On the robust isolated calmness of a class of nonsmooth optimizations on Riemannian manifolds and its applications

Yuexin Zhou · Chenglong Bao · Chao Ding

This version: August 16, 2022

Abstract This paper studies the robust isolated calmness property of the KKT solution mapping of a class of nonsmooth optimization problems on Riemannian manifold. The manifold version of the Robinson constraint qualification, the strict Robinson constraint qualification, and the second order conditions are defined and discussed. We show that the robust isolated calmness of the KKT solution mapping is equivalent to the M-SRCQ and M-SOSC conditions. Furthermore, under the above two conditions, we show that the Riemannian augmented Lagrangian method has a local linear convergence rate. Finally, we verify the proposed conditions and demonstrate the convergence rate on two minimization problems over the sphere and the manifold of fixed rank matrices.

1 Introduction

In recent years, manifold optimization has become an important class of constrained optimization problems and has various applications in many tasks [32,38,44,48]. See [2,8,22] for the comprehensive study of manifold optimization. In this paper, we consider the following nonsmooth

Y.X. Zhou
Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, P.R. China, School of Mathematical Sciences, University of Chinese Academy of Science, Beijing, P.R. China.
E-mail: zhouyuexin19@mails.ucas.ac.cn

C.L. Bao
Yau Mathematical Sciences Center, Tsinghua University, P.R. China and Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, P.R. China.
E-mail: clbao@mail.tsinghua.edu.cn

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

manifold optimization problem:

$$
\begin{aligned}
\min \quad & f(x) + \theta(g_1(x)) \\
\text{s.t.} \quad & g_2(x) \in \mathcal{Q}, \\
& x \in \mathcal{M},
\end{aligned}
\tag{1.1}
$$

where $\mathcal{M}$ is a smooth Riemannian manifold, $f : \mathcal{M} \to \mathbb{R}$, $g_1 : \mathcal{M} \to \mathbb{Y}$ and $g_2 : \mathcal{M} \to \mathbb{Z}$ are twice continuously differentiable functions, $\mathbb{Y}$ and $\mathbb{Z}$ are two Euclidean spaces each equipped with a scalar product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$, $\theta : \mathbb{Y} \to \mathbb{R}$ is a proper closed convex function, and $\mathcal{Q} \subset \mathbb{Z}$ is a nonempty closed convex set. Many applications arising from emerging fields can be cast into the form (1.1), e.g., compressed modes [36], sparse principal component analysis [48], constrained sparse principle analysis [34] and robust matrix completion [9]. See [1] for more details.

Many algorithms for solving the nonsmooth optimization problems have been extended from Euclidean space to Riemannian manifolds, such as the subgradient methods [17,19], proximal gradient methods [11,23,24], alternating direction methods of multipliers (ADMM) [29,30], proximal point methods [10,18] and augmented Lagrangian methods (ALM) [26,47]. However, the theoretical results for the nonsmooth manifold optimization problem seem much less than those under Euclidean settings. To the best of our knowledge, perturbation analysis for the problem (1.1), which is closely related to the convergence analysis of numerical algorithms, has not been established yet.

One of the essential perturbation properties for optimization problems is the robust isolated calmness of the Karush-Kuhn-Tucker (KKT) solution mapping under perturbations [15, Definition 2]. The isolated calmness is a Lipschitz-like property of set-value mappings, which is firstly proposed by Robinson [39] and generalized in [16] as follows: a mapping $F : \mathbb{R}^m \to \mathbb{R}^n$ is said to have the isolated calmness property if there exists a constant $\kappa \geq 0$ and neighborhoods $\mathcal{V}$ of $\bar{y}$ and $\mathcal{U}$ of $\bar{x}$ such that $\| y - \bar{y} \| \leq \kappa \| x - \bar{x} \|$ when $y \in F(x) \cap \mathcal{V}$ and $x \in \mathcal{U}$. The robust version is defined in [15] with additional requirement: $F(x) \cap \mathcal{V} \neq \emptyset$ for all $x \in \mathcal{U}$. This property is crucial for establishing the linear convergence rate of numerical algorithms. For example, it is used to analyze the convergence of sequential quadratic programming (SQP) in [5] for nonlinear programs (NLPs), where the property is obtained under the strict Mangasarian-Fromovitz constraint qualification (SMFCQ) and the second order sufficient condition (SOSC). The recent work [15] successfully characterizes the robust isolated calmness for optimization problems with $\mathcal{C}^2$-cone reducible constraints (cf. Definition 3) in Euclidean space. Specifically, under the Robinson constraint qualification (RCQ), it is proved that the robust isolated calmness of the KKT mapping holds at a local solution if and only if the strict Robinson constraint qualification (SRCQ) and the SOSC are satisfied at the given point. The main challenge for establishing the perturbation analysis for manifold optimization is to construct a proper perturbed problem since the canonical perturbation $\langle a, \cdot \rangle$ is no longer a linear function on manifolds. In this work, we first construct the perturbed manifold problem through a locally equivalent Euclidean problem using the normal coordinate chart around a KKT point and define the SRCQ and SOSC conditions on manifolds. Furthermore, we show that the robust isolated calmness of the KKT solution mapping of (1.1) is equivalent to the manifold SRCQ (M-SRCQ) condition and the manifold SOSC (M-SOSC) condition (the definitions are given in Section 3) hold at solution points, which is the manifold extension of [15].

An important application of the robust isolated calmness of the KKT solution mapping is the local convergence analysis of the augmented Lagrangian method (ALM). The classical ALM is proposed by Hestenes [20] and Powell [37] for equality constraints and is extended to nonlinear programming by Rockafellar [40]. The convergence analysis of ALM under the Euclidean settings has been extensively studied for decades. The classical result of the local linear convergence rate of ALM for NLPs often requires the linear independence constraint qualification (LICQ) and the

second order sufficient condition, e.g. [4,13,45]. For conic programs such as the nonlinear second order cone programs and semidefinite programs, Liu and Zhang [33] and Sun et al. [43] obtain the local linear convergence rate of ALM under the constraint non-degeneracy condition and the strong second order sufficient condition at KKT points, respectively. Recently, for the general nonlinear optimizations involving $\mathcal{C}^2$-cone reducible constraints, Kanzow and Steck [27] obtain the primal-dual linear convergence result under the assumption of the robust isolated calmness of the KKT solution mapping, which is equivalent to the SRCQ and SOSC conditions by [15].

The Riemannian ALM for nonsmooth constrained manifold optimization is recently proposed in [47] (see also [26]). In [47], the iteration sequence of the Riemannian ALM has been proved to converge to a KKT point under some suitable constraint qualifications, and numerical experiments have shown a better performance than other existing methods. Motivated by the promising numerical results in [47], one natural question is whether the local convergence rate of the Riemannian ALM for the nonsmooth manifold optimization (1.1) can be similarly obtained as in [27] under the robust isolated calmness for the KKT solution, or equivalently, the assumption of the M-SRCQ and M-SOSC conditions. First, we consider a simple example:

$$\begin{aligned} \min \quad & x_2^2 + |x_1 - x_2| \\ \text{s.t.} \quad & 2x_1 + x_2 \geq 0, \\ & x_1^2 + x_2^2 = 1. \end{aligned} \qquad (1.2)$$

It can be verified that the unique optimal solution of (1.2) is $x^* = (\sqrt{2}/2, \sqrt{2}/2)^T$ with the corresponding multipliers $y^* = \sqrt{2}/2$ and $z^* = 0$. By applying the Riemannian ALM (see Algorithm 1 for details) and noting that the corresponding ALM subproblems can be solved exactly, it is observed from Figure 1 that the distance between the iteration (x^k, y^k, z^k) and the solution (x^*, y^*, z^*) converges linearly when k is sufficiently large. Moreover, Figure 1 indicates that the linear convergence rate decreases if the penalty parameter ρ_k increases. On the other hand, it
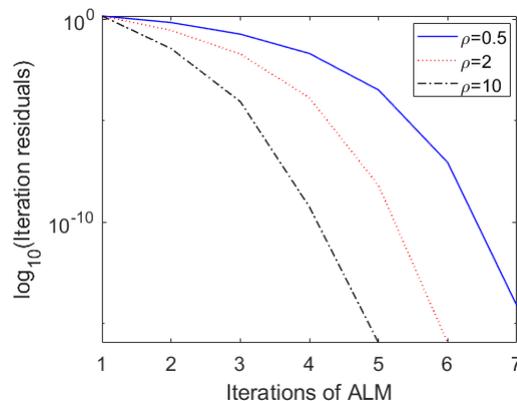

Fig. 1: The Riemannian ALM for solving (1.2) with different penalty parameters ρ

will be verified in Section 3 (Remark 9) that for problem (1.2), the M-SRCQ holds at x^* with respect to (y^*, z^*) and the M-SOSC holds at x^*. Inspired by [27], in this paper, we show that the Riemannian ALM indeed has a local linear convergence rate under the M-SRCQ and M-SOSC assumptions. To the best of our knowledge, this is the first time obtaining the local convergence rate of ALM on manifold. Moreover, we use two examples of nonsmooth optimizations over sphere and fixed rank manifold to verify the M-SRCQ and M-SOSC conditions for different nonsmooth

problems, and to illustrate the obtained theoretical results in Section 5. Moreover, we will show that the M-SRCQ condition is satisfied at all KKT points for a class of nonsmooth optimization problems on sphere.

When the manifold $\mathcal{M}$ in problem (1.1) is embedded in a Euclidean space $\mathbb{X}$ (cf. [8, Definition 3.10]), it can be locally written as the equality constraint $\{x \mid h(x) = 0\}$. In applications, many manifold constraints for optimization are embedded, for example, the Euclidean space itself, the Stiefel manifold $\{X \in \mathbb{R}^{n \times p} \mid X^\top X = I_p\}$ and the fixed rank manifold $\{X \in \mathbb{R}^{m \times n} \mid \mathrm{rank}(X) = r\}$. In this case, the perturbation properties of problem (1.1) can be studied through the classical approaches. If $h'(x)$ is onto, we will show in Section 3 (Remarks 5 and 8) that the Euclidean SRCQ and the Euclidean SOSC conditions for problem (1.1) are equivalent to the M-SRCQ and M-SOSC conditions. However, if $h'(x)$ is not of full row rank, we find that the Euclidean SRCQ condition cannot be satisfied even when the problem has no other constraints. A typical case is the optimization problems involving the Stiefel manifold, which include many important applications (e.g., [34, 36, 48]). Moreover, when $\mathcal{M}$ is no longer an embedded submanifold, applying the traditional ALM to problem (1.1) seems challenging. Based on these two observations, in this paper, we study the perturbation properties of problem (1.1) in manifold sense and the local convergence rate of Riemannian augmented Lagrangian method directly.

The rest of the paper is organized as follows. In Section 2 we review some background of smooth manifolds and set-valued mappings. In Section 3, we define the constraint qualifications and second order optimality conditions on manifold, and characterize the robust isolated calmness of the KKT solution mapping for the perturbed problem defined in this section. The Riemannian augmented Lagrangian method and its local convergence analysis are given in Section 4. The applications and numerical results are contained in Section 5. Finally, we make some remarks in Section 6.

2 Preliminaries and notations

We first introduce some basic concepts of manifolds and the set value mappings that will be used in our discussion. The manifold properties mentioned below can be found in the books [2, 28, 31]. The content about set-valued mapping and directional derivative can be found in [7, 16].

Let $\mathcal{M}$ be an n-dimensional smooth manifold and $x \in \mathcal{M}$. $\mathfrak{F}_x(\mathcal{M})$ is defined as the set of all smooth real-valued functions on a neighborhood of x. The mapping ξ_x from $\mathfrak{F}_x(\mathcal{M})$ to $\mathbb{R}$ such that there exists a curve γ on $\mathcal{M}$ with $\gamma(0) = x$ satisfying $\xi_x f := \dot{\gamma}(0)f := \left. \dfrac{d(f(\gamma(t)))}{dt} \right|_{t=0}$ for all $f \in \mathfrak{F}_x(\mathcal{M})$ is called a tangent vector, and the tangent space $T_x\mathcal{M}$ is the set of all tangent vectors to $\mathcal{M}$ at x. If $\mathcal{M}$ is embedded in Euclidean space $\mathbb{X}$, the normal space $N_x\mathcal{M}$ is defined as the orthogonal complement of $T_x\mathcal{M}$ in $\mathbb{X}$. If given a chart $(\mathcal{U}, \varphi)$ at x, a basis of $T_x\mathcal{M}$ is given by $(\dot{\gamma}_1(0), \dot{\gamma}_2(0), \cdots, \dot{\gamma}_n(0))$, where $\gamma_i(t) := \varphi^{-1}(\varphi(x) + te_i)$, with e_i denoting the i th canonical vector of $\mathbb{R}^n$. The tangent bundle is defined as $T\mathcal{M} := \bigcup_x T_x\mathcal{M}$, which is the set of all tangent vectors to $\mathcal{M}$. A map $V : \mathcal{M} \to T\mathcal{M}$ is called a vector field on $\mathcal{M}$ if $V(x) \in T_x\mathcal{M}$ for all $x \in \mathcal{M}$.

Let $F : \mathcal{M} \to \mathbb{X}$ be a smooth mapping. 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. The canonical identification $T_{F(x)}\mathbb{X} \simeq \mathbb{X}$ yields $DF(x)\xi_x = \sum_i (\xi_x F^i)e_i$, where $F(x) = \sum_i F^i(x)e_i$ is the decomposition of $F(x)$ in a basis $(e_i)_{i=1,2,\cdots,n}$ of $\mathbb{X}$. If $\xi = \sum_i \xi_i \dot{\gamma}_i(0) \in T_x\mathcal{M}$, we have $D\varphi(x)\xi = (\xi_1, \xi_2, \cdots, \xi_n)^\top \in \mathbb{R}^n$, which means that $D\varphi(x)\xi$ is the coordinate representation of $\xi \in T_x\mathcal{M}$. It is known that if $\mathcal{M}$ is an embedded submanifold

of a Euclidean space, then $DF(x)$ is reduced to the classical definition of directional derivative, i.e., $DF(x)\xi_x = \lim_{t \to 0} \dfrac{F(x + t\xi_x) - F(x)}{t}$.

To distinguish it with Riemannian differential, we use $h'(x)\xi$ to represent the traditional directional derivative for a Euclidean function h in the direction ξ and $\nabla h(x)$ to be the Euclidean gradient of h. We also use these notations for convenience: $\hat{x} := \varphi(x)$, $\hat{\xi} := D\varphi(x)\xi$ and $\hat{f} := f \circ \varphi^{-1}$. Then for $\xi_x \in T_x\mathcal{M}$ and $f \in \mathfrak{F}_x(\mathcal{M})$, it holds that $\xi_x f = \left\langle \hat{\xi}_x, \nabla \hat{f}(\hat{x}) \right\rangle_{\mathbb{R}^n}$.

Riemannian metric $\langle \cdot, \cdot \rangle_x$ is a smoothly varying inner product with respect to x defined for tangent vectors. A differentiable manifold whose tangent spaces are endowed with Riemannian metrics is called a Riemannian manifold. Let $(E_1, E_2, \cdots, E_n)$ denotes the basis of the vector field, and $g_{ij}(x) := \langle E_i(x), E_j(x) \rangle_x$. Then $g_{ij}(\cdot)$ is a smooth function on $\mathcal{M}$. For vector fields $\xi = \sum_i \xi_i E_i$ and $\zeta = \sum_i \zeta_i E_i$, we have $\langle \xi(x), \zeta(x) \rangle_x = \sum_{i,j} g_{ij}(x)\xi^i(x)\zeta^j(x)$. Let $G_{\varphi(x)}$ denotes the matrix-valued function such that the (i, j)-th element of $G_{\varphi(x)}$ is $g_{ij}(x)$, we will then obtain

$$\langle \xi(x), \zeta(x) \rangle_x = \hat{\xi}(x)^\top G_{\varphi(x)} \hat{\zeta}(x). \tag{2.1}$$

When no confusions arise, we will use $\langle \cdot, \cdot \rangle$ instead of $\langle \cdot, \cdot \rangle_x$ for simplicity. The induced norm of this inner product is denoted by $\| \cdot \|$ with the subscript being omitted.

Given $f \in \mathfrak{F}_x(\mathcal{M})$, the gradient of f at x, denoted by $\operatorname{grad} f(x)$, is defined as the unique tangent vector that satisfying

$$\langle \operatorname{grad} f(x), \xi \rangle := \xi_x f = \langle \nabla \hat{f}(\hat{x}), \hat{\xi} \rangle_{\mathbb{R}^n} \quad \forall \xi \in T_x\mathcal{M}. \tag{2.2}$$

The coordinate expression of $\operatorname{grad} f(x)$ is hence given by

$$D\varphi(x) \operatorname{grad} f(x) = G_{\varphi(x)}^{-1} \nabla \hat{f}(\hat{x}). \tag{2.3}$$

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 that satisfies $\operatorname{Hess} f(x)\xi = \nabla_\xi \operatorname{grad} f(x)$ for all $\xi \in T_x\mathcal{M}$, where ∇ is the Riemannian connection on $\mathcal{M}$. If $\mathcal{M}$ is embedded in Euclidean space $\mathbb{X}$ [2, Section 3.3], the Riemannian gradient of $f \in \mathfrak{F}_x(\mathcal{M})$ at a point x is equal to the projection of the gradient of f onto $T_x\mathcal{M}$, which means that $\operatorname{grad} f(x) = \Pi_x(\nabla f(x))$. The Riemannian Hessian of f at x in the direction ξ can be computed by $\operatorname{Hess} f(x)\xi = \Pi_x((\operatorname{grad} f)'(x)\xi)$.

