Y}}\theta(y) + \frac{\rho}{2}\|p - y\|^2$, *for any* p*;*

(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. The equivalence of (3.5) and (3.6) is obtained by using Proposition 2.2. While taking $u = 0$, it is obvious that $\partial\varphi(\bar{x}, u) = \partial\varphi^{\rho}(\bar{x}, u)$ and $\partial\varphi^{\rho}_{R_{\bar{x}}}(0_{\bar{x}}, 0) = \partial\varphi_{R_{\bar{x}}}(0_{\bar{x}}, 0)$, which implies that $(i) \iff (ii) \iff (iii) \iff (iv)$. By the chain rules of the subdifferential of manifold functions [43, Proposition 1], we have

$$\begin{aligned}(v, y) \in \partial\varphi(x, u) &\iff y \in \partial\theta(g(x) + u),\ v = \operatorname{grad}_x L(x, y)\\ &\iff v \in \partial_x l(x, y),\ u \in \partial_y[-l](x, y).\end{aligned}$$

Moreover, Proposition 2.2 yields

$$\begin{aligned}(v, y) \in \partial\varphi(x, u) \iff (v, y) \in \partial\varphi_{R_x}(0_x, u) &\iff y \in \partial\theta(g_{R_x}(0_x) + u),\ v = \nabla_\xi l_{R_x}(0_x, y)\\ &\iff v \in \partial_\xi l_{R_x}(0_x, y),\ u \in \partial_y[-l_{R_x}](0_x, y).\end{aligned}$$

Similarly, we can prove the equivalence between (i), (vii) and (viii). ∎

4. Manifold variational sufficient condition. As mentioned in the introduction, the variational sufficient condition is closely related to the local maximal monotonicity of the augmented objective function and is crucial for the linear convergence analysis of ALM in Euclidean settings. In the Riemannian case, we first define the manifold (strong) variational sufficient condition through $\varphi^{\rho}_{R_{\bar{x}}}$ as follows.

DEFINITION 4.1. *The manifold (strong) variational sufficient condition for local optimality in (1.2) under retraction $R_{\bar{x}}$ is said to hold with respect to $(\bar{x}, \bar{y})$ satisfying condition* (3.5) *if the (strong) variational sufficiency condition holds for problem (3.3) under $R_{\bar{x}}$.*

We will show in Section 4.2 that this definition of manifold strong variational sufficiency is actually independent of the choice of retraction $R_{\bar{x}}$.

4.1. Local augmented duality. Inspired by [35] and [36], this section explores the local augmented dual property for the problem (1.2). We present a proposition that establishes the equivalence between the manifold (strong) variational sufficiency and the (strong) convexity of the augmented Lagrangian. This equivalence directly follows from [35, Theorem 1] and Proposition 2.2.

PROPOSITION 4.2. *Let $\bar{x}$ and $\bar{y}$ satisfy the condition* (3.5). *For a given retraction $R_{\bar{x}}$, the manifold (strong) variational sufficiency with respect to $(\bar{x}, \bar{y})$ for problem* (1.2) *at level $\bar{\rho}$ is equivalent to the existence of a closed convex neighborhood $\mathcal{W} \times \mathcal{Y}$ of $(0_{\bar{x}}, \bar{y})$ such that:*

(i) The function $l^{\bar{\rho}}_{R_{\bar{x}}}(\xi, y)$ is locally (strongly) convex at $0_{\bar{x}}$ for $y \in \mathcal{Y}$.

(ii) The function $l^{\bar{\rho}}_{R_{\bar{x}}}(\xi, y)$ is concave in $y \in \mathcal{Y}$ for $\xi \in \mathcal{W} \subseteq B_{\bar{x}}(r_{R_{\bar{x}}})$.

Under these conditions, $(\bar{x}, \bar{y})$ is a saddle point of $l^{\bar{\rho}}(x, y)$ with respect to minimizing over $x \in R_{\bar{x}}(\mathcal{W})$ and maximizing over $y \in \mathcal{Y}$. Furthermore, for any $\rho \geq \bar{\rho}$, the function $l^{\rho}_{R_{\bar{x}}}(\xi, y)$ retains these properties, making $(\bar{x}, \bar{y})$ a saddle point of $l^{\rho}(x, y)$ relative to $R_{\bar{x}}(\mathcal{W}) \times \mathcal{Y}$.

Proof. By applying [35, Theorem 1] and Proposition 2.2 to (3.3) at $(0_{\bar{x}}, \bar{y})$, $l^{\bar{\rho}}_{R_{\bar{x}}}(\xi, y)$ is convex in $\xi \in \mathcal{W}$ when $y \in \mathcal{Y}$ as well as concave in $y \in \mathcal{Y}$ when $\xi \in \mathcal{W}$. It remains to show that $(\bar{x}, \bar{y})$ is a saddle point of $l^{\bar{\rho}}(x, y)$ in $R_{\bar{x}}(\mathcal{W}) \times \mathcal{Y}$, or $l^{\bar{\rho}}(x, y)$ attains its minimum at $\bar{x}$ in $R_{\bar{x}}(\mathcal{W})$. This is shown by

$$l^{\bar{\rho}}(\bar{x}, \bar{y}) = l^{\bar{\rho}}_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y}) \leq l^{\bar{\rho}}_{R_{\bar{x}}}\left(R^{-1}_{\bar{x}}(x), \bar{y}\right) = l^{\bar{\rho}}(x, \bar{y}) \quad \forall\, x \in R_{\bar{x}}(\mathcal{W}).$$

Hence we complete the proof. ∎

If we require the strong variational sufficiency condition to be satisfied at $(0_{\bar{x}}, \bar{y})$ for problem (3.3) at level $\bar{\rho}$, by [35, Theorem 2], the augmented tilt stability holds at $0_{\bar{x}}$ for problem (3.3). Additionally, we define the augmented tilt stability on manifold with respect to retraction.

Definition 4.3. *The manifold augmented tilt stability is said to hold at $\bar{x} \in \mathcal{M}$ with respect to retraction $R_{\bar{x}}$ if there is a neighborhood $\mathcal{V}$ of $0_{\bar{x}}$ such that the mapping*

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

is single-valued and Lipschitz continuous. Here, $\mathcal{W}$ is the neighborhood defined in Definition 2.3.

The next proposition is an augmented tilt characterization of manifold strong variational sufficiency.

Proposition 4.4. *Suppose the manifold strong variational sufficient condition holds at $(\bar{x}, \bar{y})$ with respect to $R_{\bar{x}}$ for local optimality of (1.2) at level $\bar{\rho}$, then 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 that $h(v, y) \subseteq R_{\bar{x}} h_{R_{\bar{x}}}(v, y)$. The converse relation is also true, hence $h(v, y) = R_{\bar{x}} h_{R_{\bar{x}}}(v, y)$. Therefore, the single-valued and Lipschitz continuous properties of $h(v, y)$ is equivalent to $h_{R_{\bar{x}}}(v, y)$ being single-valued and Lipschitz continuous. Thus, the augmented tilt stability will hold for $\bar{x}$ and $0_{\bar{x}}$ simultaneously. By [35, Theorem 2], we obtain the conclusion. ∎

In [36], the convergence of ALM is proved by applying local PPA to the local dual problem and using the convergence of PPA. Now by Propositions 4.2 and 4.4, we are able to establish the local augmented dual problem for problem (1.1) under the manifold (strong) variational sufficiency. Assume that the manifold variational sufficient condition holds at the stationary point $(\bar{x}, \bar{y})$ under retraction $R_{\bar{x}}$. Denote $\mathcal{S}$ as the set of all (x, y) satisfying condition (3.5). By Proposition 4.2, there is a closed convex neighborhood $\mathcal{W} \times \mathcal{Y}$ of $(0_{\bar{x}}, \bar{y})$ such that $l^{\bar{\rho}}_{R_{\bar{x}}}(\xi, y)$ is convex in $\xi \in \mathcal{W}$ when $y \in \mathcal{Y}$ as well as concave in $y \in \mathcal{Y}$ when $\xi \in \mathcal{W}$, and $(\bar{x}, \bar{y})$ is a saddle point of $l^{\bar{\rho}}(x, y)$ as well as $(0_{\bar{x}}, \bar{y})$ is the saddle point of $l^{\bar{\rho}}_{R_{\bar{x}}}(\xi, y)$. The choice of $\mathcal{W}$ and $\mathcal{Y}$ ensures that $\mathcal{S} \cap [\operatorname{int} R_{\bar{x}}(\mathcal{W}) \times \operatorname{int} \mathcal{Y}] \neq \emptyset$.

To align with the traditional convex analysis of the primal and dual problems, we define the local perturbed objective function as in [36]:

$$\begin{aligned} &\Phi(x, u) = \sup\nolimits_{y \in \mathcal{Y}} \{ l^{\bar{\rho}}(x, y) + \langle y, u \rangle \} \text{ for } x \in R_{\bar{x}}(\mathcal{W}), \qquad \Phi(x, u) = \infty \text{ for } x \notin R_{\bar{x}}(\mathcal{W}), \\ &\Psi(v, y) = \inf\nolimits_{x \in R_{\bar{x}}(\mathcal{W})} \{ l^{\bar{\rho}}(x, y) - \left\langle v, R_{\bar{x}}^{-1} x \right\rangle \} \text{ for } y \in \mathcal{Y}, \quad \Psi(v, y) = -\infty \text{ for } y \notin \mathcal{Y}. \end{aligned}$$

The following definition establishes the local primal and dual problems for manifold optimization.

Definition 4.5. *The associated local primal problem for (1.2) is defined as*

$$(P) \qquad \min_{x \in R_{\bar{x}}(\mathcal{W})} F(x), \text{ where } F(x) = \Phi(x, 0) = \sup_{y \in \mathcal{Y}} l^{\bar{\rho}}(x, y) \text{ for } x \in R_{\bar{x}}(\mathcal{W}),$$

while the local dual problem is defined as

$$(D) \qquad \max_{y \in \mathcal{Y}} H(y), \text{ where } H(y) = \Psi(0_{\bar{x}}, y) = \inf_{x \in R_{\bar{x}}(\mathcal{W})} l^{\bar{\rho}}(x, y) \text{ for } y \in \mathcal{Y}.$$

The primal and dual connections of the optimal values and solutions between (P) and (D) are given in the following theorem.