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

$$d : \mathcal{M} \times \mathcal{M} \to \mathbb{R} : d(y, z) := \inf_\Gamma L(\gamma), \tag{2.4}$$

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 $\delta > 0$. With the distance function defined above, the Lipschitz property 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

$$|f(y) - f(z)| \le Ld(y, z) \quad \forall y, z \in \mathcal{U}.$$

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 funtion f at $x \in \mathcal{M}$ in the direction $v \in T_x\mathcal{M}$, is defined in [21] as

$$f^\circ(x; v) := \limsup_{y \to x, t \downarrow 0} \frac{f \circ \varphi^{-1}(\varphi(y) + tD\varphi(x)v) - f \circ \varphi^{-1}(\varphi(y))}{t}, \tag{2.5}$$

where $(\mathcal{U}, \varphi)$ is a chart containing x. The Clarke subdifferential of a locally Lipschitz function f at $x \in \mathcal{M}$, denoted by $\partial f(x)$, is defined as

$$\partial f(x) = \{\xi \in T_x \mathcal{M} \mid \langle \xi, v \rangle \le f^\circ(x; v) \text{ for all } v \in T_x \mathcal{M}\}. \tag{2.6}$$

The following proposition is the chain rule for composite function $\theta \circ g$, where θ is locally Lipschitz and g is continuously differentiable. This extends the result in [12, Theorem 2.3.10].

Proposition 1 *Suppose that $g : \mathcal{M} \to \mathbb{Y}$ is continuously differentiable at x and $\theta : \mathbb{Y} \to \mathbb{R}$ is locally Lipschitz near $g(x)$. Then $h = \theta \circ g$ is locally Lipschitz near x, and one has*

$$\partial h(x) \subset Dg(x)^* \partial \theta(g(x)). \tag{2.7}$$

Moreover, the equality holds if θ is regular (cf. [12, Definition 2.3.4]) at $g(x)$.

Proof The locally Lipschitz of h near x is a direct result of the Lipschitz of θ and the continuously differentiable of g. Suppose that $(\mathcal{U}, \varphi)$ is a chart around x. From [46, Proposition 3.1], we know that

$$\partial h(x) = (D\varphi(x))^{-1} G_{\varphi(x)}^{-1} \partial \hat{h}(\hat{x}). \tag{2.8}$$

For any $y \in \mathbb{Y}$ and $\xi \in T_x \mathcal{M}$, we have

$$\begin{aligned}
\langle (D\varphi(x))^{-1} G_{\varphi(x)}^{-1} \hat{g}'(\hat{x})^* y, \xi \rangle &= \left(D\varphi(x)(D\varphi(x))^{-1} G_{\varphi(x)}^{-1} \hat{g}'(\hat{x})^* y \right)^\top G_{\varphi(x)} \left(D\varphi(x)\xi \right) \\
&= (\hat{g}'(\hat{x})^* y)^\top D\varphi(x)\xi \qquad \text{(by taking } \hat{\xi} = D\varphi(x)\xi) \\
&= \langle \hat{g}'(\hat{x})^* y, \hat{\xi} \rangle = \langle \hat{g}'(\hat{x})\hat{\xi}, y \rangle = \langle Dg(x)\xi, y \rangle = \langle Dg(x)^* y, \xi \rangle,
\end{aligned}$$

which implies that

$$Dg(x)^* y = (D\varphi(x))^{-1} G_{\varphi(x)}^{-1} \hat{g}'(\hat{x})^* y. \tag{2.9}$$

Since for any $x \in \mathcal{M}$, $\partial \theta(g(x)) = \partial \theta(\hat{g}(\hat{x}))$, using [12, Theorem 2.3.10] we get

$$\partial \hat{h}(\hat{x}) \subset \hat{g}'(\hat{x})^* \partial \theta(\hat{g}(\hat{x})), \tag{2.10}$$

and the equality holds when θ is regular. Combining (2.8), (2.9) and (2.10), we obtain the conclusion. $\qquad \square$

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 $\mathrm{Exp}_x : T_x \mathcal{M} \to \mathcal{M}$, $\xi \mapsto \mathrm{Exp}_x \xi = \gamma(1; x, \xi)$ is called the exponential mapping on $x \in \mathcal{M}$. Let $E : \mathbb{R}^n \to T_x \mathcal{M}$ be a linear bijection such that $\{E(e_1), E(e_2), \cdots, E(e_n)\}$ is an orthogonal basis for $T_x \mathcal{M}$, and let $\mathcal{V}$ be a neighborhood of the origin of $T_x \mathcal{M}$ and $\mathcal{U}$ a neighborhood of x such that Exp_x is a diffeomorphism between $\mathcal{V}$ and $\mathcal{U}$. If we define $\varphi = E^{-1} \mathrm{Exp}_x^{-1}$, then $(\mathcal{U}, \varphi)$ is known as the Riemannian normal coordinate chart around x. It is known that $G_{\varphi(x)} = I_n$ under the normal coordinate chart. Therefore, for any $f \in \mathfrak{F}_x(\mathcal{M})$ and $\xi \in T_x \mathcal{M}$, it holds that

$$\langle \xi, \mathrm{Hess}\, f(x)\xi \rangle = \langle \hat{\xi}, \nabla_{xx}^2 \hat{f}(\hat{x})\hat{\xi} \rangle \tag{2.11}$$

under the normal coordinate chart.

In this paper, we are concerned about the continuity and Lipschitz properties of set-valued mappings. For the set-valued mapping from a Euclidean space $\mathbb{E}$ to a manifold $\mathcal{M}$, we define the continuity and (robust) isolated calmness as follows.

Definition 1 The set-valued mapping $\Psi : \mathbb{E} \to \mathcal{M}$ is said to be lower semi-continuous at $\bar{p}$ for $\bar{x}$ if for any open neighborhood $\mathcal{V}$ of $\bar{x}$ there exists an open neighborhood $\mathcal{U}$ of $\bar{p}$ such that

$$\emptyset \neq \Psi(p) \cap \mathcal{V} \quad \forall\, p \in \mathcal{U}.$$

The mapping Ψ is said to be upper semi-continuous at $\bar{p}$ if for any open set $\mathcal{O} \supset \Psi(\bar{p})$ there exists an open neighborhood $\mathcal{U}$ such that for any $p \in \mathcal{U}$, $\Psi(p) \subset \mathcal{O}$. Furthermore, if Ψ is lower semi-continuous at $\bar{p}$ for $\bar{x}$ and is upper semi-continuous at $\bar{p}$, then Ψ is said to be continuous at $(\bar{p}, \bar{x}) \in \operatorname{gph} \Psi$.

Definition 2 The set-valued mapping $\Psi : \mathbb{E} \to \mathcal{M}$ is said to be isolated calm at $\bar{p}$ for $\bar{x}$ if there exist a constant $\kappa > 0$ and open neighborhoods $\mathcal{U}$ of $\bar{p}$ and $\mathcal{V}$ of $\bar{x}$ such that

$$d(x, \bar{x}) \leq k\|p - \bar{p}\| \quad \forall\, x \in \Psi(p) \cap \mathcal{V} \quad \text{and} \quad p \in \mathcal{U}. \tag{2.12}$$

Moreover, Ψ is said to be robustly isolated calm at $\bar{p}$ for $\bar{x}$ if (2.12) holds and for each $p \in \mathcal{U}$, $\Psi(p) \cap \mathcal{V} \neq \emptyset$.

For a given cone $\mathcal{C}$, the largest linear space contained in $\mathcal{C}$ is called the lineality space of $\mathcal{C}$. A cone $\mathcal{C}$ is said to be pointed if and only if its lineality space contains only zero. The following $\mathcal{C}^2$-cone reducibility of a closed convex set is taken from [7, Definition 3.135].

Definition 3 A closed convex set $\mathcal{K}$ is said to be $\mathcal{C}^2$-cone reducible at $y \in \mathcal{K}$, if there exist an open neighborhood $\mathcal{U} \subset \mathbb{Y}$ of y, a pointed closed convex cone $\mathcal{C}$ in Euclidean space $\mathbb{Z}$, and a twice continuously differentiable mapping $\Xi : \mathcal{U} \to \mathbb{Z}$ such that: (i) $\Xi(y) = 0 \in \mathbb{Z}$, (ii) the derivative mapping $\Xi'(y) : \mathbb{Y} \to \mathbb{Z}$ is onto, and (iii) $\mathcal{K} \cap \mathcal{U} = \{y \in \mathcal{U} \mid \Xi(y) \in \mathcal{C}\}$. We say that $\mathcal{K}$ is $\mathcal{C}^2$-cone reducible if $\mathcal{K}$ is $\mathcal{C}^2$-cone reducible at every $y \in \mathcal{K}$.

It is worth noting that the class of $\mathcal{C}^2$-cone reducible sets is rich, which notably includes all the polyhedral convex sets and many nonpolyhedral sets such as the second-order cone and positive semidefinite matrices cone [7,42]. Furthermore, a proper closed convex function $\theta : \mathbb{Y} \to \mathbb{R}$ is said to be $\mathcal{C}^2$-cone reducible at $y \in \operatorname{dom} \theta$ if its epigraph $\operatorname{epi} \theta$ is $\mathcal{C}^2$-cone reducible at $(y, \theta(y))$, and is said to be $\mathcal{C}^2$-cone reducible if it is reducible at every $y \in \operatorname{dom} \theta$.

Let $\mathcal{D}$ be a closed set in $\mathbb{Y}$. The radial cone at a point y is defined as

$$\mathcal{R}_{\mathcal{D}}(y) := \left\{ d \in \mathbb{Y} \mid \exists\, t^* > 0 \text{ such that } y + td \in \mathcal{D} \text{ for any } t \in [0, t^*] \right\}, \tag{2.13}$$

and the tangent cone is defined as

$$\mathcal{T}_{\mathcal{D}}(y) = \left\{ d \in \mathbb{Y} \mid \exists\, t^k \downarrow 0, \ \operatorname{dist}\left(y + t^k d, \mathcal{D} \right) = o\left(t^k \right) \right\},$$

where for any $u \in \mathbb{Y}$, $\operatorname{dist}(u, \mathcal{D}) := \inf\{\|u - p\| \mid p \in \mathcal{D}\}$ is the Euclidean distance function. The normal cone to $\mathcal{D}$ at y is defined by $\mathcal{N}_{\mathcal{D}}(y) = (\mathcal{T}_{\mathcal{D}}(y))^{\circ}$. The inner and outer second order tangent sets to the given closed set $\mathcal{D}$ in the direction $h \in \mathbb{Y}$ are defined respectively by

$$\mathcal{T}_{\mathcal{D}}^{i,2}(y, h) := \left\{ w \in \mathbb{Y} \mid \operatorname{dist}\left(y + th + \frac{1}{2} t^2 w, \mathcal{D} \right) = o\left(t^2 \right), t \geq 0 \right\},$$

and

$$\mathcal{T}_{\mathcal{D}}^{2}(y, h) := \left\{ w \in \mathbb{Y} \mid \exists\, t_k \downarrow 0 \text{ such that } \operatorname{dist}\left(y + t_k h + \frac{1}{2} t_k^2 w, \mathcal{D} \right) = o\left(t_k^2 \right) \right\}.$$

It follows from [7, Proposition 3.136] that if $\mathcal{D}$ is a $\mathcal{C}^2$-cone reducible convex set, then $\mathcal{T}_{\mathcal{D}}^{i,2}(y, h) = \mathcal{T}_{\mathcal{D}}^{2}(y, h)$. In this case, $\mathcal{T}_{\mathcal{D}}^{2}(y, h)$ will be simply called the second order tangent set to $\mathcal{D}$ at $y \in \mathcal{D}$ in the direction $h \in \mathbb{Y}$.

For a given function $\theta : \mathbb{Y} \to (-\infty, +\infty]$, the lower and upper directional epiderivatives (cf. e.g., [7, (2.68) and (2.69)]) of θ at $y \in \operatorname{dom} \theta$ in the direction $h \in \mathbb{Y}$ are defined as

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

respectively, where Σ is the set of all positive sequences $\{t_n\}$ converging to zero. Since θ is a proper closed convex function, it is clear that $\theta_-^\downarrow(y; \cdot) = \theta_+^\downarrow(y; \cdot)$. Moreover, it is well-known [7, Propositioin 2.58] that for any $y \in \operatorname{dom} \theta$,

$$\mathcal{T}_{\operatorname{epi} \theta}(y, \theta(y)) = \operatorname{epi} \theta_-^\downarrow(y; \cdot). \tag{2.14}$$

If $\theta_-^\downarrow(y; h)$ is finite for $x \in \operatorname{dom} \theta$ and $h \in \mathbb{Y}$, the lower second order epiderivatives [7, (2.76)] for $w \in \mathbb{Y}$ is defined as:

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

Finally, it follows from [7, Proposition 3.41] that

$$\mathcal{T}_{\operatorname{epi} \theta}^2 \left((y, \theta(y)); h, \theta_-^\downarrow(y; h) \right) = \operatorname{epi} \theta_-^{\downarrow\downarrow}(y; h, \cdot). \tag{2.15}$$

3 The robust isolated calmness of KKT solution mapping

In this section, we shall study the robust isolated calmness of KKT solution mapping for problem (1.1). The Robinson constraint qualification, the strict Robinson constraint qualification and the second order optimality conditions will be introduced to Riemannian manifolds in the following subsection. After defining the perturbed problem using the normal coordinate chart, we will characterize the robust isolated calmness property for (1.1).

3.1 The constraint qualifications and second order optimality conditions

Consider the generalized form of optimization problem on manifold as follows:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) \in \mathcal{K}, \\ & x \in \mathcal{M}, \end{aligned} \tag{3.1}$$

where $f : \mathcal{M} \to \mathbb{R}$ and $g : \mathcal{M} \to \mathbb{Y}$ are twice continuously differentiable, and $\mathcal{K} \subset \mathbb{Y}$ is a nonempty closed convex set in $\mathbb{Y}$. When $g(x) = (g_1(x), g_2(x))$ and $\mathcal{K} = \operatorname{epi} \theta \times \mathcal{Q}$, problem (3.1) is equivalent to the nonsmooth problem (1.1). In addition, we also assume $\mathcal{K}$ is C^2-cone reducible (Definition 3). Using the normal coordinate chart, we can locally transform (3.1) into the traditional minimizing problem in Euclidean space. For $x \in \mathcal{F} := \{x \in \mathcal{M} \mid g(x) \in \mathcal{K}\}$, suppose $(\mathcal{U}, \varphi)$ is the normal coordinate chart around x. Then, we locally obtain an equivalent problem in $\mathbb{R}^n$

$$\begin{aligned} \min \quad & \hat{f}(\hat{x}) \\ \text{s.t.} \quad & \hat{g}(\hat{x}) \in \mathcal{K}, \\ & \hat{x} \in \varphi(\mathcal{U}) \in \mathbb{R}^n, \end{aligned} \tag{3.2}$$

where $\hat{f} = f \circ \varphi^{-1}$, $\hat{g} = g \circ \varphi^{-1}$ and $\hat{x} = \varphi(x) = (\varphi_1(x), \varphi_2(x), \cdots, \varphi_n(x))$. Suppose that $\hat{x}$ is a feasible solution to problem (3.2). The critical cone $\widehat{C}(\hat{x})$ of (3.2) at $\hat{x}$ is defined by

$$\widehat{C}(\hat{x}) := \left\{ \hat{d} \in \mathbb{R}^n \mid \hat{g}'(\hat{x})\hat{d} \in \mathcal{T}_{\mathcal{K}}(\hat{g}(\hat{x})), \hat{f}'(\hat{x})\hat{d} \le 0 \right\}. \tag{3.3}$$

For the corresponding $x \in \mathcal{M}$ for problem (3.1), we can define the critical cone $\mathcal{C}(x)$ of (3.1) at x as

$$\mathcal{C}(x) := \{ d \in T_x\mathcal{M} \mid Dg(x)d \in \mathcal{T}_{\mathcal{K}}(g(x)), Df(x)d \le 0 \}. \tag{3.4}$$

It is easy to see that $\widehat{C}(\hat{x}) = D\varphi(x)\mathcal{C}(x)$.

Let $L : \mathcal{M} \times \mathbb{Y} \to \mathbb{R}$ be the Lagrangian function of problem (3.1) defined by

$$L(x; y) := f(x) + \langle y, g(x) \rangle, \quad (x, y) \in \mathcal{M} \times \mathbb{Y}. \tag{3.5}$$

and $\widehat{L} : \mathbb{R}^n \times \mathbb{Y} \to \mathbb{R}$ be the Lagrangian function of problem (3.2) defined by

$$\widehat{L}(\hat{x}; y) := \hat{f}(\hat{x}) + \langle y, \hat{g}(\hat{x}) \rangle, \quad (\hat{x}, y) \in \mathbb{R}^n \times \mathbb{Y}. \tag{3.6}$$

We say that $x \in \mathcal{M}$ is a stationary point of (3.1) and $y \in \mathbb{Y}$ is a Lagrange multiplier at x, if (x, y) satisfies the Karush-Kuhn-Tucker (KKT) condition:

$$\begin{cases} \text{grad}_x L(x; y) = 0, \\ y \in \mathcal{N}_{\mathcal{K}}(g(x)), \end{cases} \tag{3.7}$$

where $\mathcal{N}_{\mathcal{K}}(g(x))$ is the normal cone to $\mathcal{K}$ at $g(x) \in \mathbb{Y}$. We denote by $M(x)$ the set of all Lagrange multipliers at x. We also use $\widehat{M}(\hat{x})$ to denote the set of multipliers at a stationary point $\hat{x}$ for problem (3.2).

It is well known (cf. e.g., [7]) that the Robinson constraint qualification (RCQ) for problem (3.2) holds at a feasible point $\hat{x}$ if

$$\hat{g}'(\hat{x})\mathbb{R}^n + \mathcal{T}_{\mathcal{K}}(\hat{g}(\hat{x})) = \mathbb{Y}, \tag{3.8}$$

the strict Robinson constraint qualification (SRCQ) for problem (3.2) holds with respect to $y \in \widehat{M}(\hat{x})$ if for the stationary point $\hat{x}$ if

$$\hat{g}'(\hat{x})\mathbb{R}^n + \mathcal{T}_{\mathcal{K}}(\hat{g}(\hat{x})) \cap y^{\perp} = \mathbb{Y}, \tag{3.9}$$

and the constraint non-degeneracy is said to hold at $\hat{x}$ if

$$\hat{g}'(\hat{x})\mathbb{R}^n + \text{lin}\left(\mathcal{T}_{\mathcal{K}}(\hat{g}(\hat{x}))\right) = \mathbb{Y},$$

where $\text{lin}\left(\mathcal{T}_{\mathcal{K}}(g(\hat{x}))\right)$ is the lineality space of $\mathcal{T}_{\mathcal{K}}(\hat{g}(\hat{x}))$. Therefore, we are able to define the Robinson constraint qualification, the strict Robinson constraint qualification, and the constraint non-degeneracy on Riemannian manifolds. These definitions can lead us to the existence or uniqueness of Lagrange multipliers for problem (3.1).

Definition 4 For the manifold optimization problem (3.1), we say that the manifold Robinson constraint qualification (M-RCQ) holds at $x \in \mathcal{F}$ if

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

the manifold strict Robinson constraint qualification (M-SRCQ) holds at a stationary point x with respect to $y \in M(x)$ if

$$Dg(x)T_x\mathcal{M} + \mathcal{T}_{\mathcal{K}}(g(x)) \cap y^{\perp} = \mathbb{Y}, \tag{3.11}$$

and the manifold constraint non-degeneracy holds at $x \in \mathcal{F}$ if

$$Dg(x)T_x\mathcal{M} + \text{lin}\left(\mathcal{T}_{\mathcal{K}}(g(x))\right) = \mathbb{Y}. \tag{3.12}$$

Remark 1 For Euclidean NLPs, the RCQ and constraint non-degeneracy will reduce to the MFCQ and LICQ conditions [7, page 71 and Example 4.78], respectively. Our definitions of M-RCQ and manifold constraint non-degeneracy can also build the equivalence with the manifold MFCQ and the manifold LICQ proposed in [3, 46] if $\mathcal{K}$ is polyhedral.

Using the normal coordinate chart, we are able to establish the following theorem on the existence and boundness of multipliers of (3.1).

Theorem 1 *Suppose that x^* is a locally optimal solution of (3.1). Then, $M(x^*)$ is a nonempty, convex, bounded, and compact subset of $\mathbb{Y}$ if and only if the M-RCQ (3.10) holds at x^*.*

Proof Recall that $\widehat{M}(\hat{x})$ is the set of all Lagrange multipliers at $\hat{x}$ for problem (3.2). Then, for any $y \in \widehat{M}(\hat{x})$, y satisfies the KKT condition

$$\begin{cases} \nabla_{\hat{x}} \widehat{L}(\hat{x}; y) = 0, \\ y \in \mathcal{N}_{\mathcal{K}}(\hat{g}(\hat{x})). \end{cases} \tag{3.13}$$

It then follows from (2.3) that $\mathrm{grad}_x L(x, y) = (D\varphi(x))^{-1} G_{\varphi(x)}^{-1} \nabla_{\hat{x}} \widehat{L}(\hat{x}, y) = 0$, which implies that $y \in M(x)$. Thus, $\widehat{M}(\hat{x}) \subseteq M(x)$. The inverse relation $M(x) \subseteq \widehat{M}(\hat{x})$ can be obtained similarly. Hence, we have $M(x) = \widehat{M}(\hat{x})$.

Suppose that x^* is a locally optimal solution of (3.1), then $\hat{x}^*$ is a local solution of the equivalent problem (3.2). Thus, for any $d \in T_{x^*}\mathcal{M}$, by taking $\hat{d} = D\varphi(x^*)d$, we obtain that $Dg(x^*)d = \hat{g}'(\hat{x}^*)\hat{d}$. Conversely, for any given $\hat{d} \in \mathbb{R}^n$, $d = (D\varphi(x^*))^{-1}\hat{d}$ is a tangent vector at x^*, which ensures that $Dg(x^*)d = \hat{g}'(\hat{x}^*)\hat{d}$. Therefore, the M-RCQ (3.10) holds at x^* if and only if RCQ (3.8) holds at $\hat{x}^*$. The conclusion is then obtained by the well-known fact [49] that $\widehat{M}(\hat{x}^*)$ is nonempty, convex, bounded and compact if and only if RCQ (3.8) holds at $\hat{x}^*$. □

Remark 2 By Theorem 1, the M-RCQ (3.10) is a necessary condition for a given stationary point x^* to be the local optimal solution of (3.1). This extends the first order necessary condition proposed in [3, Theorem 4.1] for $\mathcal{K}$ being polyhedral to more general manifold constrained optimization.

The following statement in Theorem 2 can also be obtained by using the normal coordinate chart, which indicates that M-SRCQ (3.11) holds at x^* with respect to y^* if and only if SRCQ (3.9) holds at $\hat{x}^*$ for y^*. However, we give a direct proof here, which follows the proof of [7, Proposition 4.47].

Proposition 2 (Uniqueness of Lagrangian multipliers) *Suppose that x^* is a feasible solution to (3.1) and $M(x^*) \neq \emptyset$. Let $y^* \in M(x^*)$. Then, the corresponding multiplier set $M(x^*)$ is a singleton if and only if*

$$[Dg(x^*)T_{x^*}\mathcal{M}]^{\perp} \cap \mathcal{R}_{\mathcal{N}_{\mathcal{K}}(g(x^*))}(y^*) = \{0\}. \tag{3.14}$$

Proof For any $y \in M(x^*)$, let $\lambda = y - y^*$. We have $Dg(x^*)^*\lambda = 0$ and $y^* + \lambda \in \mathcal{N}_{\mathcal{K}}(g(x^*))$. Since $y^* \in \mathcal{N}_{\mathcal{K}}(g(x^*))$, by the definition of radial cone (2.13), we obtain that $\lambda \in \mathcal{R}_{\mathcal{N}_{\mathcal{K}}(g(x^*))}(y^*)$. $Dg(x^*)^*\lambda = 0$ implies that for any $\xi \in T_{x^*}\mathcal{M}$, $\langle Dg(x^*)\xi, \lambda \rangle = 0$. Thus, $\lambda \in [Dg(x^*)T_{x^*}\mathcal{M}]^{\perp}$. Moreover, by taking $\tilde{\lambda} \in [Dg(x^*)T_{x^*}\mathcal{M}]^{\perp}$, for any $\xi \in T_{x^*}\mathcal{M}$, we have $0 = \langle Dg(x^*)\xi, \tilde{\lambda} \rangle = \langle \xi, Dg(x^*)^*\tilde{\lambda} \rangle$. Hence, $Dg(x^*)^*\tilde{\lambda} = 0$. Therefore, $M(x^*)$ is not a singleton if and only if there exists $0 \neq \lambda$ such that $\lambda \in [Dg(x^*)T_{x^*}\mathcal{M}]^{\perp}$ and $\lambda \in \mathcal{R}_{\mathcal{N}_{\mathcal{K}}(g(x^*))}(y^*)$. This completes the proof. □

By Proposition 2, we obtain the following result on the uniqueness of Lagrangian multipliers of (3.1), immediately.

Theorem 2 *Suppose that x^* is a feasible solution to (3.1) and $M(x^*) \neq \emptyset$. Let $y^* \in M(x^*)$, then $M(x^*)$ is a singleton if M-SRCQ (3.11) holds at x^* with respect to y^*.*

Proof Now suppose that M-SRCQ condition (3.11) holds at (x^*, y^*). Let $S = Dg(x^*)T_{x^*}\mathcal{M} + \mathcal{T}_{\mathcal{K}}(g(x^*))$. Then,

$$S^\circ = (Dg(x^*)T_{x^*}\mathcal{M})^\circ \cap \left(\mathcal{T}_{\mathcal{K}}(g(x^*)) \cap y^{*\perp}\right)^\circ.$$

Since $(Dg(x^*)T_{x^*}\mathcal{M})^\circ = (Dg(x^*)T_{x^*}\mathcal{M})^\perp$ and

$$\left(\mathcal{T}_{\mathcal{K}}(g(x^*)) \cap y^{*\perp}\right)^\circ = \mathrm{cl}\left(\mathcal{N}_{\mathcal{K}}(g(x^*)) + \mathrm{span}\,(y^*)\right) = \mathcal{T}_{\mathcal{N}_{\mathcal{K}}(g(x^*))}(y^*),$$

it follows that

$$S^\circ = (Dg(x^*)T_{x^*}\mathcal{M})^\perp \cap \mathcal{T}_{\mathcal{N}_{\mathcal{K}}(g(x^*))}(y^*).$$

By noting that $\mathcal{R}_{\mathcal{N}_{\mathcal{K}}(g(x^*))}(y^*) \subset \mathcal{T}_{\mathcal{N}_{\mathcal{K}}(g(x^*))}(y^*)$, we know that (3.14) is satisfied, which implies that $M(x^*) = \{y^*\}$. This completes the proof. $\square$

Remark 3 It is easy to see that the manifold constraint non-degeneracy condition is stronger than the M-SRCQ condition, since $\mathrm{lin}\,(\mathcal{T}_{\mathcal{K}}(g(x^*))) \subset \mathcal{T}_{\mathcal{K}}(g(x^*)) \cap y^{*\perp}$ (cf. proof of [7, Proposition 4.73]). Therefore, constraint non-degeneracy (3.12) also implies the uniqueness of the multiplier.

Remark 4 We mention that equivalence between M-RCQ or M-SRCQ for problem (3.1) and the Euclidean RCQ or SRCQ for (3.2) can be obtained under any chart around the given point. However, we use the normal coordinate chart here for simplicity.

Remark 5 When $\mathcal{M}$ reduce to a special kind of embedded manifold which is written as $\mathcal{M} = \{x \in \mathbb{X} \mid h(x) = 0\}$, where $h : \mathbb{X} \to \mathbb{R}^m$ is smooth and $h'(x)$ has full row rank for all $x \in \mathbb{X}$, then by [2, Section 3.5.7] the tangent space at $x \in \mathcal{M}$ is $T_x\mathcal{M} = \mathrm{Ker}(h'(x))$. If we regard (3.1) as a constrained Euclidean optimization, the Euclidean RCQ condition at a feasible point x for this problem is given by

$$\begin{bmatrix} g'(x) \\ h'(x) \end{bmatrix} \mathbb{X} + \begin{bmatrix} \mathcal{T}_{\mathcal{K}}(g(x)) \\ \mathcal{T}_{\{0\}^m}(h(x)) \end{bmatrix} = \begin{bmatrix} \mathbb{Y} \\ \mathbb{R}^m \end{bmatrix}, \tag{3.15}$$

the SRCQ condition at x with respect to the corresponding multiplier (y, z) is

$$\begin{bmatrix} g'(x) \\ h'(x) \end{bmatrix} \mathbb{X} + \begin{bmatrix} \mathcal{T}_{\mathcal{K}}(g(x)) \cap y^\perp \\ \mathcal{T}_{\{0\}^m}(h(x)) \cap z^\perp \end{bmatrix} = \begin{bmatrix} \mathbb{Y} \\ \mathbb{R}^m \end{bmatrix}, \tag{3.16}$$

and the constraint non-degeneracy at x is

$$\begin{bmatrix} g'(x) \\ h'(x) \end{bmatrix} \mathbb{X} + \begin{bmatrix} \mathrm{lin}\,(\mathcal{T}_{\mathcal{K}}(g(x))) \\ \mathrm{lin}\,(\mathcal{T}_{\{0\}^m}(h(x))) \end{bmatrix} = \begin{bmatrix} \mathbb{Y} \\ \mathbb{R}^m \end{bmatrix}. \tag{3.17}$$

We will show that the RCQ condition (3.15) (SRCQ (3.16), constraint non-degeneracy (3.17)) is equivalent to the M-RCQ (M-SRCQ, manifold constraint non-degeneracy) condition.

It is obvious that $\mathcal{T}_{\{0\}^m}(h(x)) = \{0\}$ for any feasible x. If M-RCQ (3.10) holds at x, then RCQ (3.15) is satisfied by $Dg(x)T_x\mathcal{M} = g'(x)T_x\mathcal{M} \subseteq g'(x)\mathbb{X}$ and the fact that $h'(x)$ is of full row rank. Conversely, if RCQ (3.15) holds at x, for any $(d_1, d_2) \in \mathbb{Y} \times \mathbb{R}^m$, there exists $\xi \in \mathbb{X}$ and $\eta \in \mathcal{T}_{\mathcal{K}}(g(x))$, such that $g'(x)\xi + \eta = d_1$ and $h'(x)\xi = d_2$. Specially taking $d_2 = 0$, we have $\xi \in \mathrm{Ker}\,(h'(x)) = T_x\mathcal{M}$, which implies that M-RCQ (3.10) is fulfilled. The equivalence between the constraint non-degeneracy conditions can be similarly obtained. For the SRCQ conditions, we only need to show that at any stationary point x, the unique multiplier to the constraint of

(3.1) is the same as the multiplier to the Euclidean form problem. The KKT condition of the constrained Euclidean problem can be written as

$$
\begin{cases}
\tilde{L}(x, y, z) = \nabla f(x) + g'(x)^*y + h'(x)^*z = 0, \\
y \in \mathcal{N}_{\mathcal{K}}(g(x)), \\
h(x) = 0.
\end{cases}
\tag{3.18}
$$

The first equality in (3.18) implies that $\nabla f(x) + g'(x)^*y \in \text{Range}(h'(x)^*) = N_x\mathcal{M}$, or equivalently, $\text{grad}\, f(x) + Dg(x)^*y = \Pi_x(\nabla f(x) + g'(x)^*y) = 0$. Thus the SRCQ condition (3.16) holds at (x, y, z) if and only if the M-SRCQ condition (3.11) holds at (x, y).

It is worth mentioning that when $h'(x)$ is not onto, the RCQ condition (3.15) cannot be satisfied, since $h'(x)\mathbb{X} \neq \mathbb{R}^m$ and $\mathcal{T}_{\{0\}^m}(h(x)) = \{0\}$. Therefore, the SRCQ condition (3.16) and the constraint non-degeneracy (3.17) can also never hold. A special but important case is the Stiefel manifold, where $\mathcal{M} = \{X \in \mathbb{R}^{n \times p} \mid h(X) = X^\top X - I_p = 0\}$. It is easy to verified that

$$
h'(X)\mathbb{R}^{n \times p} = \{X^\top Z + Z^\top X \mid Z \in \mathbb{R}^{n \times p}\},
$$

which is contained in the symmetric matrices set and strictly contained in $\mathbb{R}^{p \times p}$ if $p > 1$. Thus the SRCQ condition (3.16) will not be satisfied for any optimization problem over the Stiefel manifold when $p > 1$. In contrast, our M-SRCQ (3.11) may hold in this case as long as $Dg(x)T_x\mathcal{M} = \mathbb{Y}$, which mostly depends on the dimension of $g(x)$ rather than the manifold constraint.

For the $\mathcal{C}^2$-cone reducible set $\mathcal{K}$, we will introduce the second order necessary and sufficient conditions to problem (3.1). Firstly, we define the quadratic growth condition, and its connection with the second order sufficient condition is shown in Theorem 3.

Definition 5 We say that the quadratic growth condition holds at $x^* \in \mathcal{M}$, if there exist a neighborhood $\mathcal{N}$ of x^* and a positive constant $c > 0$ such that for all $x \in \mathcal{F} \cap \mathcal{N}$, the following inequality holds:
$$
f(x) \geq f(x^*) + cd^2(x, x^*).
$$

Theorem 3 *Suppose that x^* is a locally optimal solution to problem (3.1) and M-RCQ (3.10) holds at x^*. Then the following manifold second order necessary condition (M-SONC) holds:*

$$
\sup_{y \in M(x^*)} \left\{ \langle \xi, \text{Hess}_x L(x^*; y)\xi \rangle - \sigma\left(y, \mathcal{T}^2_{\mathcal{K}}(g(x^*), Dg(x^*)\xi)\right) \right\} \geq 0 \quad \forall \xi \in \mathcal{C}(x^*), \tag{3.19}
$$

where for any $y \in \mathbb{Y}$, $\text{Hess}_x L(x^*; y)$ is the Hessian of $L(\cdot; y)$ at x^* and $\sigma(y, \mathcal{D})$ is the support function of set $\mathcal{D}$ at y. Conversely, suppose x^* is a stationary point of problem (3.1) and M-RCQ (3.10) holds at x^*. Then the following manifold second order sufficient condition (M-SOSC)

$$
\sup_{y \in M(x^*)} \left\{ \langle \xi, \text{Hess}_x L(x^*; y)\xi \rangle - \sigma\left(y, \mathcal{T}^2_{\mathcal{K}}(g(x^*), Dg(x^*)\xi)\right) \right\} > 0 \quad \forall \xi \in \mathcal{C}(x^*)\backslash\{0\} \tag{3.20}
$$

is necessary and sufficient for the quadratic growth condition at the point x^* for problem (3.1).

Proof Let $(\mathcal{U}, \varphi)$ be the normal coordinate chart at x^*. Note that M-RCQ (3.10) holds at x^* indicates that RCQ (3.8) holds at $\hat{x}^*$. It is well known [6, Theorem 5.2] in Euclidean space that the traditional second order necessary condition holds for problem (3.2) at the locally optimal solution $\hat{x}^* = \varphi(x^*)$ if RCQ (3.8) holds. Thus, by (2.11), we have

$$
\langle \xi, \text{Hess}_x L(x^*; y)\xi \rangle = \left\langle \hat{\xi}, \nabla^2_{xx}\widehat{L}(\hat{x}^*; y)\hat{\xi} \right\rangle \quad \forall \xi \in \mathcal{C}(x^*) \quad \text{and} \quad \hat{\xi} = D\varphi(x^*)\xi,
$$

which implies that the inequality (3.19) holds if x^* is a local solution and M-RCQ (3.10) holds.