THEOREM 4.6. *Let $(\bar{x}, \bar{y})$ satisfy the condition* (3.5)*. Suppose the manifold variational sufficient condition for local optimality holds at $(\bar{x}, \bar{y})$ with respect to $R_{\bar{x}}$. Then the problems (P) and (D) are defined in the neighborhood $R_{\bar{x}}(\mathcal{W}) \times \mathcal{Y}$ of $(\bar{x}, \bar{y})$ and have optimal solutions. Additionally, x solves (P) if and only if x minimizes the problem (1.2) relative to $R_{\bar{x}}(\mathcal{W})$. Furthermore, for a pair $(\bar{x}, y^*) \in \operatorname{int} R_{\bar{x}}(\mathcal{W}) \times \operatorname{int} \mathcal{Y}$, the following conditions are equivalent and ensure that $\bar{x}$ is locally optimal relative to $R_{\bar{x}}(\mathcal{W})$ in (1.2). These conditions also guarantee that the objective value $\varphi(\bar{x}, 0)$ agrees with the optimal values in (P) and (D), as well as with $l^{\bar{\rho}}(\bar{x}, y^*)$ and $l^{\bar{\rho}}(\bar{x}, \bar{y})$:*

(a) $\bar{x}$ minimizes in (P) and y^* maximizes in (D),

(b) $(\bar{x}, y^*)$ is a saddle point of $l^{\bar{\rho}}$ on $R_{\bar{x}}(\mathcal{W}) \times \mathcal{Y}$,

(c) $(\bar{x}, y^*)$ is a saddle point of l^{ρ} on $R_{\bar{x}}(\mathcal{W}) \times \mathcal{Y}$ for any $\rho \geq \bar{\rho}$.

Proof. Define the primal and dual problems for (3.3) as follows:

$$\min_{\xi \in \mathcal{W}} F_{R_{\bar{x}}}(\xi), \text{ where } F_{R_{\bar{x}}}(\xi) = \Phi_{R_{\bar{x}}}(\xi, 0) = \sup_{y \in \mathcal{Y}} l^{\bar{\rho}}_{R_{\bar{x}}}(\xi, y) \text{ for } \xi \in \mathcal{W},$$

$$(D1) \qquad \max_{y \in \mathcal{Y}} H_{R_{\bar{x}}}(y), \text{ where } H_{R_{\bar{x}}}(y) = \Psi_{R_{\bar{x}}}(0_{\bar{x}}, y) = \inf_{\xi \in \mathcal{W}} l^{\bar{\rho}}_{R_{\bar{x}}}(\xi, y) \text{ for } y \in \mathcal{Y},$$

where $\Phi(R_{\bar{x}}\xi, u) = \Phi_{R_{\bar{x}}}(\xi, u)$ and $\Psi(v, y) = \Psi_{R_{\bar{x}}}(v, y)$. By applying the proof of [36, Theorem 2.1] to the above problems at the point $(R_{\bar{x}}^{-1}\bar{x}, y^*)$ we obtain the results. □

4.2. Manifold strong variational sufficiency and second-order sufficient condition. Another important aspect of strong variational sufficiency is its equivalence to the strong second-order sufficient condition (SSOSC) in Euclidean settings. In our context, we also explore the relationship between the manifold strong variational sufficient condition and the manifold strong second-order sufficient condition (M-SSOSC).

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

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

Given the augmented Lagrangian function $l^{\rho}_{R_{\bar{x}}}(\xi, y)$ of (3.2) for any matrix H belonging to the Hessian bundle of $l^{\rho}_{R_{\bar{x}}}(\xi, y)$, it can be separated into four parts as $H_{\xi\xi}$, $H_{\xi y}$, $H_{y\xi}$ and H_{yy}. Let $\overline{\nabla}^2_{\xi\xi} l^{\rho}_{R_{\bar{x}}}(0_{\bar{x}}, y) := \{H_{\xi\xi} \mid H \in \overline{\nabla}^2 l^{\rho}_{R_{\bar{x}}}(\xi, y)\}$. The critical cone of function θ and g at x and y is defined by $\mathcal{C}_{\theta,g}(x, y) := \{d \in \mathbb{Y} \mid \theta'(g(x); d) = \langle d, y \rangle\}$. It is obvious that $\mathcal{C}_{\theta,g}(x, y) = \mathcal{C}_{\theta,g_{R_x}}(0_x, y)$ if $\mathcal{C}_{\theta,g_{R_x}}(0_x, y) := \{d \in \mathbb{Y} \mid \theta'(g_{R_x}(0_x); d) = \langle d, y \rangle\}$. Using [35, Theorem 3], we are able to connect the following manifold strong second-order condition with the manifold strong variational sufficiency for (1.2).

THEOREM 4.7. *Let $\bar{x}$ and $\bar{y}$ satisfy the condition (3.5). The manifold strong variational sufficient condition with respect to $(\bar{x}, \bar{y})$ under retraction $R_{\bar{x}}$ for (1.2) holds 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 $H_{\xi\xi} \in \overline{\nabla}^2_{\xi\xi} l^{\rho}_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y})$ has the form of*

$$(4.1) \qquad \operatorname{Hess}_x L(\bar{x}, \bar{y}) + Dg(\bar{x})^* G \operatorname{grad} g(\bar{x}) \textit{ for some } G \in \overline{\nabla}^2 \operatorname{env}_{\rho} \theta \left(g(\bar{x}) + \rho^{-1}\bar{y} \right).$$

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

$$(4.2) \qquad \langle \xi, \operatorname{Hess}_x L(\bar{x}; \bar{y})\, \xi \rangle > 0 \quad \forall Dg(\bar{x})\xi \in \operatorname{aff} \mathcal{C}_{\theta,g}(\bar{x}, \bar{y}) \setminus \{0\},$$

where $\operatorname{aff}\mathcal{C}_{\theta,g}(\bar{x},\bar{y})$ *represents the affine hull of the critical cone* $\mathcal{C}_{\theta,g}(\bar{x},\bar{y})$*.*

Moreover, if θ *is an indicator function of a second-order cone or positive semidefinite cone, then the manifold strong variational sufficient condition is equivalent to the M-SSOSC at* $(\bar{x},\bar{y})$*:*

$$\langle \xi, \operatorname{Hess}_x L(\bar{x};\bar{y})\,\xi\rangle - \sigma\left(\bar{y}, \mathcal{T}^2_{\mathcal{K}}\left(g(\bar{x}), Dg(\bar{x})\,\xi\right)\right) > 0 \quad \forall Dg(\bar{x})\xi \in \operatorname{aff}\mathcal{C}_{\theta,g}(\bar{x},\bar{y})\setminus\{0\}, \tag{4.3}$$

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* $\mathcal{D}$ *at* y *in direction* h*.*

Proof. Applying [35, Theorem 3], all $H_{\xi\xi} \in \overline{\nabla}^2_{\xi\xi} l^{\rho}_{R_{\bar{x}}}(0_{\bar{x}},\bar{y})$ are positive definite and of the form

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

Note that $\operatorname{grad}_x L(\bar{x},\bar{y}) = 0$. Therefore, by the definition of retraction and (2.1), we know that the above form is equivalent to (4.1). Furthermore, if θ is polyhedral convex, [35, Theorem 4] shows that the strong variational sufficiency holds if and only if

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

which is equivalent to (4.2) because $\nabla^2_{\xi\xi} l_{R_{\bar{x}}}(0_{\bar{x}},\bar{y}) = \operatorname{Hess}_x L(\bar{x};\bar{y})$, $g'_{R_{\bar{x}}}(0_{\bar{x}})\xi = Dg(\bar{x})\xi$ and $\mathcal{C}_{\theta,g_{R_{\bar{x}}}}(0_{\bar{x}},\bar{y}) = \mathcal{C}_{\theta,g}(\bar{x},\bar{y})$. If θ is an indicator function of a second-order cone or positive semidefinite cone, the manifold strong variational sufficiency is equivalent to (4.3) by [41] and the above analysis. Therefore the proof is completed. ∎

Remark 4.8. Theorem 4.7 establishes that the matrices in the Hessian bundle of $l^{\rho}_{R_{\bar{x}}}(0_{\bar{x}},\bar{y})$ remain unaffected by the choice of retraction $R_{\bar{x}}$. Consequently, we observe that the strong variational sufficient condition is inherently independent of the retraction. This property allows us to consider it as an intrinsic characteristic of manifold optimization problems.

Remark 4.9. Inspired by Remark 4.8, we aim to establish a special local convexity property that is also independence of the choice of retraction. Given an lsc function $f : \mathcal{M} \to (-\infty,\infty]$, we say the function is retractional (strongly) convex at x if for a retraction R_x, f_{R_x} is locally (strongly) convex on $T_x\mathcal{M}$. Moreover, we say the function is retractional variationally (strongly) convex with respect to a pair $(x,z) \in \operatorname{gph}\partial f$ if for a retraction R_x the Euclidean variationally (strongly) convexity of f_{R_x} holds with respect to $(0_x,z) \in \operatorname{gph}\partial f_{R_x}$ on $T_x\mathcal{M}\times T_x\mathcal{M}$. Remarkably, for a smooth function f, by [2, Proposition 5.5.6] we observe that the definitions remain independent of the chosen retraction at a critical point x (i.e., $\operatorname{grad} f(x) = 0$) for the strong cases.

Remark 4.10. It seems that our definition of retractional convexity is quite similar to the retraction-convexity defined in [22, Definition 3.2]. However, there is a crucial distinction: retraction-convexity in [22, Definition 3.2] necessitates holding on a subset of the manifold, whereas our definition of retractional convexity is defined at a specific point on the manifold. It appears that requiring convexity to be held on a subset of the manifold might be unnecessary for the local convergence analysis around the stationary point.

5. Convergence analysis of Riemannian ALM. In this section, we analyze the local convergence of RALM. Suppose that the manifold variational sufficient condition under retraction $R_{\bar{x}}$ holds at level $\bar{\rho}$ with respect to $\bar{x}$ and $\bar{y}$ satisfying condition (3.5). Let $\mathcal{W}$ be given by Proposition 4.2.

We take $\mathcal{U} = R_{\bar{x}}(\mathcal{W})$ and $\tilde{\rho}_k = \rho_k - \bar{\rho}$. Then (1.4) can be achieved by

$$\left\{\begin{array}{l} \xi^{k+1} \approx \bar{\xi}^{k+1} := \underset{\xi\in\mathcal{W}}{\operatorname{argmin}}\, l^{\rho_k}_{R_{\bar{x}}}(\xi, y^k), \\ x^{k+1} = R_{\bar{x}}(\xi^{k+1}), \\ y^{k+1} = y^k + \tilde{\rho}_k \nabla_y l^{\rho_k}_{R_{\bar{x}}}(\xi^{k+1}, y^k). \end{array}\right. \tag{5.1}$$

The first iteration of (5.1) 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 manifold using retraction $R_{\bar{x}}$.

We first apply PPA to the local dual problem (D). Let the solution set of (D) be $Z = \underset{y}{\operatorname{argmax}}\, H(y) = \underset{y}{\operatorname{argmax}}\, H_{R_{\bar{x}}}(y)$. Note that the dual problems (D) and $(D1)$ are identical ($H(y) = H_{R_{\bar{x}}}(y)$ for any $y \in \mathcal{Y}$). Therefore, applying PPA to (D) or $(D1)$ is the same. The PPA iteration is

$$\begin{aligned} y^{k+1} \approx x^k\left(y^k\right) \text{ with } x^k\left(y^k\right) &= \underset{y}{\operatorname{argmax}}\{H^k(y) := H(y) - \frac{1}{2c_k}\left\|y - y^k\right\|^2\} \\ &= \underset{y}{\operatorname{argmax}}\{H^k_{R_{\bar{x}}}(y) := H_{R_{\bar{x}}}(y) - \frac{1}{2c_k}\left\|y - y^k\right\|^2\}. \end{aligned} \tag{5.2}$$

The proximal parameters c_k satisfy $1 \le c_k \le c_\infty \le \infty$, the approximation is controlled by three stopping criteria

$$\left\|y^{k+1} - x^k\left(y^k\right)\right\| \le \left\{\begin{array}{ll} \text{(a)} & \varepsilon_k \\ \text{(b)} & \varepsilon_k \min\left\{1, \left\|y^{k+1} - y^k\right\|\right\} \\ \text{(c)} & \varepsilon_k \min\left\{1, \left\|y^{k+1} - y^k\right\|^2\right\}, \end{array}\right. \tag{5.3}$$

in which the error parameters satisfy

$$\varepsilon_k \in (0,1) \quad \text{and} \quad \sum_{k=0}^{\infty} \varepsilon_k = \sigma < \infty. \tag{5.4}$$

The next theorem presents the convergence of local PPA, as established in [36] and [34].