Conversely, if x^* is a stationary point of problem (3.1), then $\varphi(x^*)$ is a stationary point of problem (3.2). Therefore, the SOSC condition (3.21) for problem (3.2), which is stated as below, is the necessary and sufficient condition for the quadratic growth condition at $\varphi(x^*)$ for problem (3.2)

$$\sup_{y \in \widehat{M}(\hat{x}^*)} \left\{ \left\langle \hat{\xi}, \nabla_{\hat{x}\hat{x}}^2 \widehat{L}(\hat{x}^*; y)\hat{\xi} \right\rangle - \sigma\left(y, \mathcal{T}_{\mathcal{K}}^2\left(\hat{g}(\hat{x}^*), \hat{g}'(\hat{x}^*)\hat{\xi}\right)\right) \right\} > 0 \quad \forall \hat{\xi} \in \widehat{\mathcal{C}}(\hat{x}^*) \backslash \{0\}. \tag{3.21}$$

The quadratic growth condition for (3.2) at $\hat{x}^*$ is that there exists a neighborhood $\widehat{\mathcal{N}}$ of $\hat{x}^*$ and a positive constant $c > 0$, such that for all $\hat{x} \in \mathcal{F} \cap \widehat{\mathcal{N}}$, the following inequality holds

$$\hat{f}(\hat{x}) \geq \hat{f}(\hat{x}^*) + c\|\hat{x} - \hat{x}^*\|^2. \tag{3.22}$$

Shrinking $\widehat{\mathcal{N}}$ until it is contained in $\varphi(\mathcal{U})$. Define $d_1(\cdot) := d(\cdot, x^*)$, we know that $d_1(\cdot)$ is a smooth function on $\mathcal{M}$. Combining with the fact that $\varphi(\cdot)$ is a homomorphism in a neighborhood $\mathcal{U}_1$ of x^*, there exists $\alpha > 0$, such that

$$d_1(x) = d_1 \circ \varphi^{-1}(\varphi(x)) \leq \alpha\|\varphi(x) - \varphi(x^*)\| \quad \forall x \in \mathcal{U}_1. \tag{3.23}$$

Combining (3.22) and (3.23), we obtain that for any $x \in \varphi^{-1}(\widehat{\mathcal{N}}) \cap \mathcal{U}_1$,

$$f(x) \geq f(x^*) + \frac{c}{\alpha^2} d^2(x, x^*). \tag{3.24}$$

It is clear that $\varphi^{-1}(\widehat{\mathcal{N}}) \cap \mathcal{U}_1$ is a neighborhood of x^*. Hence, we complete the proof. $\qquad \square$

Remark 6 Another way to prove Theorem 3 is using the intrinsic method that does not rely on chart, which is similar to the proof of [7, Theorem 3.86]. Thus, the neighborhood of x^* where the quadratic growth condition holds is independent of chart. The reason that we use the chart-relying approach is to build the equivalence between M-SOSC for the manifold problem (3.1) and traditional SOSC for the Euclidean problem (3.2).

Remark 7 It is worth mentioning that if the set $\mathcal{K}$ in (1.1) is the non-negative orthant, then the manifold second order necessary condition (3.19) and the manifold second order sufficient condition (3.20) reduce to the conditions proposed in [46, Theorem 4.2] and [46, Theorem 4.3], respectively, since the support function $\sigma\left(\cdot, \mathcal{T}_{\mathcal{K}}^2\left(g(x^*), Dg(x^*)\xi\right)\right)$ in (3.19) and (3.20) both equal to zero (cf. e.g., [7, page 177]).

Remark 8 If $\mathcal{M}$ is again the embedded submanifold $\mathcal{M} = \{h(x) = 0\}$ with $h'(x)$ of full row rank, then for all $\xi \in T_x\mathcal{M}$, $\langle \xi, \text{Hess}_x L(x; y)\xi \rangle = \left\langle \xi, \nabla_{xx}^2 \tilde{L}(x; y, z)\xi \right\rangle$, where $\tilde{L}$ and z are defined in (3.18). Combining with the fact that the critical cones for these two problems are equal and the support function $\sigma\left(\cdot, \mathcal{T}_{\{0\}}^2\left(h(x), h'(x)\xi\right)\right)$ for manifold constraint equals to zero since the constraint is polyhedral, the traditional Eucildean SOSC holds at x if and only if M-SOSC (3.20) holds at x.

3.2 The equivalent perturbed problem

Before characterizing the robust isolated calmness property of the KKT solution for problem (1.1), we first define the perturbed optimization problem by employing the normal coordinate chart of the Riemannian manifold $\mathcal{M}$.

For a given point $x \in \mathcal{M}$, let $(\mathcal{U}, \varphi)$ be the normal coordinate chart defined around x. Then we are able to transform problem (1.1) into the optimization problem on $\mathbb{R}^n$ around x locally and define the corresponding perturbed optimization problem by:

$$\begin{array}{ll} \min & \hat{f}(\hat{x}) + \theta(\hat{g_1}(\hat{x}) + b) - \langle \hat{a}, \hat{x} \rangle \\ \text{s.t.} & \hat{g_2}(\hat{x}) + c \in \mathcal{Q}, \\ & \hat{x} \in \varphi(\mathcal{U}), \end{array} \tag{3.25}$$

where $\hat{f} := f \circ \varphi^{-1}$, $\hat{g_1} := g_1 \circ \varphi^{-1}$, $\hat{g_2} := g_2 \circ \varphi^{-1}$, $\hat{x} = \varphi(x) = (\varphi_1(x), \varphi_2(x), \cdots, \varphi_n(x))$, and $(\hat{a}, b, c) \in \mathbb{R}^n \times \mathbb{Y} \times \mathbb{Z}$ is the perturbation parameter. Thus, we can transform (3.25) into the following perturbed problem on $\mathcal{M}$:

$$\begin{array}{ll} \min & f(x) + \theta(g_1(x) + b) - \langle \hat{a}, \varphi(x) \rangle \\ \text{s.t.} & g_2(x) + c \in \mathcal{Q}, \\ & x \in \mathcal{U}. \end{array} \tag{3.26}$$

Note that problem (3.26) can be equivalently written as follow

$$\begin{array}{ll} \min & f(x) + t - \langle \hat{a}, \varphi(x) \rangle \\ \text{s.t.} & (g_1(x) + b, t) \in \operatorname{epi} \theta, \\ & g_2(x) + c \in \mathcal{Q}, \\ & x \in \mathcal{U}, \end{array} \tag{3.27}$$

and problem (3.25) can be correspondingly written locally as

$$\begin{array}{ll} \min & \hat{f}(\hat{x}) + t - \langle \hat{a}, \hat{x} \rangle \\ \text{s.t.} & (\hat{g_1}(\hat{x}) + b, t) \in \operatorname{epi} \theta, \\ & \hat{g_2}(\hat{x}) + c \in \mathcal{Q}, \\ & \hat{x} \in \varphi(\mathcal{U}). \end{array} \tag{3.28}$$

Let $\widehat{\mathcal{F}}^e(\hat{a}, b, c)$ be the set of all feasible points of the epigraph form problem (3.28) with a given $(\hat{a}, b, c)$, and $\mathcal{F}^e(\hat{a}, b, c)$ be the set of all feasible points of the epigraph form problem (3.27) with $(\hat{a}, b, c)$. The Lagrangian function of problem (3.27) with $(\hat{a}, b, c) = (0, 0, 0)$ is defined by

$$L^e(x, t; y, z, \tau) := f(x) + t + \langle y, g_1(x) \rangle + \langle z, g_2(x) \rangle + t\tau, \quad (x, t, y, z, \tau) \in \mathcal{M} \times \mathbb{R} \times \mathbb{Y} \times \mathbb{Z} \times \mathbb{R}, \tag{3.29}$$

and the Lagrangian function of problem (3.28) with $(\hat{a}, b, c) = (0, 0, 0)$ is defined by

$$\widehat{L}^e(\hat{x}, t; y, z, \tau) := \hat{f}(\hat{x}) + t + \langle y, \hat{g_1}(\hat{x}) \rangle + \langle z, \hat{g_2}(\hat{x}) \rangle + t\tau, \quad (\hat{x}, t, y, z, \tau) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{Y} \times \mathbb{Z} \times \mathbb{R}. \tag{3.30}$$

For a given perturbation parameter $(\hat{a}, b, c)$, the Karush–Kuhn–Tucker (KKT) optimality condition for problem (3.28) takes the following form:

$$\begin{cases} \hat{a} = \nabla_{\hat{x}} \widehat{L}^e(\hat{x}, t; y, z, \tau) = \nabla \hat{f}(\hat{x}) + \hat{g_1}'(\hat{x})^* y + \hat{g_2}'(\hat{x})^* z, \\ 0 = \nabla_t \widehat{L}^e(\hat{x}, t; y, z, \tau) = 1 + \tau, \\ (y, \tau) \in \mathcal{N}_{\operatorname{epi} \theta}(\hat{g_1}(\hat{x}) + b, t), \\ z \in \mathcal{N}_{\mathcal{Q}}(\hat{g_2}(\hat{x}) + c). \end{cases} \tag{3.31}$$

By (2.3), we know that $\nabla_{\hat{x}} \widehat{L}^e(\hat{x}, t; y, z, \tau) = G_{\varphi(x)} D\varphi(x) \operatorname{grad}_x L^e(x, t; y, z, \tau)$. Hence, from (3.31) we obtain the following system:

$$\begin{cases} (D\varphi(x))^{-1} G_{\varphi(x)}^{-1} \hat{a} = \operatorname{grad}_x L^e(x, t; y, z, \tau) = \operatorname{grad} f(x) + Dg_1(x)^* y + Dg_2(x)^* z, \\ 0 = \nabla_t L^e(x, t; y, z, \tau) = 1 + \tau, \\ (y, \tau) \in \mathcal{N}_{\operatorname{epi} \theta}(g_1(x) + b, t), \\ z \in \mathcal{N}_{\mathcal{Q}}(g_2(x) + c). \end{cases} \tag{3.32}$$

This is acutally the KKT system of problem (3.27), since $(D\varphi(x))^{-1} G_{\varphi(x)}^{-1} \hat{a} = \sum_{i=1}^{n} \hat{a}_i \operatorname{grad} \varphi_i(x)$ if we write $\hat{a} = (\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_n) \in \mathbb{R}^n$. Indeed, from (2.1), we know that for any $\xi \in T_x \mathcal{M}$,

$$\langle (D\varphi(x))^{-1} G_{\varphi(x)}^{-1} \hat{a}, \xi \rangle = (D\varphi(x)(D\varphi(x))^{-1} G_{\varphi(x)}^{-1} \hat{a})^T G_{\varphi(x)} D\varphi(x)\xi = \hat{a}^T D\varphi(x)\xi$$
$$= \left\langle \sum_{i=1}^{n} \hat{a}_i \operatorname{grad} \varphi_i(x), \xi \right\rangle.$$

For a given $(\hat{a}, b, c) \in \mathbb{R}^n \times \mathbb{Y} \times \mathbb{Z}$, the set of all solutions $(\hat{x}, t; y, z, \tau)$ to (3.31) is denoted by $\widehat{\mathcal{S}}_{\mathrm{KKT}}^e(\hat{a}, b, c)$. The set of Lagrangian multipliers with $(\hat{x}, t, \hat{a}, b, c)$ is defined by

$$\widehat{M}^e(\hat{x}, t, \hat{a}, b, c) := \left\{ (y, z, \tau) \in \mathbb{Y} \times \mathbb{Z} \times \mathbb{R} \mid (\hat{x}, t, y, z, \tau) \in \widehat{\mathcal{S}}_{\mathrm{KKT}}^e(\hat{a}, b, c) \right\}.$$

We can also define the solution set of (3.32) as $\mathcal{S}_{\mathrm{KKT}}^e(\hat{a}, b, c)$ and denote the multiplier set as

$$M^e(x, t, \hat{a}, b, c) := \{ (y, z, \tau) \in \mathbb{Y} \times \mathbb{Z} \times \mathbb{R} \mid (x, t, y, z, \tau) \in \mathcal{S}_{\mathrm{KKT}}^e(\hat{a}, b, c) \}.$$

Clearly, there are some relationships between the solution set or multiplier set of (3.31) and (3.32) as below:

$$\widehat{\mathcal{S}}_{\mathrm{KKT}}^e(\hat{a}, b, c) = \tilde{\varphi}(\mathcal{S}_{\mathrm{KKT}}^e(\hat{a}, b, c) \cap (\mathcal{U} \times \mathbb{R} \times \mathbb{Y} \times \mathbb{Z} \times \mathbb{R})) \, (\tilde{\varphi} := (\varphi, \mathrm{id})), \tag{3.33a}$$

$$\widehat{M}^e(\hat{x}, t, \hat{a}, b, c) = M^e(x, t, \hat{a}, b, c). \tag{3.33b}$$

The following proposition shows that the continuity and the isolated calmness of $\widehat{\mathcal{S}}_{\mathrm{KKT}}^e$ are inherited by $\mathcal{S}_{\mathrm{KKT}}^e$.

Proposition 3 *Consider a feasible solution $(x, t) \in \mathcal{F}^e(\hat{a}, b, c)$. If $\widehat{\mathcal{S}}_{\mathrm{KKT}}^e(\hat{a}, b, c)$ is lower semicontinuous or (robustly) isolated calm at $(\hat{a}, b, c)$ for $(\hat{x}, t, y, z, -1)$, then $\mathcal{S}_{\mathrm{KKT}}^e(\hat{a}, b, c)$ is lower semicontinuous or (robustly) isolated calm at $(\hat{a}, b, c)$ for $(x, t, y, z, -1)$.*

Proof For any open neighborhood $\mathcal{V}$ of $(x, t, y, z, -1)$, by (3.33a), we know that $\tilde{\varphi}(\mathcal{V})$ is an open neighborhood of $(\hat{x}, t, y, z, -1)$. If $\widehat{\mathcal{S}}_{\mathrm{KKT}}^e(\hat{a}, b, c)$ is lower semicontinuous at $(\hat{a}, b, c)$ for $(x, t, y, z, -1)$, then there exist an open neighborhood $\mathcal{N}$ of $(\hat{a}, b, c)$, such that

$$\emptyset \neq \widehat{\mathcal{S}}_{\mathrm{KKT}}^e(a_1, b_1, c_1) \cap \tilde{\varphi}(\mathcal{V}) \quad \forall (a_1, b_1, c_1) \in \mathcal{N}.$$

Since φ is a homomorphism, by (3.33a), we have

$$\widehat{\mathcal{S}}_{\mathrm{KKT}}^e(a_1, b_1, c_1) \cap \tilde{\varphi}(\mathcal{V}) \subseteq \tilde{\varphi}\left(\mathcal{S}_{\mathrm{KKT}}^e(a_1, b_1, c_1)\right) \cap \tilde{\varphi}(\mathcal{V}) = \tilde{\varphi}\left(\mathcal{S}_{\mathrm{KKT}}^e(a_1, b_1, c_1) \cap \mathcal{V}\right). \tag{3.34}$$

Thus $\emptyset \neq \mathcal{S}_{\mathrm{KKT}}^e(a_1, b_1, c_1) \cap \mathcal{V}$, $\forall (a_1, b_1, c_1) \in \mathcal{N}$. This means that $\mathcal{S}_{\mathrm{KKT}}^e(\hat{a}, b, c)$ is lower semicontinuous at $(\hat{a}, b, c)$ for $(x, t, y, z, -1)$.

If $\widehat{\mathcal{S}}^e_{\mathrm{KKT}}(\hat{a}, b, c)$ is isolated calm at $(\hat{a}, b, c)$ for $(\hat{x}, t, y, z, -1)$, then there exist open neighborhoods $\mathcal{N}$ of $(\hat{a}, b, c)$ and $\widehat{\mathcal{V}}$ of $(\hat{x}, t, y, z, -1)$ and a constant $\kappa > 0$, such that for any $(\hat{x}_1, t_1, y_1, z_1, \tau_1) \in \widehat{\mathcal{S}}^e_{\mathrm{KKT}}(a_1, b_1, c_1) \cap \widehat{\mathcal{V}}$ and $(a_1, b_1, c_1) \in \mathcal{N}$,

$$\|(\hat{x}_1, t_1, y_1, z_1, \tau_1) - (\hat{x}, t, y, z, -1)\| \leq \kappa \|(a_1, b_1, c_1) - (\hat{a}, b, c)\|.$$

By (3.23), there exist $\alpha > 0$, such that for any $\tilde{x} \in \mathcal{U}$, $d(\tilde{x}, x) \leq \alpha \|\varphi(\tilde{x}) - \varphi(x)\|$. Therefore, for any $(x_1, t_1, y_1, z_1, \tau_1) \in \mathcal{S}^e_{\mathrm{KKT}}(a_1, b_1, c_1) \cap (\mathcal{U} \times \mathbb{R} \times \mathbb{Y} \times \mathbb{Z} \times \mathbb{R}) \cap \tilde{\varphi}^{-1}(\widehat{\mathcal{V}})$ and $(a_1, b_1, c_1) \in \mathcal{N}$,

$$d\left((x_1, t_1, y_1, z_1, \tau_1), (x, t, y, z, -1)\right) \leq (\alpha + 1)\kappa \|(a_1, b_1, c_1) - (\hat{a}, b, c)\|.$$

This implies that $\mathcal{S}^e_{\mathrm{KKT}}(\hat{a}, b, c)$ is isolated calm at $(\hat{a}, b, c)$ for $(x, t, y, z, -1)$.