THEOREM 5.1. *Suppose the manifold variational sufficient condition holds at $(\bar{x}, \bar{y})$ with respect to the retraction $R_{\bar{x}}$. Let the initial point y^0 and the value σ in (5.4) satisfy the following closeness condition relative to Z:*

$$\exists\, \eta > \operatorname{dist}\left(y^0, Z\right) + \sigma \text{ such that } \mathcal{Y} \supset \left\{y \mid \|y - y^0\| < 3\eta\right\}. \tag{5.5}$$

Then the sequence $\{y^k\}$ generated by the proximal point iterations (5.2) under (5.3a) will remain within the interior of $\mathcal{Y}$ and converge to a point $y^* \in Z$ within the ball $\{y \mid \|y - y_0^*\| < \eta\} \subset \operatorname{int} \mathcal{Y}$, where y_0^* is the point in Z closest to y^0. Additionally, neither y^k nor $x^k(y^k)$ will exit this ball, and the dual objective values $H(y^k)$ will converge to the optimal value $H(y^*)$ in (D).

Proof. By [37, 11.48], $H_{R_{\bar{x}}}$, or equally, H is upper semicontinuous and concave in y, thus $\partial(-H)$ is maximal monotone [37, 12.17]. By taking $T = \partial(-H)$ in [34, Theorem 2.1] the results hold. ∎

From the convexity of $l^{\bar{\rho}}_{R_{\bar{x}}}(\cdot, y)$ we know that

$$\underset{\xi\in\mathcal{W}}{\operatorname{argmin}}\, l^{\bar{\rho}}_{R_{\bar{x}}}(\xi, y) = \left\{\xi \mid -\nabla_\xi l^{\bar{\rho}}_{R_{\bar{x}}}(\xi, y) \in N_{\mathcal{W}}(\xi)\right\}, \tag{5.6}$$

where $N_{\mathcal{W}}(\xi)$ is the normal cone to $\mathcal{W}$ at ξ. For the approximate minimization in (1.4) (where we take $\mathcal{U} = R_{\bar{x}}(\mathcal{W})$), three stopping criteria are given in [36, (1.15)] for the acceptability of x^{k+1}:

$$(5.7)\qquad \left(2\tilde{\rho}_k \left[l^{\rho_k}\left(x^{k+1}, y^k\right) - \inf_{R_{\bar{x}}(\mathcal{W})} l^{\rho_k}\left(\cdot, y^k\right)\right]\right)^{1/2} \leq \begin{cases} \text{(a)} & \varepsilon_k \\ \text{(b)} & \varepsilon_k \min\left\{1, \left\|\tilde{\rho}_k \nabla_y l^{\rho_k}\left(x^{k+1}, y^k\right)\right\|\right\} \\ \text{(c)} & \varepsilon_k \min\left\{1, \left\|\tilde{\rho}_k \nabla_y l^{\rho_k}\left(x^{k+1}, y^k\right)\right\|^2\right\}. \end{cases}$$

As a counterpart, three stopping criteria for the acceptability of ξ^{k+1} in the approximate minimization in (5.1) are

$$(5.8)\qquad \left(2\tilde{\rho}_k \left[l^{\rho_k}_{R_{\bar{x}}}\left(\xi^{k+1}, y^k\right) - \inf_{\mathcal{W}} l^{\rho_k}_{R_{\bar{x}}}\left(\cdot, y^k\right)\right]\right)^{1/2} \leq \begin{cases} \text{(a}'\text{)} & \varepsilon_k \\ \text{(b}'\text{)} & \varepsilon_k \min\left\{1, \left\|\tilde{\rho}_k \nabla_y l^{\rho_k}_{R_{\bar{x}}}\left(\xi^{k+1}, y^k\right)\right\|\right\} \\ \text{(c}'\text{)} & \varepsilon_k \min\left\{1, \left\|\tilde{\rho}_k \nabla_y l^{\rho_k}_{R_{\bar{x}}}\left(\xi^{k+1}, y^k\right)\right\|^2\right\}. \end{cases}$$

The iteration (1.4) under the stopping criteria (5.7) is equivalent to the iteration (5.1) under the stopping criteria (5.8).

THEOREM 5.2. *Suppose the manifold variational sufficient condition is satisfied at* $(\bar{x}, \bar{y})$ *at level* $\bar{\rho}$ *under retraction* $R_{\bar{x}}$*, and the sets (5.6) are nonempty and bounded when* $y \in \operatorname{int} \mathcal{Y}$*. Let the RALM (1.4) be initiated with* y^0 *satisfying the conditions in Theorem 5.1 and* $\tilde{\rho}_k = \rho_k - \bar{\rho}$*. Using the stopping criterion (5.7a), error parameters* ε_k *as in* (5.4)*, and stepsizes* $\tilde{\rho}_k$ *such that* $\tilde{\rho}_k \to \rho_\infty - \bar{\rho} \in (0, \infty]$*, the following estimate holds:*

$$(5.9)\qquad \left\|y^{k+1} - x^k\left(y^k\right)\right\|^2 \leq 2\tilde{\rho}_k\left[l^{\rho_k}\left(x^{k+1}, y^k\right) - \inf_{R_{\bar{x}}(\mathcal{W})} l^{\rho_k}\left(\cdot, y^k\right)\right],$$

With this estimate, the sequence $\left\{y^k\right\}$ *can be viewed as being generated by PPA in (5.2) with* $c_k = \tilde{\rho}_k$ *under the stopping criterion (5.3a), using the same error parameters* ε_k*.*

Moreover, the sequence $\left\{x^k\right\}$ *in* $R_{\bar{x}}(\mathcal{W})$ *is bounded, and each of its accumulation points will be a solution* $\bar{x}$ *to (P) and a minimizer of (1.2) relative to* $R_{\bar{x}}(\mathcal{W})$*. Therefore,* $\bar{x}$ *is locally optimal in (P) if it belongs to* $\operatorname{int} R_{\bar{x}}(\mathcal{W})$*.*

When implementing RALM with stopping criteria (5.7b) or (5.7c), this corresponds to implementing PPA with the respective stopping criteria (5.3b) or (5.3c).

Proof. The proof is quite similar with the proof of [36, Theorem 2.3]. Let the parameter c_k in the k-th PPA iteration (5.2) be $c_k = \tilde{\rho}_k$. Define $H^{c_k}_{R_{\bar{x}}}\left(y^k\right) := \max_y H^k_{R_{\bar{x}}}(y)$. By definition we have

$$(5.10)\qquad \nabla H^{c_k}_{R_{\bar{x}}}\left(y^k\right) = c_k^{-1}\left[x^k\left(y^k\right) - y^k\right], \text{ so that } x^k\left(y^k\right) = y^k + c_k \nabla H^{c_k}_{R_{\bar{x}}}\left(y^k\right).$$

Define the convex-concave function

$$(5.11)\qquad \hat{l}^k(\xi, y) := l^{\bar{\rho}}_{R_{\bar{x}}}(\xi, y) - \frac{1}{2c_k}\left\|y - y^k\right\|^2 \text{ for } \xi \in \mathcal{W} \text{ and } y \in \mathcal{Y}.$$

The k-th PPA iteration is therefore associated with the following primal and dual problems:

$$(P1\text{-}k)\qquad \min_{\xi \in \mathcal{W}} F^k_{R_{\bar{x}}}(\xi), \text{ where } F^k_{R_{\bar{x}}}(\xi) := \sup_{y \in \mathcal{Y}} \hat{l}^k(\xi, y),$$

$$(D1\text{-}k)\qquad \max_{y \in \mathcal{Y}}\left\{\inf_{\xi \in \mathcal{W}} \hat{l}^k(\xi, y) = H_{R_{\bar{x}}}(y) - \frac{1}{2c_k}\left\|y - y^k\right\|^2\right\}.$$

The unique solution of ($D1$-k) is $x^k(y^k)$ with the optimal value $H^{c_k}_{R_{\bar{x}}}(y^k)$. Similar to the proof in [36, Theorem 2.3], optimal solutions to both ($P1$-k) and ($D1$-k) exist, and these solutions are characterized by forming saddle points in (5.11). Furthermore, the optimal values in these two problems are equal. Therefore, there exists $\tilde{\xi}^k \in \mathcal{W}$, such that

$$\begin{aligned} H^{c_k}_{R_{\bar{x}}}\left(y^k\right) &= \widehat{F}^k\left(\tilde{\xi}^k\right) = \hat{l}^k\left(\tilde{\xi}^k, x^k\left(y^k\right)\right) = \max_{y\in\mathcal{Y}} \hat{l}^k\left(\tilde{\xi}^k, y\right) \\ &= \max_{y\in\mathcal{Y}} \{ l^{\bar{\rho}}_{R_{\bar{x}}}\left(\tilde{\xi}^k, y\right) - \frac{1}{2c_k}\left\|y - y^k\right\|^2 \}. \end{aligned} \tag{5.13}$$

If it further holds that $x^k(y^k) \in \operatorname{int}\mathcal{Y}$, the concavity of $l^{\bar{\rho}}_{R_{\bar{x}}}(\xi, y)$ in y yields that

$$\max_{y\in\mathcal{Y}} \{ l^{\bar{\rho}}_{R_{\bar{x}}}\left(\tilde{\xi}^k, y\right) - \frac{1}{2c_k}\left\|y - y^k\right\|^2 \} = \max_{y\in\mathbb{Y}} \{ l^{\bar{\rho}}_{R_{\bar{x}}}\left(\tilde{\xi}^k, y\right) - \frac{1}{2c_k}\left\|y - y^k\right\|^2 \},$$

and the latter part equals to $l^{\rho_k}_{R_{\bar{x}}}(\tilde{\xi}^k, y^k)$ by [37, 11.23]. Now we obtain that

$$\exists\, \tilde{\xi}^k \in \mathcal{W} \text{ such that } H^{c_k}_{R_{\bar{x}}}\left(y^k\right) = l^{\rho_k}_{R_{\bar{x}}}\left(\tilde{\xi}^k, y^k\right) \text{ if } x^k\left(y^k\right) \in \operatorname{int}\mathcal{Y}. \tag{5.14}$$

For any $y \in \mathcal{Y}$ and $\xi \in \mathcal{W}$, $H_{R_{\bar{x}}}(y) = \inf_{\xi'\in\mathcal{W}} l^{\bar{\rho}}_{R_{\bar{x}}}(\xi', y) \leq l^{\bar{\rho}}_{R_{\bar{x}}}(\xi, y)$. Consequently,

$$\begin{aligned} H^{c_k}_{R_{\bar{x}}}(y) &= \max_{y'} \{ H_{R_{\bar{x}}}(y') - \frac{1}{2\rho_k}\|y' - y\|^2 \} \\ &\leq \max_{y'} \{ l^{\bar{\rho}}_{R_{\bar{x}}}(\xi, y') - \frac{1}{2c_k}\|y' - y\|^2 \} = l^{\rho_k}_{R_{\bar{x}}}(\xi, y). \end{aligned} \tag{5.15}$$

Combining (5.14) with (5.15), we know that if $x^k\left(y^k\right) \in \operatorname{int}\mathcal{Y}$,

$$H^{c_k}_{R_{\bar{x}}}\left(y^k\right) = \min_{\xi\in\mathcal{W}} l^{\rho_k}_{R_{\bar{x}}}\left(\xi, y^k\right) = \min_{x\in R_{\bar{x}}(\mathcal{W})} l^{\rho_k}\left(x, y^k\right).$$

Now consider the vectors y^{k+1} in (5.1). By the concavity of $l^{\rho_k}\left(x^{k+1}, y\right)$ in y we have

$$l^{\rho_k}\left(x^{k+1}, y\right) \leq l^{\rho_k}\left(x^{k+1}, y^k\right) + \langle \nabla_y l^{\rho_k}\left(x^{k+1}, y^k\right), y - y^k\rangle \text{ for all } y \in \mathbb{Y}.$$

Moreover, by [37, 12.60] and the definition of $H^{c_k}_{R_{\bar{x}}}(y^k)$,

$$H^{c_k}_{R_{\bar{x}}}(y) \geq H^{c_k}_{R_{\bar{x}}}\left(y^k\right) + \langle \nabla H^{c_k}_{R_{\bar{x}}}\left(y^k\right), y - y^k\rangle - \frac{1}{2c_k}\left\|y - y^k\right\|^2 \text{ for all } y \in \mathbb{Y}.$$

Similar with (5.15), the definition of $H^{c_k}_{R_{\bar{x}}}(y)$ also implies $H^{c_k}_{R_{\bar{x}}}(y) \leq l^{\rho_k}(x, y)$ for any $x \in R_{\bar{x}}(\mathcal{W})$ and $y \in \mathcal{Y}$. Therefore,

$$H^{c_k}_{R_{\bar{x}}}\left(y^k\right) + \langle \nabla H^{c_k}_{R_{\bar{x}}}\left(y^k\right), y - y^k\rangle - \frac{1}{2c_k}\left\|y - y^k\right\|^2 \leq l^{\rho_k}\left(x^{k+1}, y^k\right) + \langle \nabla_y l^{\rho_k}\left(x^{k+1}, y^k\right), y - y^k\rangle,$$

which means $l^{\rho_k}\left(x^{k+1}, y^k\right) - H^{c_k}_{R_{\bar{x}}}\left(y^k\right) \geq \langle \nabla_y l^{\rho_k}\left(x^{k+1}, y^k\right) - \nabla H^{c_k}_{R_{\bar{x}}}\left(y^k\right), y - y^k\rangle - \frac{1}{2c_k}\left\|y - y^k\right\|^2$. The iterations (1.4) and (5.10) yield that

$$\nabla_y l^{\rho_k}\left(x^{k+1}, y^k\right) - \nabla H^{c_k}_{R_{\bar{x}}}\left(y^k\right) = c_k^{-1}\left[y^{k+1} - x^k\left(y^k\right)\right].$$

Therefore,

$$
\begin{aligned}
&c_k\left[l^{\rho_k}\left(x^{k+1}, y^k\right)-H_{R_{\bar{x}}}^{c_k}\left(y^k\right)\right] \\
&\geq \max_y\left\{\left\langle y^{k+1}-x^k\left(y^k\right), y-y^k\right\rangle-\frac{1}{2}\|y-y^k\|^2\right\}=\frac{1}{2}\left\|y^{k+1}-x^k\left(y^k\right)\right\|^2 .
\end{aligned} \tag{5.16}
$$

Theorem 5.1 ensures that $x^k(y^k) \in \operatorname{int} \mathcal{Y}$ and $\{y^k\}$ will converge to a solution $y^* \in \operatorname{int} \mathcal{Y}$ if y^0 is chosen through (5.5). Hence by (5.14), (5.16) is the estimation (5.9). Moreover, the approximation (5.7)(a,b,c) will lead to the PPA approximation (5.3)(a,b,c).

We now consider the sequence x^{k+1}. Define $F^k(x) := \sup_{y \in \mathcal{Y}}\ l(x, y)-\frac{1}{2 c_k}\left\|y-y^k\right\|^2$. it is true that $F^k(x)=F_{R_{\bar{x}}}^k(\operatorname{Exp}_{\bar{x}}^{-1} x)$, implying $\min_{\xi \in \mathcal{W}} F_{R_{\bar{x}}}^k(\xi)=\min_{x \in R_{\bar{x}}(\mathcal{W})} F^k(x)$. Together with (5.13) and (5.14), we have $\min_{x \in R_{\bar{x}}(\mathcal{W})} F^k(x)=\min_{x \in R_{\bar{x}}(\mathcal{W})} l^{\rho_k}\left(x, y^k\right)=H_{R_{\bar{x}}}^{c_k}\left(y^k\right)$. Since $\xi^{k+1}=R_{\bar{x}}^{-1} x^{k+1}$ and x^{k+1} is chosen under the stopping criterion (5.7a),

$$
\begin{aligned}
F^k\left(x^{k+1}\right)=F_{R_{\bar{x}}}^k(\xi^{k+1})=\sup_{y \in \mathcal{Y}}\ l_{R_{\bar{x}}}^{\bar{\rho}}(\xi, y)-\frac{1}{2 c_k}\left\|y-y^k\right\|^2 &\leq \max_{y \in \mathcal{Y}}\ l_{R_{\bar{x}}}^{\bar{\rho}}(\xi, y)-\frac{1}{2 c_k}\left\|y-y^k\right\|^2 \\
&\leq l_{R_{\bar{x}}}^{\rho_k}\left(\xi^{k+1}, y^k\right)=l^{\rho_k}\left(x^{k+1}, y^k\right) \leq \alpha_k:=H_{R_{\bar{x}}}^{c_k}\left(y^k\right)+\frac{\varepsilon_k^2}{2 c_k} .
\end{aligned} \tag{5.17}
$$

The definition of $H_{R_{\bar{x}}}^{c_k}\left(y^k\right)$ implies $H_{R_{\bar{x}}}\left(y^k\right) \leq H_{R_{\bar{x}}}^{c_k}\left(y^k\right) \leq \max H_{R_{\bar{x}}}$. By Theorem 5.1, $H_{R_{\bar{x}}}\left(y^k\right) \rightarrow \max H_{R_{\bar{x}}}$, hence $\alpha_k \rightarrow \alpha^*=\max(D_{\bar{x}})=\min(P_{\bar{x}})=\min(P)=\min_{x \in R_{\bar{x}}(\mathcal{W})} l^{\bar{\rho}}(x, y^*)$. Moreover, from the definition of F^k we have $F^k(x) \geq l^{\bar{\rho}}\left(x, y^k\right)$, which implies

$$
\begin{aligned}
\left\{x \in R_{\bar{x}}(\mathcal{W}) \mid F^k(x) \leq \alpha\right\} &\subset\left\{x \in R_{\bar{x}}(\mathcal{W}) \mid l^{\bar{\rho}}(x, y^*) \leq \alpha\right\} \\
&=R_{\bar{x}}\left\{\xi \in \mathcal{W} \mid l_{R_{\bar{x}}}^{\bar{\rho}}(\xi, y^*) \leq \alpha\right\} \text { for all } \alpha \in \mathbb{R},
\end{aligned}
$$

where sets on the right are bounded under the argmin assumption in [37, 3.23]. From (5.17), we have $x^{k+1} \in \{x \in R_{\bar{x}}(\mathcal{W}) \mid l^{\bar{\rho}}(x, y^*) \leq \alpha\}$ for any $\alpha \geq \alpha_k$. Thus, the sequence $\left\{x^k\right\}$ is bounded, and all its accumulation points belong to $\operatorname{argmin}_{R_{\bar{x}}(\mathcal{W})} l^{\bar{\rho}}(\cdot, y^*)$. ∎

COROLLARY 5.3. *Suppose the manifold strong variational sufficiency holds at* $(\bar{x}, \bar{y})$ *under retraction* $R_{\bar{x}}$. *The sequence* $\left\{x^k\right\}$ *generated by RALM must converge to that local solution* $\bar{x}$.

Proof. The manifold strong variational sufficiency holding at $(\bar{x}, \bar{y})$ refers to the isolated minimizing property of $\bar{x}$. By Theorem 5.2, $\{x^k\}$ will converge to the unique local solution $\bar{x}$. ∎

The local linear convergence rate of PPA iterations is obtained by [34, Theorems 3.2 and 3.3] as follows.

THEOREM 5.4. *In the circumstances of Theorem 5.1 with stopping criterion (5.3b) to get* $y^k \rightarrow \bar{y} \in Z=\operatorname{argmax}_{\mathcal{Y}} H_{R_{\bar{x}}}$, *suppose* $\exists\, b>0, \lambda>0$, *such that* $H_{R_{\bar{x}}}(y) \leq\left[\max_y H_{R_{\bar{x}}}\right]-b \operatorname{dist}^2(y, Z)$ *when* $\|y-\bar{y}\|<\lambda$. *Then* $\operatorname{dist}\left(y^k, Z\right) \rightarrow 0$ *at the Q-linear rate* $\rho=1 / \sqrt{1+b^2 \rho_{\infty}^2}$, *which is 0 when* $\rho_{\infty}=\infty$. *If the still tighter stopping criterion (5.3c) is used, then* $y^k \rightarrow \bar{y}$ *at that Q-linear rate* ρ.

A condition that ensures the fulfillment of the conditions stated in Theorem 5.4 is provided in [36, Theorem 4.2], and we now expand it to encompass our specific scenario in manifold optimization.