Moreover, if for each $(a_1, b_1, c_1) \in \mathcal{N}$, $\tilde{\varphi}\left(\mathcal{S}^e_{\mathrm{KKT}}(a_1, b_1, c_1) \cap (\mathcal{U} \times \mathbb{R} \times \mathbb{Y} \times \mathbb{Z} \times \mathbb{R}) \cap \tilde{\varphi}^{-1}(\widehat{\mathcal{V}})\right) \neq \emptyset$, then $\mathcal{S}^e_{\mathrm{KKT}}(a_1, b_1, c_1) \cap \mathcal{V} \neq \emptyset$. Thus, $\mathcal{S}^e_{\mathrm{KKT}}(\hat{a}, b, c)$ is robustly isolated calm at $(\hat{a}, b, c)$ for $(x, t, y, z, -1)$.   □

3.3 The characterization of the robust isolated calmness of KKT solution mapping

In this subsection, we will characterize the robust isolated calmness property of KKT solution mapping for problem (3.26). We first present the analysis of the perturbation properties of $\mathcal{S}^e_{\mathrm{KKT}}(\hat{a}, b, c)$ for (3.27). Let $(\hat{a}, b, c) = (0, 0, 0)$ in (3.27), recalling Definition 4, we say the M-RCQ holds at a feasible solution $(x, \theta(g_1(x)))$ if

$$\begin{bmatrix} (Dg_1(x), 1) \\ (Dg_2(x), 0) \end{bmatrix} (T_x\mathcal{M} \times \mathbb{R}) + \begin{bmatrix} \mathcal{T}_{\mathrm{epi}\,\theta}(g_1(x), \theta(g_1(x))) \\ \mathcal{T}_{\mathcal{Q}}(g_2(x)) \end{bmatrix} = \begin{bmatrix} \mathbb{Y} \times \mathbb{R} \\ \mathbb{Z} \end{bmatrix} \tag{3.35}$$

and say the M-SRCQ holds at $(x, \theta(g_1(x)))$ with respect to $(y, z, -1) \in M^e(x, \theta(g_1(x)), 0, 0, 0)$ if

$$\begin{bmatrix} (Dg_1(x), 1) \\ (Dg_2(x), 0) \end{bmatrix} (T_x\mathcal{M} \times \mathbb{R}) + \begin{bmatrix} \mathcal{T}_{\mathrm{epi}\,\theta}(g_1(x), \theta(g_1(x))) \cap (y, -1)^\perp \\ \mathcal{T}_{\mathcal{Q}}(g_2(x)) \cap z^\perp \end{bmatrix} = \begin{bmatrix} \mathbb{Y} \times \mathbb{R} \\ \mathbb{Z} \end{bmatrix}. \tag{3.36}$$

The critical cone at a feasible point $(x, \theta(g_1(x)))$ for problem (3.27) takes the form of

$$\mathcal{C}^e(x, \theta(g_1(x))) := \{(d_1, d_2) \in T_x\mathcal{M} \times \mathbb{R} \mid (Dg_1(x)d_1, d_2) \in \mathcal{T}_{\mathrm{epi}\,\theta}(g_1(x), \theta(g_1(x))),$$
$$Dg_2(x)d_1 \in \mathcal{T}_{\mathcal{Q}}(g_2(x)), \ Df(x)d_1 + d_2 \leq 0\}. \tag{3.37}$$

If x is further a stationary point for (3.28) with $(\hat{a}, b, c) = (0, 0, 0)$ and there exists $(y, z, -1) \in M^e(x, \theta(g_1(x)), 0, 0, 0)$, then the critical cone can be written as

$$\mathcal{C}^e(x, \theta(g_1(x)))$$
$$= \{(d_1, d_2) \mid (Dg_1(x)d_1, d_2) \in \mathcal{T}_{\mathrm{epi}\,\theta}(g_1(x), \theta(g_1(x))), \ Dg_2(x)d_1 \in \mathcal{T}_{\mathcal{Q}}(g_2(x)), \ Df(x)d_1 + d_2 = 0\}$$
$$= \{(d_1, d_2) \mid (Dg_1(x)d_1, d_2) \in \mathcal{T}_{\mathrm{epi}\,\theta}(g_1(x), \theta(g_1(x))) \cap (y, -1)^\perp, \ Dg_2(x)d_1 \in \mathcal{T}_{\mathcal{Q}}(g_2(x)) \cap z^\perp\}. \tag{3.38}$$

From now on, we always assume that the function θ is $\mathcal{C}^2$-cone reducible at $g_1(x)$ and the set $\mathcal{Q}$ is $\mathcal{C}^2$-cone reducible at $g_2(x)$ (Definition 3). Then, we know that the M-SOSC at $(x, \theta(g_1(x)))$ for problem (3.27) with $(\hat{a}, b, c) = (0, 0, 0)$ holds if for any $d \in \mathcal{C}^e(x, \theta(g_1(x)))\backslash\{0\}$,

$$\sup_{(y, z, -1) \in M^e} \left\langle d, \mathrm{Hess}_{(x,t)} L^e(x, \theta(g_1(x)), y, z, -1)d \right\rangle$$
$$- \sigma\left((y, -1, z), \mathcal{T}^2_{\mathrm{epi}\,\theta \times \mathcal{Q}}\left((g_1(x), \theta(g_1(x)), g_2(x)), ((Dg_1(x), 1)\,d, Dg_2(x)d))\right)\right) > 0. \tag{3.39}$$

Now let us return to the original perturbed problem (3.26). Let $(\hat{a}, b, c)$ be given. We say that x is a feasible solution to problem (3.26) if

$$x \in \mathcal{F}(\hat{a}, b, c) := \{g_1(x) + b \in \operatorname{dom}\theta \mid g_2(x) + c \in \mathcal{Q}\}. \tag{3.40}$$

Denote $l : \mathcal{M} \times \mathbb{Z} \to \mathbb{R}$ by

$$l(x, z) := f(x) + \langle z, g_2(x) \rangle, \quad (x, z) \in \mathcal{M} \times \mathbb{Z}, \tag{3.41}$$

then the KKT condition takes the form of

$$\begin{cases} (D\varphi(x))^{-1} G_{\varphi(x)}^{-1} \hat{a} \in \operatorname{grad}_x l(x; z) + \partial\theta \circ (g_1(x) + b), \\ z \in \mathcal{N}_{\mathcal{Q}}(g_2(x) + c). \end{cases} \tag{3.42}$$

By Proposition 1 and [12, Proposition 2.3.6], we know that $\partial\theta \circ (g_1(x) + b) = Dg_1(x)^* \partial\theta(g_1(x) + b)$. Therefore, the KKT condition (3.42) can be rewritten as

$$\begin{cases} (D\varphi(x))^{-1} G_{\varphi(x)}^{-1} \hat{a} = \operatorname{grad} f(x) + Dg_1(x)^* y + Dg_2(x)^* z, \\ y \in \partial\theta(g_1(x) + b), \\ z \in \mathcal{N}_{\mathcal{Q}}(g_2(x) + c). \end{cases} \tag{3.43}$$

Let $\mathcal{S}_{\mathrm{KKT}} : \mathbb{R}^n \times \mathbb{Y} \times \mathbb{Z} \to \mathcal{M} \times \mathbb{Y} \times \mathbb{Z}$ be the following KKT solution mapping:

$$\mathcal{S}_{\mathrm{KKT}}(\hat{a}, b, c) = \{(x, y, z) \in \mathcal{M} \times \mathbb{Y} \times \mathbb{Z} \mid (x, y, z) \text{ satisfies } (3.43)\}. \tag{3.44}$$

We also denote $M(x, \hat{a}, b, c)$ as the set of Lagrange multipliers with respect to x.

The following proposition establishes the relation between $\mathcal{S}_{\mathrm{KKT}}^e$ and $\mathcal{S}_{\mathrm{KKT}}$. The proof is similar with that of Proposition 3.2 in [14], and we include the proof here for completion.

Proposition 4 *Let $(x^*, \theta(g_1(x^*))) \in \mathcal{M} \times \mathbb{R}$ be a local optimal solution of problem (3.27) with $M^e(x^*, \theta(g_1(x^*)), 0, 0, 0) \neq \emptyset$. Let $(y^*, z^*, -1) \in M^e(x^*, \theta(g_1(x^*)), 0, 0, 0)$. Then the KKT solution mapping $\mathcal{S}_{\mathrm{KKT}}^e$ is robustly isolated calm at the origin for $(x^*, \theta(g_1(x^*)), y^*, z^*, -1)$ if and only if the KKT solution mapping $\mathcal{S}_{\mathrm{KKT}}$ is robustly isolated calm at the origin for (x^*, y^*, z^*).*

Proof It follows from [12, Corollary 2.4.9] that

$$(y, -1) \in \mathcal{N}_{\mathrm{epi}\,\theta}(g_1(x) + b, \theta(g_1(x) + b)) \Leftrightarrow y \in \partial\theta(g_1(x) + b).$$

Thus, if $(x, \theta(g_1(x)), y, z, -1) \in \mathcal{S}_{\mathrm{KKT}}^e(\hat{a}, b, c)$, then $(x, y, z) \in \mathcal{S}_{\mathrm{KKT}}(\hat{a}, b, c)$. The "only if" part now follows directly from the definition of the robust isolated calmness.

Conversely, consider any $(\hat{a}, b, c) \in \mathbb{R}^n \times \mathbb{Y} \times \mathbb{Z}$ and $(x, y, z) \in \mathcal{S}_{\mathrm{KKT}}(\hat{a}, b, c)$. We can obtain that $(x, \theta(g_1(x)), y, z, -1) \in \mathcal{S}_{\mathrm{KKT}}^e(\hat{a}, b, c)$. Since $g_1 \in \mathcal{C}^2$ and θ is convex, there exists $\kappa_1, \kappa_2 > 0$, such that for any x sufficiently close to x^*, we have $\|\theta(g_1(x)) - \theta(g_1(x^*))\| \leq \kappa_1 d(x, x^*)$. Thus the robustly isolated calmness of $\mathcal{S}_{\mathrm{KKT}}$ at the origin for (x^*, y^*, z^*) indicates the robustly isolated calmness of $\mathcal{S}_{\mathrm{KKT}}^e$ at the origin for $(x^*, \theta(g_1(x^*)), y^*, z^*, -1)$. □

By noting that θ is convex on $\mathbb{Y}$, we know from [41, Proposition 8.12, Exercise 8.4] and [12, Proposition 2.2.7] that

$$\partial\theta(g_1(x) + b) = \{y \mid \langle y, d \rangle \leq \theta^{\downarrow}(g(x), d), \text{ for all } d \in \mathbb{Y}\}.$$

Thus, the KKT condition (3.43) can be equivalently written as

$$\begin{cases} \operatorname{grad}_x l(x; z) - (D\varphi(x))^{-1} G_{\varphi(x)}^{-1} \hat{a} = -Dg_1(x)^* y, \\ \theta^{\downarrow}(g_1(x) + b, d) - \langle y, d \rangle \geq 0 \quad \forall d \in \mathbb{Y}, \\ z \in \mathcal{N}_{\mathcal{Q}}(g_2(x) + c). \end{cases} \tag{3.45}$$

Let $(\hat{a}, b, c) = (0, 0, 0)$. Define the critical cone of the function θ and g by

$$\mathcal{C}_{\theta,g}(x, y) := \left\{ d \in \mathbb{Y} \mid \theta^{\downarrow}(g(x); d) = \langle d, y \rangle \right\}. \tag{3.46}$$

The M-RCQ is said to hold at a feasible solution x of the problem (3.26) if

$$\begin{bmatrix} Dg_1(x) \\ Dg_2(x) \end{bmatrix} T_x \mathcal{M} + \begin{bmatrix} \mathcal{T}_{\mathrm{dom}\,\theta}(g_1(x)) \\ \mathcal{T}_{\mathcal{Q}}(g_2(x)) \end{bmatrix} = \begin{bmatrix} \mathbb{Y} \\ \mathbb{Z} \end{bmatrix}, \tag{3.47}$$

and by (2.14), the M-SRCQ is said to hold at a stationary point x with respect to $(y, z) \in M(x, 0, 0, 0)$ if

$$\begin{bmatrix} Dg_1(x) \\ Dg_2(x) \end{bmatrix} T_x \mathcal{M} + \begin{bmatrix} \mathcal{C}_{\theta,g_1}(x, y) \\ \mathcal{T}_{\mathcal{Q}}(g_2(x)) \cap z^{\perp} \end{bmatrix} = \begin{bmatrix} \mathbb{Y} \\ \mathbb{Z} \end{bmatrix}. \tag{3.48}$$

The critical cone at a stationary point x of problem (3.26) with $z \in M(x, 0, 0, 0)$ is

$$\mathcal{C}(x) = \left\{ \xi \in T_x \mathcal{M} \mid Dg_2(x)\xi \in \mathcal{T}_{\mathcal{Q}}(g_2(x)) \cap z^{\perp}, Df(x)\xi + (\theta \circ g_1)^{\circ}(x; \xi) = 0 \right\}. \tag{3.49}$$

By (2.15) and the $\mathcal{C}^2$-cone reducibility of θ and $\mathcal{Q}$, we know that the M-SOSC at x is said to hold if for any $\xi \in \mathcal{C}(x) \backslash \{0\}$,

$$\sup_{y, z \in M(x, 0, 0, 0)} \left\{ \langle \xi, \mathrm{Hess}_x\, l(x, z)\xi \rangle - \psi^*_{(g_1(x), Dg_1(x)\xi)}(y) - \sigma\left(z, \mathcal{T}^2_{\mathcal{Q}}(g_2(x); Dg_2(x)\xi)\right) \right\} > 0, \tag{3.50}$$

where $\psi^*_{(g_1(x), Dg_1(x)\xi)}(\cdot)$ is the conjugate function of $\psi_{(g_1(x), Dg_1(x)\xi)}(\cdot) = \theta^{\downarrow\downarrow}_{-}(g_1(x); Dg_1(x)\xi, \cdot)$ for any $g_1(x) \in \mathrm{dom}\,\theta$ and any $\xi \in T_x \mathcal{M}$.

Remark 9 Now let us consider the M-SRCQ and M-SOSC conditions for problem (1.2), where $f(x) = x_2^2$, $g_1(x) = x_1 - x_2$, $\theta(\cdot) = |\cdot|$ and $g_2(x) = 2x_1 + x_2$. We focus on the properties at the global solution $x^* = \left(\sqrt{2}/2, \sqrt{2}/2\right)^{\top}$ with respect to the multipliers $y^* = \sqrt{2}/2$ and $z^* = 0$. The tangent space of $\mathcal{S}^1 = \{x = (x_1, x_2) \mid x_1^2 + x_2^2 = 1\}$ at x^* is $T_{x^*}\mathcal{S}^1 = \left\{\xi \in \mathbb{R}^2 \mid x^{*\top}\xi = 0\right\} = \left\{(v, -v)^{\top} \mid v \in \mathbb{R}\right\}$. Thus, for any $u \in Dg_1(x^*)T_{x^*}\mathcal{S}^1$, there exists $(v, -v)^{\top} \in \mathbb{R}^2$, such that

$$u = \langle \nabla g_1(x^*), (v, -v)^{\top} \rangle = \left\langle (1, -1)^{\top}, (v, -v)^{\top} \right\rangle = 2v.$$

Since v can be taken arbitrary, $Dg_1(x^*)T_{x^*}\mathcal{S}^1 = \mathbb{R}$. Combining with the fact that $\mathcal{T}_{\mathcal{Q}}(g_2(x^*)) = \mathbb{R}$ and $(z^*)^{\perp} = (0)^{\perp} = \mathbb{R}$, the M-SRCQ condition is satisfied at x^* with respect to (y^*, z^*). Moreover, using [46, Lemma 5.1 and Definition 5.1] and the fact that $\theta \circ g_1(\cdot)$ is convex on $\mathbb{R}^2$, for any $\xi = (v, -v) \in T_{x^*}S^1$,

$$(\theta \circ g_1)^{\circ}(x^*, \xi) = (\theta \circ g_1)'(x^*, \xi) = \lim_{t \downarrow 0} \frac{|x_1^* + tv - (x_2^* - tv)| - |x_1^* - x_2^*|}{t} = 2|v|.$$

Therefore, the critical cone at x^* is

$$\mathcal{C}(x^*) = \left\{ \xi = \langle v, -v \rangle \mid Dg_2(x^*)\xi \in \mathbb{R}, Df(x^*)\xi + (\theta \circ g_1)^{\circ}(x^*, \xi) = 0 \right\}$$
$$= \left\{ \langle v, -v \rangle \mid -\sqrt{2}v + 2|v| = 0 \right\} = \{(0, 0)\},$$

and the M-SOSC condition is thus satisfied.

Next, we shall study the characterization of the robust isolated calmness of $\mathcal{S}_{\mathrm{KKT}}$ at $(0,0,0)$ for a local optimal solution x^* of problem (3.26). First, let $(\hat{x},t) \in \mathbb{R}^n$ be a feasible solution to problem (3.28) with $(\hat{a},b,c) = (0,0,0)$. It is known (cf. e.g., [15, Theorem 24] and [14, Proposition 3.1]) that the mapping $\widehat{\mathcal{S}}^e_{\mathrm{KKT}}$ has the isolated calmness property at $(0,0,0)$ for $(\hat{x},t)$ if the SOSC holds at $(\hat{x},t)$ and SRCQ holds at $(\hat{x},t)$ with respect to the multiplier. Therefore, the following results for the semicontinuity and isolated calmness of $\mathcal{S}^e_{\mathrm{KKT}}$ of the problem (3.27) extends [15] to manifold optimization, which can be easily obtained by using the normal coordinate chart around x^*. Combining Proposition 3 with [15, Theorem 17], we obtain the following theorem.

Theorem 4 *Let $(x^*, \theta(g_1(x^*)))$ be a feasible solution of problem (3.27) with $(\hat{a},b,c) = (0,0,0)$. Suppose that the M-SRCQ (3.36) holds at $(x^*, \theta(g_1(x^*)))$ with respect to $(y^*, z^*, -1) \in M^e(x^*, \theta(g_1(x^*)), 0,0,0)$ and the M-SOSC (3.39) holds at $(x^*, \theta(g_1(x^*)))$ for problem (3.27) with respect to $(\hat{a},b,c) = (0,0,0)$. Then the set-valued mapping $\mathcal{S}^e_{\mathrm{KKT}}$ is lower semicontinuous at $(0,0,0, x^*, \theta(g_1(x^*)), y^*, z^*, -1) \in \mathrm{gph}\,\mathcal{S}^e_{\mathrm{KKT}}$.*

By applying [15, Theorem 24] to $\widehat{S}^e_{\mathrm{KKT}}$ and using Proposition 3, we obtain the following result on the characterization of the (robust) isolated calmness of $\mathcal{S}^e_{\mathrm{KKT}}$.

Theorem 5 *Let $(x^*, \theta(g_1(x^*)))$ be a feasible solution to problem (3.27) with $(\hat{a},b,c) = (0,0,0)$. Suppose that the M-RCQ (3.35) holds at $(x^*, \theta(g_1(x^*)))$. Let $(y^*, z^*, -1) \in M^e(x^*, \theta(g_1(x^*)), 0,0,0) \neq \emptyset$. Then the following statements are equivalent:*

(i) the M-SRCQ (3.36) holds at $(x^*, \theta(g_1(x^*)))$ with respect to $(y^*, z^*, -1)$ and the M-SOSC (3.39) holds at $(x^*, \theta(g_1(x^*)))$ for problem (3.27) with $(\hat{a},b,c) = (0,0,0)$;
(ii) $(x^*, \theta(g_1(x^*)))$ is a locally optimal solution to problem (3.27) with $(\hat{a},b,c) = (0,0,0)$ and $\mathcal{S}^e_{\mathrm{KKT}}$ is (robustly) isolated calm at the origin for $(x^*, \theta(g_1(x^*)), y^*, z^*, -1)$.