PROPOSITION 5.5. *Let* $G=\partial \theta(g(\bar{x}))$ *and* $M=\{y \mid 0=\operatorname{grad}_x L(\bar{x}, y)=\operatorname{grad} f(\bar{x})+D g(\bar{x})^* y\}$, *noting that* $G=\{y \mid \theta^*(y)=\theta^*(\bar{y})+\langle g(\bar{x}), y-\bar{y}\rangle\}$. *Suppose that* G *is polyhedral, and there exist* $b_0>0$ *and* $\lambda_0>0$ *such that, when* $\|y-\bar{y}\|<\lambda_0$, *it holds that*

$$
\theta^*(y) \geq \theta^*(\bar{y})+D g(\bar{x})^*(y-\bar{y})+b_0 \operatorname{dist}^2(y, G) .
$$

Furthermore, assume $\operatorname{grad} g(\bar{x}) \neq 0$ *and denote* $\beta(\operatorname{grad} g(\bar{x})) := \min\left\{\|Dg(\bar{x})^*\eta\| \mid \eta \in M^\perp, \|\eta\| = 1\right\}$. *Then* $\beta(\operatorname{grad} g(\bar{x})) > 0$ *and there exists* $\kappa_{G,M} > 0$ *such that condition in Theorem 5.4 holds for*

$$b = \frac{\kappa_{G,M}}{a_G + a_M} \quad \textit{with} \quad a_G = b_0^{-1} + 2\bar{\rho} \quad \textit{and} \quad a_M = \frac{2\left\|\operatorname{Hess} L(\bar{x}, \bar{y}) + \bar{\rho} I\right\|}{\beta(\operatorname{grad} g(\bar{x}))^2}.$$

Proof. Applying [36, Theorem 4.2] to problem (3.2) and using $\operatorname{grad}_x L(\bar{x}, y) = \nabla_\xi L_{R_{\bar{x}}}(0_{\bar{x}}, y)$ and $\operatorname{Hess} L(\bar{x}, \bar{y}) = \nabla_{\xi\xi} L_{R_{\bar{x}}}(0_{\bar{x}}, y)$ we are able to obtain the conclusion. ∎

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 iteration (1.4) has a unique solution, which we denote by $\bar{x}^{k+1}$. Next we give the local convergence rate of $\{x^k\}$ and $\{\bar{x}^k\}$ under the dual convergence.

THEOREM 5.6. *The convergence* $y^k \to \bar{y} \in Z$ *in the augmented Lagrangian method (1.4), as implemented in Theorem 5.2, induces both* $x^k \to \bar{x}$ *and* $\bar{x}^k \to \bar{x}$. *If* $\operatorname{dist}\left(y^k, Z\right) \to 0$ *Q-linearly at a rate* ρ *as* $y^k \to \bar{y}$, *then* $\bar{x}^k \to \bar{x}$ *R-linearly at that rate. Moreover, by requiring*

$$\left\|\operatorname{grad}_x l^{\rho_k}\left(x^{k+1}, y^k\right)\right\| \leq c\left\|y^{k+1} - y^k\right\| \text{ for some fixed } c, \tag{5.18}$$

if $y^k \to \bar{y}$ *Q-linearly at a rate* ρ, *then* $x^k \to \bar{x}$ *R-linearly at that rate.*

Proof. By Corollary 5.3, $x^k \to \bar{x}$. Let $\bar{\xi}^{k+1}$ denote the unique exact solution of the ALM subproblem in (5.1) and $\bar{x}^{k+1}$ denote the exact solution of the RALM subproblem in (1.4), we have $\bar{x}^{k+1} = R_{\bar{x}}(\bar{\xi}^{k+1})$ and $l^{\rho_k}_{R_{\bar{x}}}(\bar{\xi}^{k+1}, y^k) = l^{\rho_k}(\bar{x}^{k+1}, y^k)$. Since $\bar{\xi}^{k+1}$ minimize $l^{\rho_k}_{R_{\bar{x}}}(\cdot, y^k)$ over $\mathcal{W}$,

$$\left\langle \nabla_\xi l^{\rho_k}_{R_{\bar{x}}}(\bar{\xi}^{k+1}, y^k), 0_{\bar{x}} - \bar{\xi}^{k+1} \right\rangle \geq 0.$$

The strong convexity of $l^{\rho_k}_{R_{\bar{x}}}(\cdot, y^k)$ gives us

$$l^{\rho_k}_{R_{\bar{x}}}\left(0_{\bar{x}}, y^k\right) \geq l^{\rho_k}_{R_{\bar{x}}}\left(\bar{\xi}^{k+1}, y^k\right) + \left\langle \nabla_\xi l^{\rho_k}_{R_{\bar{x}}}\left(\bar{\xi}^{k+1}, y^k\right), 0_{\bar{x}} - \bar{\xi}^{k+1} \right\rangle + \frac{s}{2}\left\|0_{\bar{x}} - \bar{\xi}^{k+1}\right\|^2.$$

Combining with (5.14), we obtain

$$l^{\rho_k}\left(\bar{x}, y^k\right) = l^{\rho_k}_{R_{\bar{x}}}\left(0_{\bar{x}}, y^k\right) \geq l^{\rho_k}_{R_{\bar{x}}}\left(\bar{\xi}^{k+1}, y^k\right) + \frac{s}{2}\left\|\bar{\xi}^{k+1}\right\|^2 = H^{c_k}_{R_{\bar{x}}}(y^k) + \frac{s}{2}\left\|\bar{\xi}^{k+1}\right\|^2.$$

Moreover, $H^{c_k}_{R_{\bar{x}}}(y^k) \to l^{\bar{\rho}}(\bar{x}, \bar{y})$ through the proof of Theorem 5.2, implying $l^{\rho_k}\left(\bar{x}^{k+1}, y^k\right) \to l^{\bar{\rho}}(\bar{x}, \bar{y})$ and hence $\bar{x}^{k+1} \to \bar{x}$.

By the proof of [35, Theorem 2], the strong convexity of $l^{\rho}_{R_{\bar{x}}}(\cdot, y)$ corresponds to the Lipschitz property with modulus s^{-1} of the mapping $\lambda(v, y) := \operatorname{argmin}_{\xi \in \mathcal{W}}\{l^{\rho}_{R_{\bar{x}}}(\xi, y) - \langle v, \xi \rangle\}$. Therefore,

$$d(\bar{x}^{k+1}, \bar{x}) = \|\bar{\xi}^{k+1}\| \leq \frac{1}{s}\operatorname{dist}(y^k, Z),$$

and the Q-linear convergence of $\{y^k\}$ means that $\bar{x}^k \to \bar{x}$ R-linearly.

Since $0_{\bar{x}} = \lambda(0, \bar{y})$ and $\xi^{k+1} = \lambda\left(\nabla_\xi l^{\rho_k}_{R_{\bar{x}}}(\xi^{k+1}, y^k), y^k\right)$, the Lipschitz property of $\lambda(v, y)$ yields

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

Using the facts that $x^{k+1} = R_{\bar{x}}\xi^{k+1}$ and $\{\xi^k\}$ sequence is generated in the closed set $\mathcal{W}$, under the assumption (5.18), there exists a positive constant $L > 0$, such that

$$\left\|\nabla_\xi l_{R_{\bar{x}}}^{\rho_k}(\xi^{k+1}, y^k)\right\| \le L \left\|\operatorname{grad}_x l^{\rho_k}\left(x^{k+1}, y^k\right)\right\| \le Lc \left\|y^{k+1} - y^k\right\|. \tag{5.19}$$

Therefore, we further have

$$\begin{aligned} s^2 d^2(x^{k+1}, \bar{x}) &\le L^2 c^2 \left\|y^{k+1} - y^k\right\|^2 + \left\|y^k - \bar{y}\right\|^2 \\ &\le L^2 c^2 \Big(\frac{\left\|y^{k+1} - \bar{y}\right\|}{\left\|y^k - \bar{y}\right\|} + 1\Big) \left\|y^k - \bar{y}\right\|^2 + \left\|y^k - \bar{y}\right\|^2, \end{aligned}$$

while the Q-linear convergence of $\{y^k\}$ implies there exists $0 < r < 1$, such that $\limsup_k \dfrac{\left\|y^{k+1} - \bar{y}\right\|}{\left\|y^k - \bar{y}\right\|} = r$. Thus, x^k converge to $\bar{x}$ R-linearly. □

Remark 5.7. In [43], a local primal-dual Q-linear convergence rate for RALM is obtained under the M-SRCQ and M-SOSC, which require the multipliers to be unique. Although our work under the manifold strong variational sufficient condition can only achieve the local primal R-linear convergence rate, it relaxes the constraint of being a singleton for the multiplier set, which provides theoretical guarantees for more real-world cases.

Remark 5.8. Assume that the manifold strong variational sufficient condition holds at $(\bar{x}, \bar{y})$ under $R_{\bar{x}}$. It follows from [36, Theorem 3.1] that the approximate error for $\bar{\xi}^{k+1}$ in (5.8) can be replaced by

$$\sqrt{\tilde{\rho}^k} \left\|\nabla_\xi l_{R_{\bar{x}}}^{\rho_k}\left(\xi^{k+1}, y^k\right)\right\| \le \begin{cases} \text{(a)} & \varepsilon_k' \\ \text{(b)} & \varepsilon_k' \min\left\{1, \left\|\tilde{\rho}_k \nabla_y l_{R_{\bar{x}}}^{\rho_k}\left(\xi^{k+1}, y^k\right)\right\|\right\} \\ \text{(c)} & \varepsilon_k' \min\left\{1, \left\|\tilde{\rho}_k \nabla_y l_{R_{\bar{x}}}^{\rho_k}\left(\xi^{k+1}, y^k\right)\right\|^2\right\}, \end{cases}$$

where $\varepsilon_k' = \varepsilon_k\sqrt{s}$ as s is the modulus of the strong convexity of $\nabla_\xi l_{R_{\bar{x}}}^{\rho_k}\left(\cdot, y^k\right)$. Moreover, from (5.19) we can replace (5.7) to

$$\sqrt{\tilde{\rho}^k} \left\|\operatorname{grad}_x l^{\rho_k}\left(x^{k+1}, y^k\right)\right\| \le \begin{cases} \text{(a)} & \varepsilon_k'' \\ \text{(b)} & \varepsilon_k'' \min\left\{1, \left\|\tilde{\rho}_k \nabla_y l^{\rho_k}\left(x^{k+1}, y^k\right)\right\|\right\} \\ \text{(c)} & \varepsilon_k'' \min\left\{1, \left\|\tilde{\rho}_k \nabla_y l^{\rho_k}\left(x^{k+1}, y^k\right)\right\|^2\right\}. \end{cases} \tag{5.20}$$

where $\varepsilon_k'' = \varepsilon_k\sqrt{s}/L$ as L is the local Lipschitz modulus given in (5.19). The stopping criteria in Theorem 5.2 can be replaced by (5.20).

Define the KKT residual mapping as $R(x, y) := \|\operatorname{grad}_x L(x, y)\| + \|g(x) - \operatorname{prox}_\theta\left(g(x) + y\right)\|$. Under stopping criterion (5.20b) we are able to obtain the following proposition, which implies that the KKT residual of (1.1) also converges R-linearly.

Proposition 5.9. *Suppose that the manifold strong variational sufficiency holds at $(\bar{x}, \bar{y})$ at level $\bar{\rho}$ and the approximation error is chosen as (5.20b). If $\sqrt{\tilde{\rho}^k}\varepsilon_k < 1$ for sufficiently large k, then there exists $L_g > 0$ such that*

$$R\left(x^{k+1}, y^{k+1}\right) \le w^k \operatorname{dist}\left(Y^k, Z\right),$$

where $w^k = \left(\varepsilon_k''\sqrt{\tilde{\rho}_k} + (1 + \bar{\rho} + L_g\bar{\rho})\left(\tilde{\rho}_k\right)^{-1}\right)\left(1 - \sqrt{\tilde{\rho}_k}\varepsilon_k\right)^{-1}$.

The proof of Proposition 5.9 is similar with the proof of [41, Proposition 4.7] with the replacement of projection with proximal mapping in our case, and we omit the details here.

6. Semismooth Newton method for RALM subproblem. After assuming obtaining the linear convergence results of the RALM, one remaining issue is how to solve the subproblem of updating x in (1.4) efficiently. In [42] the authors propose a globalized semismooth Newton method on Riemannian manifold, which could be well-suited for our problem. To begin, we provide the definition in [11] of generalized covariant derivative for vector field of manifold.

DEFINITION 6.1. *Let X be a locally Lipschitz vector field on $\mathcal{M}$. The B-derivative is a set-valued map $\partial_B X : \mathcal{M} \rightrightarrows \mathcal{L}(T\mathcal{M})$ with*

$$\partial_B X(x) := \left\{ H \in \mathcal{L}\left(T_x\mathcal{M}\right) : \exists \left\{x^k\right\} \subset \mathcal{D}_X, \quad \lim_{k\to+\infty} x^k = x, \quad H = \lim_{k\to+\infty} \nabla X\left(x^k\right) \right\},$$

where the last limit means that $\left\|\nabla X\left(x^k\right)\left[P_{xx^k} v\right] - P_{xx^k} H v\right\| \to 0$ for all $v \in T_x\mathcal{M}$. The Clarke generalized covariant derivative 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 $\partial_B X(x)$.

Algorithm 6.1 Globalized semismooth Newton method for solving (1.4) at (x^j, y^j, ρ_j)

Input: Choose $x^0 \in \mathcal{M}, \bar{\nu} \in (0,1]$ and let $\{\eta_k\} \subset \mathbb{R}_+$be a sequence converging to 0. Set $\mu \in (0,1/2), \delta \in (0,1), m_{\max} \in \mathbb{N}$, and $p, \beta_0, \beta_1 > 0$, and set $k := 0$.

1: Let $X^j(x^k) := \mathrm{grad}_x\, l^{\rho_j}\left(x^k, y^j\right)$. Choose $G^k \in \partial X^j(x^k)$ and find $V^k \in T_{x^k}\mathcal{M}$ such that

$$\left\| \left(G^k + \omega_k I\right) V^k + X^j\left(x^k\right) \right\| \leq \tilde{\eta}_k,$$

where $\omega_k := \left\|X^j\left(x^k\right)\right\|^{\bar{\nu}}, \tilde{\eta}_k := \min\left\{\eta_k, \left\|X^j\left(x^k\right)\right\|^{1+\bar{\nu}}\right\}$.

2: If V^k is not a sufficient descent direction of φ, i.e. it does not satisfy

$$\left\langle -X^j\left(x^k\right), V^k \right\rangle \geq \min\left\{\beta_0, \beta_1 \left\|V^k\right\|^p\right\} \left\|V^k\right\|^2,$$

then, we set V^k to be $-X^j\left(x^k\right)$. Next, find the minimum non-negative integer m_k such that

$$l^{\rho_j}\left(R_{x^k}\left(\delta^{m_k} V^k\right), y^j\right) \leq l^{\rho_j}\left(x^k, y^j\right) + \mu\delta^{m_k}\left\langle X\left(x^k\right), V^k\right\rangle.$$

3: Set $x^{k+1} = R_{x^k}\left(\delta^{m_k} V^k\right)$.

4: Set $k = k + 1$ and go to step 2.

If a retraction R additionally satisfies $\frac{\mathrm{D}^2}{\mathrm{d}t^2} R(t\xi)\Big|_{t=0} = 0$ for any $\xi \in T_x\mathcal{M}$, where $\dfrac{\mathrm{D}^2}{\mathrm{d}t^2}\gamma$ denotes acceleration of the curve γ, then we call it a second-order retraction. The second-order retraction can be interpreted as the approximation of exponential mappings and includes retractions such as the polar retraction for Stiefel manifold and the projective retraction for fixed rank manifold, see [3, Exmaple 4.12] for more details. In this section, we always assume that the retraction is second-order.

The globalized semismooth Newton method for solving RALM subproblem (1.4) is given in Algorithm 6.1. Let $R_{\bar{x}}$ be a second-order retraction. Suppose that θ is a polyhedral convex function or is the indicator function of second-order cone or positive semidefinite cone, it follows Theorem

4.7 that the M-SSOSC is equivalent to the positive definiteness of the elements in Hessian bundle of $l^{\rho}_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y})$ for ρ sufficiently large. Moreover, we shall show that this condition also is sufficient to the superlinear convergence of Algorithm 6.1 by employing following proposition.

PROPOSITION 6.2. *Let $(\bar{x}, \bar{y})$ satisfy the condition* (3.5)*. Assume that the manifold $\mathcal{M}$ is an embedded submanifold, and retraction $R_{\bar{x}}$ is second-order, then we have*

$$\partial \nabla_{\xi} l^{\rho^j}_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y}) = \partial X^j(\bar{x}),$$

where $X^j(\bar{x})$ is given in Algorithm 6.1.

Proof. Now for any $G \in \partial X^j(\bar{x})$, we claim that $G \in \partial \nabla_{\xi} l^{\rho^j}_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y})$, or equivalently,

$$\lim_{\xi_k \to 0_{\bar{x}}} \left\| \nabla^2_{\xi\xi} l^{\rho^j}_{R_{\bar{x}}}\left(\xi^k\right)[v] - Gv \right\| = 0.$$

Since $l^{\rho^j}_{R_{\bar{x}}}(\xi) = l^{\rho^j} \circ R_{\bar{x}}(\xi)$, for any $v \in T_{\bar{x}}\mathcal{M}$ we have

$$\begin{aligned}
&\left\langle \nabla^2_{\xi\xi} l^{\rho^j}_{R_{\bar{x}}}\left(\xi^k\right)[v], v\right\rangle \\
&= \frac{d^2}{dt^2}(l^{\rho^j} \circ R_{\bar{x}})(\xi + tv)\Big|_{t=0} = \frac{d}{dt}\Big(\frac{d}{dt} l^{\rho^j}\left(R_{\bar{x}}(\xi + tv)\right)\Big)\Big|_{t=0} \\
&= \frac{d}{dt} Dl^{\rho^j}\left(R_{\bar{x}}(\xi + tv)\right)\Big[\frac{d}{dt} R_{\bar{x}}(\xi + tv)\Big]\Big|_{t=0} \qquad (6.1)\\
&= \Big\langle \frac{D}{dt} \operatorname{grad} l^{\rho^j}\big(R_{\bar{x}}(\xi + tv)\big), DR_{\bar{x}}(\xi)[v]\Big\rangle + \Big\langle \operatorname{grad} l^{\rho^j}\big(R_{\bar{x}}(\xi + tv)\big), \frac{D^2}{dt^2} R_{\bar{x}}(\xi + tv)\Big\rangle\Big|_{t=0} \\
&= \Big\langle \nabla \operatorname{grad} l^{\rho^j}(R_{\bar{x}}(\xi))\big[DR_{\bar{x}}(\xi)v\big], DR_{\bar{x}}(\xi)v\Big\rangle + \Big\langle \operatorname{grad} l^{\rho^j}\big(R_{\bar{x}}(\xi + tv)\big), \frac{D^2}{dt^2} R_{\bar{x}}(\xi + tv)\Big\rangle\Big|_{t=0}.
\end{aligned}$$

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 in (6.1) can be written as

$$\begin{aligned}
&\Big\langle \nabla \operatorname{grad} l^{\rho^j}(R_{\bar{x}}(\xi))\left[D\exp_{\bar{x}}(\xi)v\right], D\exp_{\bar{x}}(\xi)v\Big\rangle + O(\|\xi\|) \\
&= \Big\langle \nabla \operatorname{grad} l^{\rho^j}(R_{\bar{x}}(\xi))\left[P_{\bar{x}R_{\bar{x}}(\xi)}v\right], P_{\bar{x}R_{\bar{x}}(\xi)}v\Big\rangle + O(\|\xi\|).
\end{aligned} \qquad (6.2)$$

Moreover, the right term of the last equation in (6.1) equals to zero if $\xi_k \to 0_{\bar{x}}$ and $t \to 0$ as $R_{\bar{x}}$ is a second-order retraction. Thus, for any $v \in T_{\bar{x}}\mathcal{M}$,

$$\begin{aligned}
\left\langle \nabla^2_{\xi\xi} l^{\rho^j}_{R_{\bar{x}}}\left(\xi^k\right)[v] - Gv, v\right\rangle = &\left\langle P_{\bar{x}R_{\bar{x}}(\xi^k)} \nabla^2_{\xi\xi} l^{\rho^j}_{R_{\bar{x}}}\left(\xi^k\right)[v] - \operatorname{Hess} l^{\rho^j}\left(R_{\bar{x}}(\xi^k)\right)\left[P_{\bar{x}R_{\bar{x}}(\xi^k)}v\right], P_{\bar{x}R_{\bar{x}}(\xi^k)}v\right\rangle \\
&+ \left\langle \operatorname{Hess} l^{\rho^j}\left(R_{\bar{x}}(\xi^k)\right)\left[P_{\bar{x}R_{\bar{x}}(\xi^k)}v\right] - P_{\bar{x}R_{\bar{x}}(\xi^k)}Gv, P_{\bar{x}R_{\bar{x}}(\xi^k)}v\right\rangle.
\end{aligned}$$

The first equality is obatined since the parallel transport is isometry. By (6.1), (6.2) and $G \in \partial X^j(\bar{x})$, the above equation converges to 0 when $\xi^k \to 0_{\bar{x}}$. The arbitrary taken v implies that

$$\lim_{\xi_k \to 0_{\bar{x}}} \left\| \nabla^2_{\xi\xi} l^{\rho^j}_{R_{\bar{x}}}\left(\xi^k\right)[v] - Gv \right\| = 0.$$

Similarly, $\partial \nabla_{\xi} l^{\rho^j}_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y}) \subseteq \partial X^j(\bar{x})$. This implies $\partial \nabla_{\xi} l^{\rho^j}_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y}) = \partial X^j(\bar{x})$. ∎

THEOREM 6.3. *Let $(\bar{x}, \bar{y})$ satisfy the condition (3.5). The manifold strong variational sufficient condition at $(\bar{x}, \bar{y})$ if and only if all elements in $\partial X^j(\bar{x})$ are positive definiteness.*

Proof. By Theorem 4.7, the M-SSOSC guarantee the positive definiteness of the elements in $\partial \nabla_{\xi} l^{\rho^j}_{R_{\bar{x}}}(0_{\bar{x}}, \bar{y})$. Follows from Proposition 6.2 we have the conculsion. ∎

Now by Theorem 6.3 and [42, Theorem 4.3] we obtain the following convergence result of Algorithm 6.1.