When $(\hat{a},b,c) = (0,0,0)$, the KKT system (3.43) is equivalent to the following system of nonsmooth equations: $0 \in F(x,y,z)$, where $F : \mathcal{M} \times \mathbb{Y} \times \mathbb{Z} \to T\mathcal{M} \times \mathbb{Y} \times \mathbb{Z}$ is the natural mapping defined by

$$F(x,y,z) := \begin{bmatrix} \mathrm{grad}\, f(x) + Dg_1(x)^*y + Dg_2(x)^*z \\ g_1(x) - \mathrm{prox}_\theta(g_1(x) + y) \\ g_2(x) - \Pi_{\mathcal{Q}}(g_2(x) + z) \end{bmatrix}, \quad (x,y,z) \in \mathcal{M} \times \mathbb{Y} \times \mathbb{Z}. \qquad (3.51)$$

where $\mathrm{prox}_\theta(\cdot)$ is the proximal mapping of θ which is defined as $\mathrm{prox}_\theta(u) := \mathrm{argmin}_{y \in \mathbb{Y}}\ \theta(y) + \frac{1}{2}\|u - y\|^2$, and $\Pi_{\mathcal{Q}}(\cdot)$ denotes the projection onto $\mathcal{Q}$. It is clear that $(0,0,0, x^*, y^*, z^*) \in \mathrm{gph}\,\mathcal{S}_{\mathrm{KKT}}$ if and only if $(0,0,0, x^*, y^*, z^*) \in \mathrm{gph}\,F^{-1}$. The proof of Lemma 1 is almost the same as the proof of [15, Lemma 18], and we omit it here.

Lemma 1 *Let $(0,0,0, x^*, y^*, z^*) \in \mathrm{gph}\,\mathcal{S}_{\mathrm{KKT}}$. The set-valued mapping $\mathcal{S}_{\mathrm{KKT}}$ is isolated calm at the origin for (x^*, y^*, z^*) if and only if the set-valued mapping F^{-1} is isolated calm at the origin for (x^*, y^*, z^*).*

Combining Lemma 1, Proposition 4 and Theorem 5, we can derive the main result of this section in the next Proposition.

Proposition 5 *Let $x^* \in \mathcal{M}$ be a local optimal solution of problem (3.26) with $(\hat{a},b,c) = (0,0,0)$. Suppose that the M-RCQ (3.47) holds at x^*. Let $(y^*, z^*) \in M(x^*, 0,0,0)$. Then the following statements are equivalent :*

(i) The M-SOSC (3.50) holds at x^* and the M-SRCQ (3.48) holds at x^* with respect to (y^*, z^*);

(ii) The KKT solution mapping $\mathcal{S}_{\mathrm{KKT}}$ is robustly isolated calm at the origin for (x^*, y^*, z^*);
(iii) The set-valued mapping F^{-1} is robustly isolated calm at the origin for (x^*, y^*, z^*).

Remark 10 The isolated calmness of F^{-1} at the origin for (x^*, y^*, z^*) implies that there exist $\kappa > 0$, and a neighborhood $\mathcal{V}$ of (x^*, y^*, z^*) such that

$$d(x, x^*) + \|y - y^*\| + \|z - z^*\| \leq \kappa \|F(x, y, z)\| \quad \forall (x, y, z) \in \mathcal{V}. \tag{3.52}$$

4 Riemannian augmented Lagrangian method

In this section, we shall present the inexact augmented Lagrangian method for solving the problem (1.1) and establish its local convergence analysis. When $\mathcal{Q}$ is polyhedral, Algorithm 1 will reduce to the Riemannian augmented Lagrangian method proposed in [47].

The Lagrangian function of (1.1) is defined by

$$L(x, y, z) := f(x) + \langle y, g_1(x) \rangle + \langle z, g_2(x) \rangle, \quad (x, y, z) \in \mathcal{M} \times \mathbb{Y} \times \mathbb{Z}, \tag{4.1}$$

and the augmented Lagrangian function $L_\rho : \mathcal{M} \times \mathbb{Y} \times \mathbb{Z} \to \mathbb{R}$ is in the form of

$$L_\rho(x, y, z) := f(x) + \theta^\rho \left(g_1(x) + \frac{y}{\rho} \right) + \frac{\rho}{2} \operatorname{dist}^2 \left(g_2(x) + \frac{z}{\rho}, \mathcal{Q} \right), \tag{4.2}$$

where θ^ρ is the Moreau-Yosida regularization of θ defined by $\theta^\rho(u) := \min_{y \in \mathbb{Y}} \theta(y) + \frac{\rho}{2} \|u - y\|^2$. Moreover, we define the auxiliary function

$$V(x, y, z, \rho) := \max \left\{ \left\| g_1(x) - \operatorname{prox}_\theta \left(g_1(x) + \frac{y}{\rho} \right) \right\|, \left\| g_2(x) - \Pi_\mathcal{Q} \left(g_2(x) + \frac{z}{\rho} \right) \right\| \right\}. \tag{4.3}$$

The detail of Riemannian ALM for solving (1.1) is presented in Algorithm 1.

4.1 Local convergence analysis of the Riemannian ALM

In this subsection, motivated by the proof in [27], we present the local convergence result of Algorithm 1. We first define the KKT residual mapping as follows.

$$R(x, y, z) := \| \operatorname{grad}_x L(x; y, z) \| + \|g_1(x) - \operatorname{prox}_\theta(g_1(x) + y)\| + \|g_2(x) - \Pi_\mathcal{Q}(g_2(x) + z)\|. \tag{4.9}$$

The next theorem characterize the local error bound for the distance between the points around a KKT solution (x^*, y^*, z^*) and this solution by a local Lipschitz property of $\mathcal{S}_{\mathrm{KKT}}$.

Theorem 6 *Let (x^*, y^*, z^*) be a KKT point of problem (1.1). Then the following assertions are equivalent:*

(i) There exist a neighborhood $\mathcal{U}$ of x^* and $\kappa > 0$ such that, for all $q = (\hat{a}, b, c) \in \mathbb{R}^n \times \mathbb{Y} \times \mathbb{Z}$ close to $(0, 0, 0)$, any solution $(x_q, y_q, z_q) \in \mathcal{U} \times \mathbb{Y} \times \mathbb{Z}$ of the perturbed KKT system (3.43) satisfies $d(x_q, x^*) + \operatorname{dist}((y_q, z_q), M(x^*, 0, 0, 0)) \leq \kappa \|q\|_{\mathbb{R}^n \times \mathbb{Y} \times \mathbb{Z}}$.
(ii) There exist a neighborhood $\mathcal{U}$ of x^* and $\kappa > 0$ such that, for all $(x, y, z) \in \mathcal{U} \times \mathbb{Y} \times \mathbb{Z}$ with $R(x, y, z)$ sufficiently small,

$$d(x, x^*) + \operatorname{dist}((y, z), M(x^*, 0, 0, 0)) \leq \kappa R(x, y, z). \tag{4.10}$$

Algorithm 1 Riemannian augmented Lagrangian method

Input: Let $(x^0, y^0, z^0) \in \mathcal{M} \times \mathbb{Y} \times \mathbb{Z}$, $B \subseteq \mathbb{Y} \times \mathbb{Z}$ be bounded, $\rho^0 > 0$, $\gamma > 1$, $\alpha \in (0, 1)$, a sequence $\{\epsilon^k\} \in \mathbb{R}_+$ convergence to 0, and set $k := 0$.

1: If (x^k, y^k, z^k) satisfies a suitable termination criterion: STOP.
2: Choose $(w^k, p^k) \in B$ and compute an x^{k+1}, such that

$$x^{k+1} \approx \operatorname*{argmin}_{x \in \mathcal{M}} L_{\rho^k}(x, w^k, p^k) := f(x) + \theta^{\rho^k}\left(g_1(x) + \frac{w^k}{\rho^k}\right) + \frac{\rho^k}{2}\operatorname{dist}^2\left(g_2(x) + \frac{q^k}{\rho^k}, \mathcal{Q}\right). \quad (4.4)$$

Specially, we need to find x^{k+1} satisfying

$$\|\operatorname{grad} L_{\rho^k}(x, w^k, p^k)\| \leq \epsilon^k. \quad (4.5)$$

3: Update the vector of multipliers to

$$y^{k+1} = \rho^k\left[g_1(x^{k+1}) + \frac{w^k}{\rho^k} - \operatorname{prox}_{\theta/\rho^k}\left(g_1(x^{k+1}) + \frac{w^k}{\rho^k}\right)\right], \quad (4.6)$$

$$z^{k+1} = \rho^k\left[g_2(x^{k+1}) + \frac{p^k}{\rho^k} - \Pi_{\mathcal{Q}}\left(g_2(x^{k+1}) + \frac{p^k}{\rho^k}\right)\right]. \quad (4.7)$$

4: If $k = 0$ or

$$V\left(x^{k+1}, w^k, p^k, \rho^k\right) \leq \tau V\left(x^k, w^{k-1}, p^{k-1}, \rho^{k-1}\right) \quad (4.8)$$

holds, set $\rho_{k+1} = \rho^k$; otherwise, set $\rho_{k+1} = \gamma\rho^k$.
5: Set $k = k + 1$ and go to step 2.

Proof (ii) $\Rightarrow$ (i) Let $q = (\hat{a}, b, c) \in \mathbb{R}^n \times \mathbb{Y} \times \mathbb{Z}$. Suppose that (x_q, y_q, z_q) is the solution of (3.43) for q, then

$$\operatorname{grad}_x L(x_q, y_q, z_q) = (D\varphi(x_q))^{-1}G_{\varphi(x_q)}^{-1}\hat{a}.$$

Since $(D\varphi(x))^{-1}$ and $G_{\varphi(x)}^{-1}$ are smooth functions of x, there exists $L > 0$ and a neighborhood $\mathcal{U}_{x^*}$ of x^*, such that for any $x \in \mathcal{U}_{x^*}$, we have

$$\|(D\varphi(x))^{-1}G_{\varphi(x)}^{-1}\hat{a}\| \leq L\|\hat{a}\|.$$

By (3.43), $g_1(x_q) + b = \operatorname{prox}_\theta(g_1(x_q) + b + y_q)$ and $g_2(x_q) + c = \Pi_{\mathcal{Q}}(g_2(x_q) + c + z_q)$. Thus we obtain that

$$\begin{aligned}
&\|g_1(x_q) - \operatorname{prox}_\theta(g_1(x_q) + y_q)\| \\
&= \|g_1(x_q) - \operatorname{prox}_\theta(g_1(x_q) + y_q)\| - \|g_1(x_q) + b - \operatorname{prox}_\theta(g_1(x_q) + b + y_q)\| \\
&\leq \|g_1(x_q) - \operatorname{prox}_\theta(g_1(x_q) + y_q) - (g_1(x_q) + b - \operatorname{prox}_\theta(g_1(x_q) + b + y_q))\| \\
&\leq \|g_1(x_q) - (g_1(x_q) + b)\| = \|b\|,
\end{aligned}$$

where the last inequality holds by Moreau decomposition [35] and the fact that the proximal operator is 1-Lipschitz (see [41, Proposition 12.19]). Similarly, we have $\|g_2(x_q) - \Pi_{\mathcal{Q}}(g_2(x_q) + z_q)\| \leq \|c\|$. Therefore,

$$\begin{aligned}
R(x_q, y_q, z_q) &= \|\operatorname{grad}_x L(x_q, y_q, z_q)\| + \|g_1(x_q) - \operatorname{prox}_\theta(g_1(x_q) + y_q)\| \\
&\quad + \|g_2(x_q) - \Pi_{\mathcal{Q}}(g_2(x_q) + z_q)\| \\
&\leq L\|\hat{a}\| + \|b\| + \|c\| \leq (L+1)\|q\|.
\end{aligned} \quad (4.11)$$

Thus, taking q sufficiently close to zero and substituting (4.11) into (4.10), we now obtain

$$d(x_q, x^*) + \operatorname{dist}((y_q, z_q), M(x^*, 0, 0, 0)) \leq \kappa R(x_q, y_q, z_q) \leq \kappa(L+1)\|q\|.$$

(i) $\Rightarrow$ (ii) Let $(x, y, z) \in \mathcal{U} \times \mathbb{Y} \times \mathbb{Z}$ and we define

$$\tilde{g}_1 := \mathrm{prox}_\theta(g_1(x) + y), \quad \tilde{y} := g_1(x) + y - \tilde{g}_1, \quad \tilde{g}_2 := \Pi_{\mathcal{Q}}(g_2(x) + z), \quad \tilde{z} := g_2(x) + z - \tilde{g}_2.$$

Now take $\hat{a} = G_{\varphi(x)} D\varphi(x) \mathrm{grad}_x L(x, \tilde{y}, \tilde{z})$, $b = g_1(x) - \tilde{g}_1$ and $c = g_2(x) - \tilde{g}_2$. It is easy to obtain that $\tilde{y} \in \partial\theta(g_1(x) + b)$ and $\tilde{z} \in \mathcal{N}_q(g_2(x) + c)$. Hence, $(x, \tilde{y}, \tilde{z})$ is the solution of (3.43) with $(\hat{a}, b, c)$. We further have

$$\|\tilde{y} - y\| = \|b\| = \|\tilde{g}_1 - g_1(x)\| = \|g_1(x) - \mathrm{prox}_\theta(g_1(x) + y)\| \le R(x, y, z)$$

$$\text{and} \quad \|\tilde{z} - z\| = \|c\| = \|\tilde{g}_2 - g_2(x)\| = \|g_2(x) - \Pi_{\mathcal{Q}}(g_2(x) + z)\| \le R(x, y, z)$$

by the definition of $\tilde{y}, \tilde{z}$ and $R(x, y, z)$. Now assume that $\|Dg_1(x)^*\| \le c_1$ and $\|Dg_2(x)^*\| \le c_2$ for all $x \in \mathcal{U}$ with some $c_1, c_2 \ge 0$, we have

$$\begin{aligned}
\|(\hat{a}, b, c)\| &= \|G_{\varphi(x)} D\varphi(x) \mathrm{grad}_x L(x, \tilde{y}, \tilde{z})\| + \|b\| + \|c\| \\
&\le L_1 \|\mathrm{grad}_x L(x, \tilde{y}, \tilde{z})\| + \|b\| + \|c\| \\
&\le L_1 \|\mathrm{grad}_x L(x, y, z)\| + (L_1 c_1 + 1)\|b\| + (L_1 c_2 + 1)\|c\| \\
&\le (L_1(1 + c_1 + c_2) + 2) R(x, y, z).
\end{aligned}$$

The first inequality is obtained from the smoothness of $D\varphi(\cdot)$ and $G_{\varphi(\cdot)}$ in $\mathcal{U}$ which implies that there exists $L_1 > 0$, such that for any $x \in \mathcal{U}$, $\|G_{\varphi(x)} D\varphi(x) \mathrm{grad}_x L(x, \tilde{y}, \tilde{z})\| \le L_1 \|\mathrm{grad}_x L(x, \tilde{y}, \tilde{z})\|$. Therefore, taking $R(x, y, z)$ close enough to 0, we can then apply (i) to $(x, \tilde{y}, \tilde{z})$ and obtain

$$d(x, x^*) + \mathrm{dist}\left((\tilde{y}, \tilde{z}), M(x^*, 0, 0, 0)\right) \le \kappa\|(\hat{a}, b, c)\| \le \kappa\left(L_1(1 + c_1 + c_2) + 2\right) R(x, y, z).$$

Moreover, by $\|\tilde{y} - y\| \le R(x, y, z)$ and $\|\tilde{z} - z\| \le R(x, y, z)$,

$$\mathrm{dist}\left((y, z), M(x^*, 0, 0, 0)\right) \le \mathrm{dist}\left((\tilde{y}, \tilde{z}), M(x^*, 0, 0, 0)\right) + 2R(x, y, z).$$

We finally obtain

$$d(x, x^*) + \mathrm{dist}\left((y, z), M(x^*, 0, 0, 0)\right) \le \left[\kappa\left(L_1(1 + c_1 + c_2) + 2\right) + 2\right]R(x, y, z).$$

The proof is then complete. $\square$

For any $\sigma \ge 0$, we denote by $M_\sigma(x)$ the set of (y, z) satisfying the following relationships:

$$\|\mathrm{grad}_x L(x, y, z)\| \le \sigma, \quad y \in \partial\theta(g_1(x)), \quad z \in \mathcal{N}_{\mathcal{Q}}(g_2(x)). \tag{4.12}$$

The following proposition is a direct conclusion from [7, Theorem 4.43].

Proposition 6 *Suppose that M-RCQ holds at x^*. Then for any $\sigma \ge 0$, the solution sets $M_\sigma(x)$ are bounded for all x in a neighborhood of x^*.*

Proof Let us denote $\widehat{M}_\sigma(\hat{x})$ as the set of (y, z) satisfying

$$\|\nabla_{\hat{x}} \widehat{L}(\hat{x}, y, z)\| \le \sigma, \quad y \in \partial\theta(\hat{g}_1(\hat{x})), \quad z \in \mathcal{N}_{\mathcal{Q}}(\hat{g}_2(\hat{x})), \tag{4.13}$$

where $\hat{x} = \varphi(x)$ and $\widehat{L} = L \circ \varphi^{-1}$. By (2.3) and the Lipschitz property of $D\varphi(x)$ and $G_{\varphi(x)}$ at x^*, if (y, z) satisfying (4.12) for some σ, then there exists $\alpha \ge 0$, such that for any x in a neighborhood $\mathcal{U}_{x^*}$, $(y, z) \in \widehat{M}_{\alpha\sigma}(\hat{x})$. Thus, $M_\sigma(x) \subseteq \widehat{M}_{\alpha\sigma}(\hat{x})$. Using [7, Theorem 4.43], we have the boundness of $\widehat{M}_{\alpha\sigma}(\hat{x})$, which implies the boundness of $M_\sigma(x)$. $\square$

With this proposition, we can establish the following local error bound around the KKT point (x^*, y^*, z^*) in terms of the KKT residual mapping.

Theorem 7 *Assume that the problem (3.26) admits a KKT point (x^*, y^*, z^*) with $(\hat{a}, b, c) = (0, 0, 0)$ which satisfies M-SOSC (3.50) and M-SRCQ (3.48). Then $M(x^*, 0, 0, 0) = \{(y^*, z^*)\}$ and there exist $c_1, c_2 > 0$ such that, for all $(x, y, z) \in \mathcal{M} \times \mathbb{Y} \times \mathbb{Z}$ with x sufficiently close to x^* and $R(x, y, z)$ sufficiently small,*

$$c_1 R(x, y, z) \leq d(x, x^*) + \|y - y^*\| + \|z - z^*\| \leq c_2 R(x, y, z). \tag{4.14}$$

Proof Let $\{(\hat{a}^k, b^k, c^k)\}$ be a sequence converging to 0 and $\{(x^k, y^k, z^k)\}$ be a sequence of the solutions of (3.43) corresponding to $\{(\hat{a}^k, b^k, c^k)\}$ such that $x^k \to x^*$. We know from Proposition 6 that $\{(y^k, z^k)\}$ is bounded, thus every accumulation point of $\{(y^k, z^k)\}$ is the multiplier corresponding to x^*. Since M-SRCQ holds at x^* with respect to (y^*, z^*), $M(x^*) = \{(y^*, z^*)\}$ by Theorem 2 and we obtain that (y^k, z^k) converge to (y^*, z^*). Using Remark 10 and the definition of $R(x^k, y^k, z^k)$, there exist $\kappa > 0$, such that for k sufficiently large,

$$d(x^k, x^*) + \|y^k - y^*\| + \|z^k - z^*\| \leq \kappa R(x^k, y^k, z^k) \leq \kappa \|(\hat{a}, b, c)\|.$$

The right term now follows from Theorem 6 for some $c_2 > 0$. Note that the function R is locally Lipschitz continuous with respect to x and globally with respect to (y, z). Therefore, the left term holds for suitable constant $c_1 > 0$. □

The following lemma extends the result [27, Lemma 4.1] from the Euclidean settings to the Riemannian settings, which is on the behavior of the local minimization of augmented Lagrangian function.

Lemma 2 *Let (x^*, y^*, z^*) be a KKT point satisfying M-SOSC and $B \subseteq \mathbb{Y} \times \mathbb{Z}$ be a bounded set. Then there exist $\bar{\rho} > 0$ and $r > 0$ such that for every $\rho \geq \bar{\rho}$ and $(w, p) \in B$, the function $L_\rho(x, w, p)$ has a local minimizer $x = x_\rho(w, p)$ in $B_r(x^*)$ ($B_r(x^*)$ is the closed ball under the Riemannian distance (2.4) on $\mathcal{M}$). Moreover, $x_\rho \to x^*$ uniformly on B as $\rho \to \infty$.*

Proof Since M-SOSC holds, there exists $r > 0$, such that x^* is a strict local solution of the problem in $B_r(x^*)$. For each $\rho > 0$ and $(w, p) \in B$, by the compactness of $B_r(x^*)$, there exists $x = x_\rho(w, p)$ to be the solution of

$$\min_x \ L_\rho(x, w, p) \quad \text{s.t. } x \in B_r(x^*).$$

Assume by contradiction, that x_ρ do not uniformly converge to x^*. Then there exists $\epsilon > 0$, $\rho^k \to \infty$ and $\{w^k, p^k\} \subseteq B$ such that $d(x_{\rho^k}(w^k, p^k), x^*) > \epsilon$ for all k. Since $B_r(x^*)$ is compact, the sequence $x_{\rho^k}(w^k, p^k)$ has an accumulation point $x^* \in B_r(x^*)$. We obtain by $g_2(x^*) \in \mathcal{Q}$ that for all k,

$$\theta^{\rho^k}\left(g_1(x^*) + \frac{w^k}{\rho^k}\right) \leq \theta(g_1(x^*)) + \frac{\|w^k\|^2}{2\rho^k} \quad \text{and} \quad \frac{\rho^k}{2}\operatorname{dist}^2\left(g_2(x^*) + \frac{q^k}{\rho^k}, \mathcal{Q}\right) \leq \frac{\|q^k\|^2}{2\rho^k}.$$

Thus, we have

$$f\left(x^k\right) + \theta^{\rho^k}\left(g_1(x^k) + \frac{w^k}{\rho^k}\right) + \frac{\rho^k}{2}\operatorname{dist}^2(g_2(x^k) + \frac{p^k}{\rho_k}, \mathcal{Q})$$

$$\leq L_{\rho^k}\left(x^*, w^k, p^k\right) \leq f(x^*) + \theta(g_1(x^*)) + \frac{\|w^k\|^2}{2\rho^k} + \frac{\|p^k\|^2}{2\rho^k}. \tag{4.15}$$

Since

$$\theta^{\rho^k}\left(g_1(x^k) + \frac{w^k}{\rho^k}\right) = \theta\left(\operatorname{prox}_{\theta/\rho^k}(g_1(x^k) + \frac{w^k}{\rho^k})\right) + \frac{\rho^k}{2}\|g_1(x^k) + \frac{w^k}{\rho^k} - \operatorname{prox}_{\theta/\rho^k}(g_1(x^k) + \frac{w^k}{\rho^k})\|,$$

by taking $\rho^k \to \infty$, we know from (4.15) that

$$\|g_1(x^k) + \frac{w^k}{\rho^k} - \text{prox}_{\theta/\rho^k}(g_1(x^k) + \frac{w^k}{\rho^k})\| \to 0 \quad \text{and} \quad \text{dist}(g_2(x^k) + \frac{p^k}{\rho_k}, \mathcal{Q}) \to 0.$$

Hence, we have $\lim_{k\to\infty} \text{prox}_{\theta/\rho^k}(g_1(x^k) + \frac{w^k}{\rho^k}) = g_1(x^*)$ and $g_2(x^*) \in \mathcal{Q}$. Moreover, (4.15) also yields $\limsup_{k\to\infty} f(x^k) + \theta\left(\text{prox}_{\theta/\rho^k}(g_1(x^k) + \frac{w^k}{\rho^k})\right) \leq f(x^*) + \theta(g_1(x^*))$. Therefore, $f(x^*) + \theta(g_1(x^*)) \leq f(x^*) + \theta(g_1(x^*))$, which means that $x^* = x^*$ since x^* is the strict solution in $B_r(x^*)$. This contradicts the assumption, and the proof is thus complete. $\square$

Remark 11 The uniform convergence implies the existence of a $\bar{\rho} > 0$ such that $x_\rho(w, p)$ lie in the interior of $B_r(x^*)$, $(w, p) \in B$. Now taking an appropriate small $\epsilon_\rho > 0$, by the continuity of $L_\rho(x, w, p)$ and $\text{grad}_x L_\rho(x, w, p)$ for x, inequality (4.4) and $L_\rho(\tilde{x}_\rho, w, p) \leq L_\rho(x_\rho, w, p) + \epsilon_\rho$ hold simultaneously for any $\tilde{x}_\rho$ sufficiently close to x_ρ. Then, we obtain that $\tilde{x}_\rho \to x^*$ if $\epsilon_\rho \to 0$ as $\rho \to \infty$.

Next, we establish the local convergence result for the inexact Riemannian ALM (Algorithm 1), which is inspired by [27, Theorem 4.2]. The proof can be applied to the linear convergence of exact ALM when we take each $\epsilon_k = 0$.

Theorem 8 *Let (x^*, y^*, z^*) be a KKT point satisfying M-SOSC and M-SRCQ and assume that, for k sufficiently large, x^{k+1} is one of the approximate minimizers from Remark 11. Then there exists $\bar{\rho} > 0$ such that, if $\rho^k \geq \bar{\rho}$ for sufficiently large k, then $(x^k, y^k, z^k) \to (x^*, y^*, z^*)$. If, in addition, $\epsilon_k = o(R(x^k, y^k, z^k))$ and $(w^k, p^k) = (y^k, z^k)$ for sufficiently large k, then there exists $c > 0$ such that*

$$d(x^{k+1}, x^*) + \|y^{k+1} - y^*\| + \|z^{k+1} - z^*\| \leq \frac{c}{\rho^k}\left(d(x^k, x^*) + \|y^k - y^*\| + \|z^k - z^*\|\right) \quad (4.16)$$

for all k sufficiently large. Moreover, $\{\rho^k\}$ remains bounded.

Proof By choosing $\bar{\rho} > 0$ sufficiently large such that for any $\rho^k \geq \bar{\rho}$, x^k lies in a neighborhood of x^* where (4.14) holds. Therefore, the convergence $R_k := R(x^k, y^k, z^k) \to 0$ implies that $(x^k, y^k, z^k) \to (x^*, y^*, z^*)$. From the the definition of (x^k, y^k, z^k), we know that

$$\text{grad}_x L(x^k, y^k, z^k) = \text{grad}_x L_{\rho^k}(x^k, w^k, p^k).$$

Now let $q^{k+1} := \text{prox}_{\theta/\rho^k}(g_1(x^{k+1}) + w^k/\rho^k)$ and $s^{k+1} := \Pi_{\mathcal{Q}}(g_2(x^{k+1}) + p^k/\rho^k)$. Then $y^{k+1} \in \partial\theta(q^{k+1})$ and $z^{k+1} \in \mathcal{N}_{\mathcal{Q}}(s^{k+1})$, which imply that $q^{k+1} = \text{prox}_{\theta/\rho^k}(q^{k+1} + y^{k+1})$ and $s^{k+1} = \Pi_{\mathcal{Q}}(s^{k+1} + z^{k+1})$. Therefore, we have

$$\left\|g_1(x^{k+1}) - \text{prox}_{\theta/\rho^k}(g_1(x^{k+1}) + y^{k+1})\right\|$$
$$= \left\|g_1(x^{k+1}) - \text{prox}_{\theta/\rho^k}(g_1(x^{k+1}) + y^{k+1})\right\| - \left\|q^{k+1} - \text{prox}_{\theta/\rho^k}(q^{k+1} + y^{k+1})\right\|$$
$$\leq \left\|g_1(x^{k+1}) - \text{prox}_{\theta/\rho^k}(g_1(x^{k+1}) + y^{k+1}) - q^{k+1} + \text{prox}_{\theta/\rho^k}(q^{k+1} + y^{k+1})\right\|$$
$$\leq \left\|g_1(x^{k+1}) - q^{k+1}\right\|. \quad (4.17)$$

The last inequality is obtained by using Moreau decomposition and the Lipschitz property of the proximal operator. Similarly, we have

$$\left\|g_2(x^{k+1}) - \Pi_{\mathcal{Q}}(g_2(x^{k+1}) + z^{k+1})\right\| \leq \left\|g_2(x^{k+1}) - s^{k+1}\right\|. \quad (4.18)$$

When $\{\rho_k\}$ is bounded, then by (4.8), $\left\|g_1\left(x^{k+1}\right) - q^{k+1}\right\|$ and $\left\|g_2\left(x^{k+1}\right) - s^{k+1}\right\|$ both converge to zero. If $\rho^k \to \infty$, by Lemma 2 and Remark 11, we have $x^{k+1} \to x^*$,

$$\left\|g_1\left(x^{k+1}\right) - q^{k+1}\right\| = \left\|g_1(x^{k+1}) - \mathrm{prox}_{\theta/\rho^k}\left(g_1(x^{k+1}) + \frac{w^k}{\rho^k}\right)\right\| \to 0$$

and

$$\left\|g_2\left(x^{k+1}\right) - s^{k+1}\right\| \leq \left\|s^{k+1} - \Pi_{\mathcal{Q}}\left(g_2\left(x^{k+1}\right)\right)\right\| + \mathrm{dist}\left(g_2\left(x^{k+1}\right), \mathcal{Q}\right) \to 0.$$

Therefore, we know that

$$R_{k+1} \leq \epsilon_{k+1} + \left\|g_1\left(x^{k+1}\right) - \mathrm{prox}_{\theta/\rho^k}(g_1(x^{k+1}) + y^{k+1})\right\| + \left\|g_2\left(x^{k+1}\right) - \Pi_{\mathcal{Q}}(g_2(x^{k+1}) + z^{k+1})\right\| \to 0,$$
$$(4.19)$$

which means that $\left(x^k, y^k, z^k\right) \to \left(x^*, y^*, z^*\right)$.

Now we are considering the convergence rate. Assume that for sufficiently large k, $(w^k, p^k) = (y^k, z^k)$ and $\epsilon_k = o(R(x^k, y^k, z^k))$. Using (4.19) and the definition of (y^{k+1}, z^{k+1}), we have

$$R_{k+1} \leq \epsilon_{k+1} + \left\|g_1\left(x^{k+1}\right) - q^{k+1}\right\| + \left\|g_2\left(x^{k+1}\right) - s^{k+1}\right\|$$
$$= \epsilon_{k+1} + \frac{\left\|y^{k+1} - y^k\right\|}{\rho^k} + \frac{\left\|z^{k+1} - z^k\right\|}{\rho^k}$$
$$\leq \epsilon_{k+1} + \frac{1}{\rho^k}\left(\left\|y^{k+1} - y^*\right\| + \left\|y^k - y^*\right\| + \left\|z^{k+1} - z^*\right\| + \left\|z^k - z^*\right\|\right). \qquad (4.20)$$

(4.14) implies that there exists $c_1 > 0$, such that $\left\|y^k - y^*\right\| + \left\|z^k - z^*\right\| \leq c_1 R_k$. Since $\epsilon_k = o(R(x^k, p^k, y^k, z^k))$, there exists $\alpha < \frac{c_1}{3\bar{\rho}}$ and $\delta > 0$, such that

$$\epsilon_k \leq \alpha R(x^k, y^k, z^k) \quad \forall\,(x^k, y^k, z^k) \text{ satisfying } d((x^k, y^k, z^k), (x^*, y^*, z^*)) \leq \delta.$$

It follows that $R_{k+1} \leq \frac{c_1}{\rho^k}\left(R_{k+1} + R_k\right) + \alpha R_{k+1}$. Now increase $\bar{\rho}$ until $1 - \frac{c_1}{\rho^k} - \alpha > 1/2$, then $\left(1 - \frac{c_1}{\rho^k} - \alpha\right) R_{k+1} \leq \frac{c_1}{\rho^k} R_k$ implies that $R_{k+1} < \left(2c_1/\rho^k\right) R_k$. The linear convergence is now obtained by (4.14).

Let $V_{k+1} := V\left(x^{k+1}, y^{k+1}, z^{k+1}, \rho^k\right)$. The boundness of $\{\rho^k\}$ can be obtained if $V_{k+1} \leq \tau V_k$ holds for k sufficiently large. By the equality in (4.20), we have $2V_{k+1} \geq R_{k+1} - \alpha R_k$ and

$$V_{k+1} = \max\left(\frac{\left\|y^{k+1} - y^k\right\|}{\rho^k}, \frac{\left\|z^{k+1} - z^k\right\|}{\rho^k}\right)$$
$$\leq \frac{1}{\rho^k}\left(\left\|y^{k+1} - y^*\right\| + \left\|y^k - y^*\right\| + \left\|z^{k+1} - z^*\right\| + \left\|z^k - z^*\right\|\right) \leq \frac{c_1}{\rho^k}\left(R_{k+1} + R_k\right).$$

Thus, these inequalities yields

$$\frac{V_{k+1}}{V_k} \leq \frac{2c_1\left(R_{k+1} + R_k\right)}{\rho^k\left(R_k - \alpha R_{k-1}\right)} \leq \frac{2c_1}{\rho^k}\left(1 + \frac{(2c_1/\rho_k)R_k + \alpha R_{k-1}}{R_k - \alpha R_{k-1}}\right).$$

When $\rho^k \to \infty$, it can be seen that $V_{k+1}/V_k \to 0$, which implies that $\{\rho^k\}$ is bounded. $\qquad \square$

Remark 12 If we require $\{(y^k, z^k)\}$ to be bounded, then B can be defined as a bounded set containing $\{(y^k, z^k)\}$ for all k. This means that the latter condition in Theorem 8, which is $(w^k, p^k) = (y^k, z^k)$ for sufficiently large k, can be fulfilled if $\{(y^k, z^k)\}$ is bounded. In this case we simply choose (w^k, p^k) to be (y^k, z^k).

Remark 13 The convergence rates given in (4.16) depend on the constant c and the penalty parameters ρ_k, where c is determined by the problem and ρ_k can be chosen dynamically. It is shown that the fast linear rate can be achieved if we increase the penalty parameters ρ_k, and the rates become asymptotically superlinear if $\rho_k \to \infty$.

Remark 14 The well behavior of Riemannian ALM solving (1.2) can now be explained, since the M-SOSC condition is satisfied at x^* and M-SRCQ holds at x^* with respect to (y^*, z^*), and by Theorem 8 the iteration sequence converges linearly for k sufficiently large.

5 Applications and numerical experiments

It is known by the two-side error bound (4.14) that the linear convergence of the KKT residual $R(x, y, z)$ can imply the linear convergence of the iteration residual. In this part, we will use $R(x, y, z)$ to illustrate the convergence rate for two reasons. The first is that it is difficult to obtain the local solution and its corresponding multipliers for randomly generated cases. Another reason is that the distance functions for most manifolds are unknown, while using the Euclidean distance instead is not convincing. In the following experiments, the ALM penalty parameters ρ_k are tuned adaptively to guarantee smooth implementation. 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 memory.