PROPOSITION 6.4. *Suppose the manifold strong variational sufficient condition holds at a stationary point $(\bar{x}, \bar{y})$. Let $\{x^k\}$ be the sequence generated by Algorithm 6.1 and the retraction $R_{\bar{x}}$ to be second-order. Suppose there exists $\delta > 0$ such that $\Omega := \{x \in \mathcal{M} : \varphi(x) \leq \varphi(x^0) + \delta\}$ is compact. Denote $\hat{x}$ be any accumulation point of $\{x^k\}$. If X^j is semismooth at $\hat{x}$ with order ν with respect to ∂X^j, then we have $x^k \to \hat{x}$ as $k \to \infty$ and $\hat{x}$ is optimal for subproblem (1.4). Moreover, for sufficiently large k, it holds*

$$d\left(x^{k+1}, \hat{x}\right) \leq O\left(d\left(x^k, \hat{x}\right)^{1+\min\{\nu, \bar{\nu}\}}\right),$$

where $\bar{\nu} \in (0, 1]$ is the parameter defined in Algorithm 6.1.

Remark 6.5. Based on Theorem 6.3, it can be concluded that 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 stationary point. This relationship underscores the pivotal role played by manifold strong variational sufficiency in ensuring the efficiency of the semi-smooth Newton method for solving the RALM subproblem.

7. Applications and numerical experiments. Based on Proposition 5.9, we are able to verify the convergence rate by using the KKT residual $R(x, y)$. In this section, we will present numerical experiments on fixed rank manifold and Stiefel manifold, and illustrate the local linear rate when the M-SSOSC is satisfied. All codes are implemented in Matlab (R2021b) and all the numerical experiments are run under a 64-bit MacOS on an Intel Cores i5 2.4GHz CPU with 16GB of memory.

7.1. Robust matrix completion. We are now considering the robust matrix completion (RMC) problem proposed in [7]. For a given $A \in \mathbb{R}^{m \times n}$, let $g(X) = P_\Omega(X - A)$ and $\theta(\cdot) = \mu\|\cdot\|_1$. Here, P_Ω is the projector defined by $(P_\Omega(X))_{ij} = X_{ij}$ if $(i, j) \in \Omega$ and 0 otherwise. By setting $\mathcal{M} = Fr(m, n, r) := \{X \in \mathbb{R}^{m \times n} : \operatorname{rank}(X) = r\}$, we obtain the following robust matrix completion problem

$$\begin{array}{ll} \min_{X \in \mathbb{R}^{m \times n}} & \|P_\Omega(X - A)\|_1 \\ \text{s.t.} & X \in Fr(m, n, r). \end{array} \tag{7.1}$$

In comparison with the matrix completion using the Frobenius norm as an objective function, the l_1-norm is expected to due with an inexact data A with some extreme outliers.

It is known in [40] that the tangent space of $\mathcal{M} = F(m, n, r)$ at a point $X = USV^\top$ is

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

and the normal space is

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

The projection to tangent space 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) can be written as $L(X,y) = \langle P_\Omega(X-A), y\rangle$. Then the KKT condition takes the form of

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

By [20, Section 3], the Hessian of a function f on $Fr(m,n,r)$ at $X = U\Sigma V$ can be written as

$$\operatorname{Hess} f(X)[\xi] = U\hat{M}V^\top + \hat{U}_p V^\top + U\hat{V}_p^\top \quad \forall \xi \in T_X Fr(m,n,r),$$

where

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

in which $M(Z;X) := U^\top Z X$, $U_p(Z;X) = P_U^\perp Z V$ and $V_p(Z;X) = P_V^\perp Z^\top U$. While $\nabla_X L(\bar{X},\bar{y}) = P_\Omega(\bar{y})$ and $\nabla^2_{XX} L(\bar{X},\bar{y}) = 0$, it holds

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

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

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

Therefore, the M-SSOSC for the RMC problem (7.1) at the KKT point $(\bar{X},\bar{y})$ holds if for any $\xi \in T_{\bar{X}}\mathcal{M}$ satisfying $P_\Omega(\xi) \in \operatorname{aff}\mathcal{C}_{\theta,g}(\bar{X},\bar{y})\backslash\{0\}$, $\operatorname{tr}\left(\xi^\top P_\Omega(\bar{y})\xi^\top U\Sigma^{-1}V^\top\right) > 0$ holds.

We consider a basic example of problem (7.1), where Ω is the full index set. Let $U = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 \\ 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 \end{bmatrix}^T$, $V = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.6 & -0.8 & 0 & 0 \\ 0 & 0.8 & 0.6 & 0 & 0 \end{bmatrix}^T$ and $S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$. The observed matrix is set to $A = A_{\text{ex}} + E_{\text{out}}$, where $A_{\text{ex}} = USV^T$ is the assumed ground truth and E_{out} is a matrix with random entries added only in the lower right 2×2 submatrix. Since A_{ex} is of rank $r = 3$, $\bar{X} = A_{\text{ex}}$ is a solution of this problem. Consider if $(\bar{X}, y)$ satisfies (7.2), then y can be chosen as $y_{ij} = \operatorname{sgn}(E^{ij}_{\text{out}})$. In this case, $\mathcal{C}_{\theta,g}(\bar{X},y) = \{d \in \mathbb{R}^{m\times n} \mid d_{ij} \in \mathbb{R} \text{ if } E^{ij}_{\text{out}} \neq 0, d_{ij} = 0 \text{ if } E^{ij}_{\text{out}} = 0\}$, and $\operatorname{aff}\mathcal{C}_{\theta,g}(\bar{X},y) = \mathcal{C}_{\theta,g}(\bar{X},y)$. The nonzero position of E_{out} implies that only $0_{\bar{X}}$ can satisfy $P_\Omega(\xi) \in \operatorname{aff}\mathcal{C}_{\theta,g}(\bar{X},y)$. Therefore, the M-SSOSC holds at $(\bar{X},y)$. Now we can apply the inexact RALM to this problem and obtain Figure 1, which is the variation of KKT residue.

7.2. Compressed modes. In this section, we will consider the compressed modes (CM) problem. Let H be a discretization of the Hamilton operator, then the CM problem is formulated as follow ([29]):

$$\begin{array}{ll} \min_{X\in\mathbb{R}^{n\times r}} & \operatorname{tr}(X^\top H X) + \mu\|X\|_1 \\ \text{s.t.} & X \in \operatorname{St}(n,r). \end{array} \tag{7.3}$$

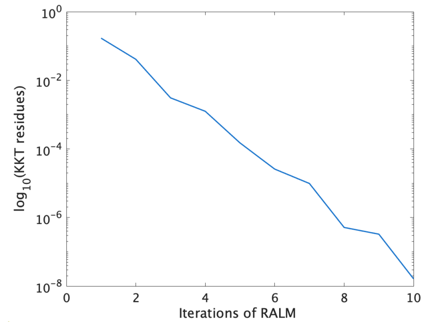


Fig. 1: KKT residues of the RMC problem generated by inexact RALM

By setting $f(X) = \mathrm{tr}(X^\top HX)$, $\theta = \mu\|\cdot\|_1$, $g(X) = X$ and $\mathcal{M} = \mathrm{St}(n,r) = \{X|X^\top X = I_r\}$, this is of the form of (1.1). The tangent space of $\mathcal{M} = \mathrm{St}(n,r)$ at X is

$$T_X\mathcal{M} = \left\{\xi \in \mathbb{R}^{n\times r} : X^\top\xi + \xi^\top X = 0\right\}, \tag{7.4}$$

and the the projection to tangent space is given by $\Pi_X Y = Y - X\,\mathrm{sym}(X^\top Y)$. The Lagrangian of (7.3) can be written as $L(X,y) = \mathrm{tr}(X^\top HX) + \langle X, y\rangle$. Then the KKT condition is

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

The Euclidean gradient of $L(X,y)$ is $\nabla_X L(X,y) = 2HX + y$, and for any $\xi \in T_X\mathcal{M}$, the Euclidean Hessian of $L(X,y)$ is $\nabla^2_{XX}L(X,y)[\xi] = 2H\xi$. Therefore, by [20] the Riemannian gradient and Hessian of $L(X,y)$ can be computed as

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

Moreover, for any $\xi \in T_w\mathcal{M}$, we have

$$\begin{aligned} \langle \xi, \mathrm{Hess}\, L(w,y)\xi\rangle &= \left\langle \xi, 2H\xi - \xi\,\mathrm{sym}(X^\top\nabla_X L(X,y)\right\rangle \\ &= \langle \xi, 2H\xi\rangle - \left\langle \xi, 2\xi X^\top HX\right\rangle - \left\langle \xi, \xi\,\mathrm{sym}(X^\top y)\right\rangle. \end{aligned} \tag{7.6}$$

By the KKT condition (7.5), we have $2HX + y \in N_X\mathcal{M}$. Equivalently, there exists a symmetric matrix S, such that $2HX + y = XS$. Combining with (7.6) we can obtain

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

We further obtain the critical cone of θ and g as $\mathcal{C}_{\theta,g}(X,y) = \{d \in \mathbb{R}^{n\times r}|\theta'(X,d) = \langle d, y\rangle\}$, where

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

The affine hull of $\mathcal{C}_{\theta,g}(X,y)$ is then given by

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

Therefore, the M-SSOSC for CM problem at the KKT point $(\bar{X},\bar{y})=(\bar{X},-2H\bar{X}+\bar{X}\bar{S})$ holds if for any $\xi\in T_{\bar{X}}\mathcal{M}$ satisfying $\xi\in\operatorname{aff}\mathcal{C}_{\theta,g}(\bar{X},\bar{y})\backslash\{0\}$, $\operatorname{tr}(\xi^\top H\xi)-\operatorname{tr}(\xi^\top\xi\bar{S})>0$ holds.

In [42], the authors consider setting the CM problem to solve Schrödinger equation of 1D free-electron model with periodic boundary condition:

$$-\frac{1}{2}\Delta\phi(x)=\lambda\phi(x),\quad x\in[0,50] \tag{7.7}$$

and numerically, they find that the smallest eigenvalue of $H_k\in\partial\operatorname{grad}L^{\rho_k}\left(X_k,y^k\right)$ is always larger than zero, which implies that the M-SSOSC may be satisfied in this case. Here we use a simple example to illustrate this conjecture.

Consider the Schrödinger equation of with boundary condition when $x\in[0,2]$. Discretize the domain$[0,2]$ into $n=4$ nodes and let H be the discretized version of $-\frac{1}{2}\Delta$. Then $H=-\begin{bmatrix}-4&2&0&2\\2&-4&2&0\\0&2&-4&2\\2&0&2&-4\end{bmatrix}$. For $r=2$, it can be verified that $\bar{X}=\begin{bmatrix}0&0&\sqrt{2}/2&\sqrt{2}/2\\\sqrt{2}/2&\sqrt{2}/2&0&0\end{bmatrix}^\top$ is a stationary point of (7.3), and one of the corresponding multiplier $\bar{y}$ is given by $\bar{y}=\mu\begin{bmatrix}0&0&1&1\\1&1&0&0\end{bmatrix}^\top$. Moreover, if rewrite $\bar{y}=\bar{X}\bar{S}-2H\bar{X}$, then $\bar{S}=\begin{bmatrix}-4+\sqrt{2}\mu&4\\4&-4+\sqrt{2}\mu\end{bmatrix}$. The affine hull of $\mathcal{C}_{\theta,g}(\bar{X},\bar{y})$ is now written as $\operatorname{aff}\mathcal{C}_{\theta,g}(\bar{X},\bar{y})=\left\{\begin{bmatrix}0&0&\xi_1&\xi_2\\\xi_3&\xi_4&0&0\end{bmatrix}^\top\Big|\xi_i\in\mathbb{R},i=1,2,3,4\right\}$. For any $\xi\in\operatorname{aff}\mathcal{C}_{\theta,g}(\bar{X},\bar{y})$ satisfying $\xi\in T_{\bar{X}}\mathcal{M}$, by (7.4) we further obtain $\xi=\begin{bmatrix}0&0&\xi_1&-\xi_1\\\xi_2&-\xi_2&0&0\end{bmatrix}^\top$. Therefore, for any $\xi\in T_{\bar{X}}\mathcal{M}$ satisfying $\xi\in\operatorname{aff}\mathcal{C}_{\theta,g}(\bar{X},\bar{y})\backslash\{0\}$, we have

$$\operatorname{tr}(\xi^\top H\xi)-\operatorname{tr}(\xi^\top\xi\bar{S})=12(\xi_1^2+\xi_2^2)-2(-4+\sqrt{2}\mu)(\xi_1^2+\xi_2^2)=(20-2\sqrt{2}\mu)(\xi_1^2+\xi_2^2).$$

As long as $\mu<5\sqrt{2}$, the M-SSOSC is satisfied for this problem. From Corollary 6.3, the general Jacobian used in the semisooth newton is positve definite. Now setting $\mu=0.8$ and apply RALM, we obtain Figure 2, which shows the linear rate of KKT residue.

For larger n and r, it seems difficult to prove the positive definiteness of the general Hessian used in Algorithm 6.1. However, The minimal eigenvalue is observed to be positive in [42, Table 5]. We also consider back to problem (7.7) with shifting n nodes and similarly observe that the smallest eigenvalues of the general Hessians used in the semismooth Newton methods are positive.

Table 1 indicates that the manifold strong variational sufficiency might be satisfied at the obtained points. Therefore, we consider applying the RALM and all settings follow [42]. The results of variations in KKT residues are reported in Figure 3[1].

[1]The code of using semismooth Newton based RALM to solve compressed modes problems is provided in the published paper [42]

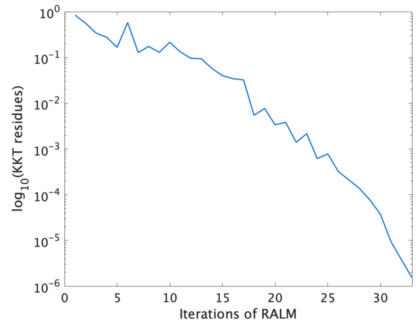


Fig. 2: KKT residues of CM generated by inexact RALM

Table 1: The minimum eigenvalue of the general Hessian which is used in Algorithm 6.1 for (7.3).

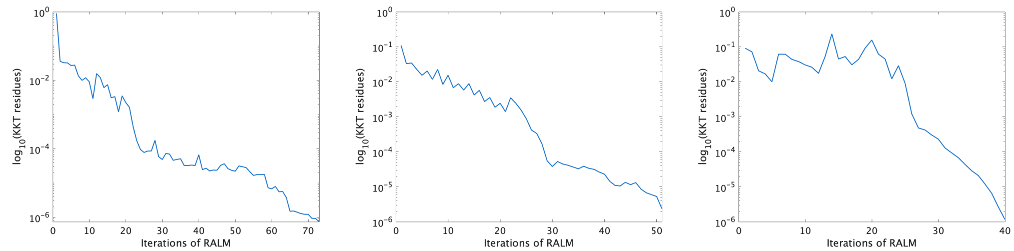
n	r	mimimum eigenvalue				
200	20	0.3049	0.3238	0.1926	0.2811	0.1447
500	20	0.2755	0.1089	0.1173	0.0982	0.1087
1000	20	0.0266	0.1658	0.0026	0.1684	0.2055

8. Conclusion. This paper investigates the local convergence of RALM without assuming the uniqueness of multiplier. We devise a local equivalent problem on tangent space and introduce the manifold variational sufficient condition. It is shown that manifold strong variational sufficient condition is equivalent to the M-SSOSC under certain circumstances. Under this condition, a local augmented dual problem is formulated, consequently establishing the linear convergence rate of RALM. Furthermore, we prove that general Hessians used in the semismooth Newton method for solving the RALM subproblem are positive definite when the manifold strong variational sufficiency holds. The numerical experiments on various applications demonstrate the linear convergence rate.

However, there remain several unresolved issues in Riemannian nonsmooth optimization. For instance, verifying the retractional convexity for manifold functions presents a significant challenge and is waiting for future study. Additionally, while it is understood that the primal proximal point algorithm is equivalent to the dual ALM in Euclidean settings, the relationship between these two algorithms remains unknown under the retactional convexity in a Riemannian setting.

REFERENCES

[1] P.-A. Absil and S. Hosseini, *A collection of nonsmooth Riemannian optimization problems*, Nonsmooth Optimization and Its Applications, (2019), pp. 1–15.
[2] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, 2009.
[3] P.-A. Absil and J. Malick, *Projection-like retractions on matrix manifolds*, SIAM Journal on Optimization, 22 (2012), pp. 135–158.
[4] D. Azagra, J. Ferrera, and F. López-Mesas, *Nonsmooth analysis and Hamilton–Jacobi equations on Rie-*



(a) n=200, r=20

(b) n=500, r=20

(c) n=1000, r=20

Fig. 3: The KKT residues of compressed modes problems generated by inexact RALM

mannian manifolds, Journal of Functional Analysis, 220 (2005), pp. 304–361.

[5] D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods*, Academic press, 2014.

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

[7] L. Cambier and P.-A. Absil, *Robust low-rank matrix completion by Riemannian optimization*, SIAM Journal on Scientific Computing, 38 (2016), pp. S440–S460.

[8] 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 (2021), pp. 4759–4773.

[9] 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 (2020), pp. 210–239.

[10] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Trust region methods*, SIAM, 2000.

[11] F. R. de 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 (2020), pp. 522–539.

[12] L. Demanet and P. Hand, *Scaling law for recovering the sparsest element in a subspace*, Information and Inference: A Journal of the IMA, 3 (2014), pp. 295–309.

[13] K. K. Deng and Z. Peng, *An inexact augmented Lagrangian method for nonsmooth optimization on Riemannian manifold*, arXiv preprint arXiv:1911.09900, (2019).

[14] O. P. Ferreira and P. R. Oliveira, *Subgradient algorithm on Riemannian manifolds*, Journal of Optimization Theory and Applications, 97 (1998), pp. 93–104.

[15] O. P. Ferreira and P. R. Oliveira, *Proximal point algorithm on Riemannian manifolds*, Optimization, 51 (2002), pp. 257–270.

[16] P. Grohs and S. Hosseini, *ε-subgradient algorithms for locally Lipschitz functions on Riemannian manifolds*, Advances in Computational Mathematics, 42 (2016), pp. 333–360.

[17] M. R. Hestenes, *Multiplier and gradient methods*, Journal of optimization theory and applications, 4 (1969), pp. 303–320.

[18] S. Hosseini and M. R. Pouryayevali, *Generalized gradients and characterization of epi-Lipschitz sets in Riemannian manifolds*, Nonlinear Analysis: Theory, Methods & Applications, 74 (2011), pp. 3884–3895.

[19] S. Hosseini and A. Uschmajew, *A Riemannian gradient sampling algorithm for nonsmooth optimization on manifolds*, SIAM Journal on Optimization, 27 (2017), pp. 173–189.

[20] 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 (2020), pp. 199–248.

[21] 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 (2022), p. e2409.

[22] W. Huang and K. Wei, *Riemannian proximal gradient methods*, Mathematical Programming, 194 (2022), pp. 371–413.

[23] C. Kanzow and D. Steck, *Improved local convergence results for augmented Lagrangian methods in $\mathcal{C}^2$-cone reducible constrained optimization*, Mathematical Programming, 177 (2019), pp. 425–438.

[24] A. Kovnatsky, K. Glashoff, and M. M. Bronstein, *MADMM: a generic algorithm for non-smooth opti-*

mization on manifolds, in European Conference on Computer Vision, Springer, 2016, pp. 680–696.

[25] R. J. Lai and S. Osher, *A splitting method for orthogonality constrained problems*, Journal of Scientific Computing, 58 (2014), pp. 431–449.

[26] J. M. Lee, *Introduction to Smooth manifolds*, Springer, 2003.

[27] 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 (2008), pp. 557–575.

[28] J. Nocedal and S. J. Wright, *Numerical optimization*, Springer, 1999.

[29] 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 (2013), pp. 18368–18373.

[30] M. J. Powell, *A method for nonlinear constraints in minimization problems*, Optimization, (1969), pp. 283–298.

[31] R. T. Rockafellar, *Convex analysis*, vol. 18, Princeton university press, 1970.

[32] R. T. Rockafellar, *A dual approach to solving nonlinear programming problems by unconstrained optimization*, Mathematical programming, 5 (1973), pp. 354–373.

[33] R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Mathematics of operations research, 1 (1976), pp. 97–116.

[34] R. T. Rockafellar, *Advances in convergence and scope of the proximal point algorithm*, J. Nonlinear and Convex Analysis, 22 (2021), pp. 2347–2375.

[35] R. T. Rockafellar, *Augmented Lagrangians and hidden convexity in sufficient conditions for local optimality*, Mathematical Programming, (2022), pp. 1–36.

[36] R. T. Rockafellar, *Convergence of augmented Lagrangian methods in extensions beyond nonlinear programming*, Mathematical Programming, (2022), pp. 1–46.

[37] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, vol. 317, Springer Science & Business Media, 2009.

[38] D. A. Spielman, H. Wang, and J. Wright, *Exact recovery of sparsely-used dictionaries*, in Conference on Learning Theory, JMLR Workshop and Conference Proceedings, 2012, pp. 37–1.

[39] 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 (2008), pp. 349–391.

[40] B. Vandereycken, *Low-rank matrix completion by Riemannian optimization*, SIAM Journal on Optimization, 23 (2013), pp. 1214–1236.

[41] S. W. Wang, C. Ding, Y. J. Zhang, and X. Zhao, *Strong Variational Sufficiency for Nonlinear Semidefinite Programming and its Implications*, arXiv preprint arXiv:2210.04448, (2022).

[42] Y. Zhou, C. Bao, C. Ding, and J. Zhu, *A semismooth Newton based augmented Lagrangian method for nonsmooth optimization on matrix manifolds*, Mathematical Programming, 201 (2023), pp. 1–61.

[43] Y. X. Zhou, C. L. Bao, and C. Ding, *On the robust isolated calmness of a class of nonsmooth optimizations on riemannian manifolds and its applications*, arXiv preprint arXiv:2208.07518, (2022).

[44] H. Zou, T. Hastie, and R. Tibshirani, *Sparse principal component analysis*, Journal of computational and graphical statistics, 15 (2006), pp. 265–286.