5.1 Nonsmooth optimization on sphere

Set $g_1(x) = x$, $\theta(\cdot) = \mu \| \cdot \|_1$ and $\mathcal{M}$ to be the sphere $\mathcal{S}^{n-1} := \{x \in \mathbb{R}^n \mid x^T x = 1\}$. Consider the following nonsmooth optimizations on the unit sphere

$$\min_{x \in \mathcal{M} = \mathcal{S}^{n-1}} f(x) + \mu \|x\|_1, \tag{5.1}$$

where $f : \mathcal{S}^{n-1} \to \mathbb{R}$ is a smooth function. It is well-known that tangent space and normal space for $\mathcal{S}^{n-1}$ at a point x are given by $T_x \mathcal{M} = \{\xi \mid x^\top \xi = 0\}$ and $N_x \mathcal{M} = \{ax \mid a \in \mathbb{R}\}$, respectively. The projection onto the tangent space is $\Pi_x(z) = z - x x^\top z$ and $\operatorname{grad} f(x) = \Pi_x(\nabla f(x))$. Follows from $g_1(x) = x$, $Dg_1(x)T_x \mathcal{M} = T_x \mathcal{M}$ and $Dg_1(x)^* y = \Pi_x(y)$ for any $y \in \mathbb{R}^n$. Since $\operatorname{dom} \theta = \mathbb{R}^n$, it can be checked directly that the M-RCQ condition (3.47) with respect to (5.1) is satisfied at any feasible point of (5.1).[1] Furthermore, the M-SRCQ condition (3.48) with respect to (5.1) is said to hold at a stationary point x with a multiplier y if

$$T_x \mathcal{M} + \mathcal{C}_{\theta, g_1}(x, y) = \mathbb{R}^n, \tag{5.2}$$

where

$$\mathcal{C}_{\theta, g_1}(x, y) = \left\{ d \in \mathbb{R}^n \mid \theta^\downarrow(x; d) = \langle d, y \rangle \right\} \text{ with } y \text{ satisfying } \Pi_x(y) = -\Pi_x(\nabla f(x)). \tag{5.3}$$

Since $\theta(\cdot) = \mu \| \cdot \|_1$, we know that

$$\theta^\downarrow(x; d) = \mu \sum_{x_i = 0} |d_i| + \mu \sum_{x_i > 0} d_i - \mu \sum_{x_i < 0} d_i.$$

---

[1] The corresponding M-RCQ condition (3.47) also holds at any feasible point of (5.1) where the manifold $\mathcal{M}$ in (5.1) is replaced by the Stiefel manifold $\operatorname{St}(n, p) = \{X \in \mathbb{R}^{n \times p} \mid X^\top X = I_p\}$, e.g., the sparse principal component analysis (SPCA) problem (see [25, 48]).

Since the M-RCQ condition holds at any feasible point of (5.1), it follows from Theorem 1 that if x^* is a local optimal solution of (5.1), then x^* must be a stationary point, i.e., there exists $y \in \mathbb{R}^n$ such that

$$\begin{cases} \Pi_{x^*}(\nabla f(x^*) + y) = 0, \\ y \in \mu \partial \|x^*\|_1. \end{cases} \tag{5.4}$$

Moreover, we know from the following proposition that the M-SRCQ condition (3.48) for problem (5.1) is indeed satisfied at any KKT pair (x^*, y^*), which implies that the multiplier is unique.

Proposition 7 *Let x^* be a local optimal solution of (5.1) and y^* be a corresponding multiplier satisfying (5.4). Then, the M-SRCQ condition (5.2) holds for (5.1) at x^* with respect to y^*, which implies the corresponding multiplier y^* is unique.*

Proof In order to show (5.2) holds, we only need to verify that the normal space $N_{x^*}\mathcal{M}$ at x^* satisfies $N_{x^*}\mathcal{M} \subseteq \mathcal{C}_{\theta, g_1}(x^*, y^*)$. Since (x^*, y^*) satisfying (5.4), there exists a such that $y^* + \nabla f(x^*) = ax^*$ and $y^* \in \mu \partial \|x^*\|_1$. Since x^* is orthogonal, it has at least one nonzero component, and we denote by x_i^*. Then a can be determined by $a = \dfrac{y_i^* + \nabla f(x^*)_i}{x_i^*} = \dfrac{\operatorname{sgn}(x_i^*)\mu + \nabla f(x^*)_i}{x_i^*}$. Therefore, y^* is uniquely defined by a. Taking any $d \in N_{x^*}\mathcal{M}$, there exist $\tilde{a}$ such that $d = \tilde{a}x^*$. Moreover, $d_i = 0$ if $x_i^* = 0$. Hence, it can be easily obtained that $\langle y^*, d \rangle = \sum_i y_i^* d_i = \sum_{x_i^* > 0} \mu d_i - \sum_{x_i^* < 0} \mu d_i = \theta^{\downarrow}(x^*, d)$, which implies that $N_{x^*}\mathcal{M} \subseteq \mathcal{C}_{\theta, g_1}(x^*, y^*)$. Thus, the M-SRCQ condition is satisfied at any KKT pair for (5.1). The uniqueness then follows from Theorem 2, directly. □

It is clear that $\operatorname{epi}\theta$ with $\theta = \mu\|\cdot\|_1$ is polyhedral. Thus, θ is C^2-cone reducible, then the M-SOSC condition (3.50) can be applied to problem (5.1). The critical cone at a stationary point x is

$$\begin{aligned} \mathcal{C}(x) &= \{\xi \in T_x\mathcal{M} \mid Df(x)\xi + (\theta \circ g_1)^{\circ}(x; \xi) = 0\} \\ &= \{\xi \in T_x\mathcal{M} \mid \langle \nabla f(x), \xi \rangle + \theta^{\downarrow}(x; \xi) = 0\}. \end{aligned}$$

It is easy to compute that for any $\xi \in T_x\mathcal{M}$,

$$\begin{aligned} \theta_{-}^{\downarrow\downarrow}(g_1(x); Dg_1(x)\xi, w) &= \liminf_{\substack{t\downarrow 0 \\ w' \to w}} \frac{\mu\|x + t\xi + \frac{1}{2}t^2 w'\|_1 - \mu\|x\|_1 - t\theta^{\downarrow}(x, \xi)}{\frac{1}{2}t^2} \\ &= \mu \sum_{x_i=0, \xi_i=0} |w_i| + \mu \sum_{\substack{x_i>0 \text{ or} \\ x_i=0, \xi_i>0}} w_i - \mu \sum_{\substack{x_i<0 \text{ or} \\ x_i=0, \xi_i<0}} w_i. \end{aligned}$$

Therefore, for any multiplier y at x,

$$\begin{aligned} \psi^*_{(g_1(x), Dg_1(x)\xi)}(y) &= \sup_w \left\{ \langle y, w \rangle - \theta_{-}^{\downarrow\downarrow}(g_1(x); Dg_1(x)\xi, w) \right\} \\ &= \sup_w \Big\{ \sum_{x_i=0, \xi_i=0} y_i w_i - \mu|w_i| + \sum_{\substack{x_i>0 \text{ or} \\ x_i=0, \xi_i>0}} y_i w_i - \mu w_i + \sum_{\substack{x_i<0 \text{ or} \\ x_i=0, \xi_i<0}} y_i w_i + \mu w_i \Big\}. \end{aligned}$$

Thus, the M-SOSC condition (3.50) for problem (5.1) at a stationary point x^* with the unique multiper y^*, takes the following form

$$\langle \xi, \operatorname{Hess} f(x^*)\xi \rangle - \psi^*_{(g_1(x^*), Dg_1(x^*)\xi)}(y^*) > 0 \quad \forall \xi \in \mathcal{C}(x^*)\backslash\{0\}. \tag{5.5}$$

Unlike the M-SRCQ condition (5.2), the M-SOSC (5.5) may not hold for some cases. Nevertheless, we could still find some examples that fulfill this condition. Let us consider the following

example of problem (5.1). Set $\mu = 0.25$ and

$$f(x) = -\operatorname{tr}(x^\top A^\top A x) \quad \text{with} \quad A = \begin{bmatrix} 10 & 0 & 0 & 0 & 0 \\ 0 & 25 & 0 & 0 & 0 \\ 0 & 0 & 1.028 & 1.104 & 0 \\ 0 & 0 & 1.104 & 1.672 & 0 \\ 0 & 0 & 0 & 0 & 8 \end{bmatrix}.$$

Then, $x^* = [0, -1, 0, 0, 0]^\top$ is an optimal solution of problem (5.1) with the multiplier $y^* = [0, -\mu, 0, 0, 0]^\top$. The M-SOSC condition is trivially satisfied since the critical cone $\mathcal{C}(x^*) = \{0\}$.

The detail implementation[2] and more numerical results of the Riemannian ALM (Algorithm 1) can be found from [47]. Here, we only use a simple example, in which the matrix $A \in \mathbb{R}^{20 \times 20}$ is randomly generated, to illustrate the numerical performance of the Riemannian ALM (Algorithm 1). It is worth noting that since we do not know the exact solution for this random example, the M-SOSC condition is difficult to verify. However, we still can observe from Figure 2 that the KKT residues converge linearly even though the M-SOSC condition may not be satisfied.
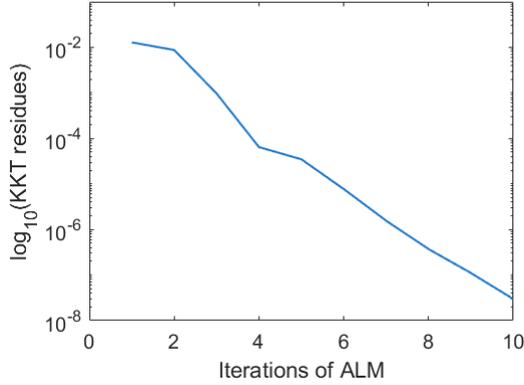

Fig. 2: the KKT residues of problem (5.1) with randomly generated $A \in \mathbb{R}^{20 \times 20}$.

5.2 Robust matrix completion

Consider the robust matrix completion problem proposed in [9]. For a given $A \in \mathbb{R}^{m \times n}$, let $g_1(X) = P_\Omega(X - A)$ 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

$$\begin{aligned} \min_{X \in \mathbb{R}^{m \times n}} & \quad \|P_\Omega(X - A)\|_1 \\ \text{s.t.} & \quad X \in Fr(m, n, r). \end{aligned} \tag{5.6}$$

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

---

[2] Matlab code is available at https://github.com/miskcoo/almssn.

It is known in [44] that the tangent space of $\mathcal{M} = Fr(m, n, r)$ 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

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

Let $P_U = UU^\top$, $P_U^\perp = I - UU^\top$ and $P_V = VV^\top$, $P_V^\perp = I - VV^\top$. The projection to tangent space can be written as

$$\Pi_X(Y) = P_U Y P_V + P_U^\perp Y P_V + P_U Y P_V^\perp.$$

For simplicity, we rewrite (5.6) into the following problem:

$$\begin{aligned} \min_{X,p\in\mathbb{R}^{m\times n}}\ & \|p\|_1 \\ \text{s.t.}\quad & P_\Omega(X - A) - p = 0, \\ & X \in Fr(m, n, r). \end{aligned} \tag{5.7}$$

The Lagrangian of (5.7) can be written as $L(X, p, y) = \|p\|_1 + \langle P_\Omega(X - A) - p, y \rangle$. It is easy to see that the KKT condition is

$$\begin{cases} \Pi_X(P_\Omega(y)) = 0, \\ 0 \in \partial\|p\|_1 - y, \\ P_\Omega(X - A) - p = 0. \end{cases} \tag{5.8}$$

By (3.48), the M-SRCQ is said to hold at a stationary point X with respect to a multiplier z if $Dg_1(X)T_X\mathcal{M} + \mathcal{C}_{\theta,g_1}(X, y) = \mathbb{R}^{m\times n}$. Since $g_1(X) = P_\Omega(X - A)$, we have $Dg_1(X)T_X\mathcal{M} = g_1'(X)T_X\mathcal{M} = P_\Omega(T_X\mathcal{M})$. We can further obtain that $\mathcal{C}_{\theta,g_1}(X, y) = \left\{ d \in \mathbb{R}^{m\times n} \mid \theta^\downarrow(P_\Omega(X - A); d) = \langle d, y \rangle \right\}$, in which

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

Therefore, the M-SRCQ condition for RMC problem at (X, y) satisfying (5.8) is given by

$$P_\Omega(T_X\mathcal{M}) + \mathcal{C}_{\theta,g_1}(X, y) = \mathbb{R}^{m\times n}. \tag{5.9}$$

The M-SOSC condition(3.50) can also be applied to problem (5.7) since $\mathcal{Q} = \{0\}^{m\times n}$ and the epigraph of θ are both polyhedral. It is easy to obtain that the critical cone at a stationary point X is

$$\mathcal{C}(X) = \left\{ \xi \in T_X\mathcal{M} \mid \sum_{P_\Omega(X-A)_{ij}=0} |\xi_{ij}| + \sum_{P_\Omega(X-A)_{ij}>0} \xi_{ij} - \sum_{P_\Omega(X-A)_{ij}<0} \xi_{ij} = 0 \right\}.$$

Since

$$\theta_-^{\downarrow\downarrow}(g(X); Dg(X)\xi, w) = \theta_-^{\downarrow\downarrow}(P_\Omega(X - A); P_\Omega(\xi), w)$$

$$= \sum_{\substack{P_\Omega(X-A)_{ij}=0, \\ P_\Omega(\xi)_{ij}=0}} |w_{ij}| + \sum_{\substack{P_\Omega(X-A)_{ij}>0 \text{ or} \\ P_\Omega(X-A)_{ij}=0, P_\Omega(\xi)_{ij}>0}} w_{ij} - \sum_{\substack{P_\Omega(X-A)_{ij}<0 \text{ or} \\ P_\Omega(X-A)_{ij}=0, P_\Omega(\xi)_{ij}<0}} w_{ij}, \tag{5.10}$$

the M-SOSC condition is said to hold at X if for any $\xi \in \mathcal{C}(X)\backslash\{0\}$,

$$\sup_{y \in M(X,0,0,0)} \left\{ -\psi^*_{(g_1(X),Dg_1(X)\xi)}(y) \right\}$$

$$= \sup_{y \in M(X,0,0,0)} \left\{ -\sup_w \left\{ \langle y, w \rangle - \theta^{\downarrow\downarrow}_-(g(X);Dg(X)\xi,w) \right\} \right\} > 0. \tag{5.11}$$

We first consider a basic example of the problem (5.6), where Ω is the full index set. Let

$$U = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 \\ 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 \end{bmatrix}^T, \quad V = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.6 & -0.8 & 0 & 0 \\ 0 & 0.8 & 0.6 & 0 & 0 \end{bmatrix}^T \quad \text{and} \quad S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$ The observed matrix is

set to $A = A_{\mathrm{ex}} + E_{\mathrm{out}}$, where $A_{\mathrm{ex}} = USV^T$ 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 $r = 3$, $X^* = A_{\mathrm{ex}}$ is a solution of this problem. The M-SOSC condition is satisfied since the critical cone contains only zero. Since $Dg_1(X^*)T_{X^*}F(m,n,r) = T_{X^*}F(m,n,r)$, the M-SRCQ condition holds if there exists a multiplier y^* at X^*, such that $N_{X^*}F(m,n,r) \subset \mathcal{C}_{\theta,g_1}(X^*,y^*)$. For any multiplier y satisfying $Dg_1(X^*)^*y = 0$, we know that $y \in N_{X^*}F(m,n,r)$, and there exists $Y \in \mathbb{R}^{m \times n}$ such that $y = P_U^\perp Y P_V^\perp$. Then for any $d = P_U^\perp \widetilde{Y} P_V^\perp \in N_{X^*}F(m,n,r)$, if $d \in \mathcal{C}_{\theta,g_1}(X^*,y)$, then

$$\sum_{i,j \in \{4,5\}} \mathrm{sgn}(E_{\mathrm{out}}^{ij})\widetilde{Y}_{ij} = \sum_{E_{\mathrm{out}}} \mathrm{sgn}(E_{\mathrm{out}})d_{ij} = \theta^\downarrow(g_1(X^*);d) = \langle d, y \rangle = \langle P_U^\perp \widetilde{Y} P_V^\perp, P_U^\perp Y P_V^\perp \rangle$$

$$= \sum_{i,j \in \{4,5\}} \widetilde{Y}_{ij} Y_{ij}.$$

It is easy to see that there exists a y^* that fulfills the above equalities. Thus, the M-SRCQ condition holds at $X^* = A_{\mathrm{ex}}$ for y^*. In Figure 3, we illustrate the linear convergence rate of the Riemannian ALM (Algorithm 1) in which a semismooth Newton method is employed for solving the subproblem (4.4) (see [47, Algorithm 4.1] for more details).
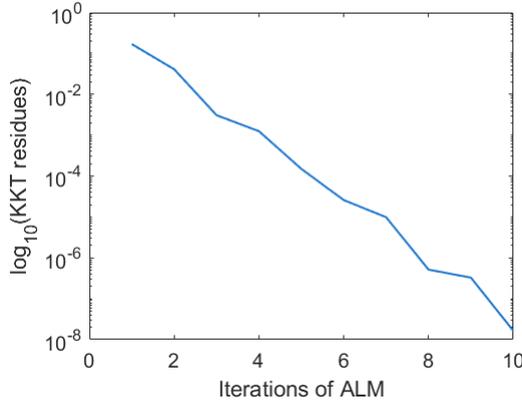

Fig. 3: the KKT residues of RMC problem generated by inexact ALM

Now, look back to the original model, where Ω is no longer to be the full index set. In our next experiment, A is generated as $A = A_{ex} + E_{out}$, where A_{ex} is the ground truth and generated by

$A_{ex} = LR^\top \in \mathbb{R}^{m \times n}$, where $L \in \mathbb{R}^{m \times r}, R \in \mathbb{R}^{n \times r}$ are two random matrices with i.i.d. standard normal distribution. E_{out} is a sparse matrix with 3% of the sample number are nonzero entries satisfying the exponential distribution with mean 10. The sample set is randomly generated on $\{1, 2, \cdots, m\} \times \{1, 2, \cdots, n\}$ with the uniform distribution of $|\Omega|/(mn)$, and the sample size is taken as $\mathrm{OS}(m + n - r)r$, where OS is the oversampling rate for A introduced in [44], and $(m + n - r)r$ is the dimension of $Fr(m, n, r)$.

The stopping criteria is based on the KKT conditions. By (5.8), we terminate our ALM algorithm when the following conditions of the KKT residuals are satisfied

$$\max \left\{ \|\Pi_X(P_\Omega(z))\|_F, \|P_\Omega(X - A) - Y\|_F, \|Y - \mathrm{prox}_{\|\cdot\|_1}(z + Y)\|_F \right\} \leq 1 \times 10^{-7}.$$

In our algorithm, a first-order method is used to find a good initial point and the semismooth Newton method is started, when $\|\mathrm{grad}\, L_{\rho^k}\| \leq 5 \times 10^{-3}$. It must be mentioned that the M-SRCQ and M-SOSC conditions for the random examples are hard to verify and may fail in some cases. Nevertheless, Figure 4 shows the linear convergence rate when applying Riemannian ALM to these randomly generated problems. These results also imply that there may exist some weaker conditions for attaining the local linear convergence rate of Riemannian ALM. The detail

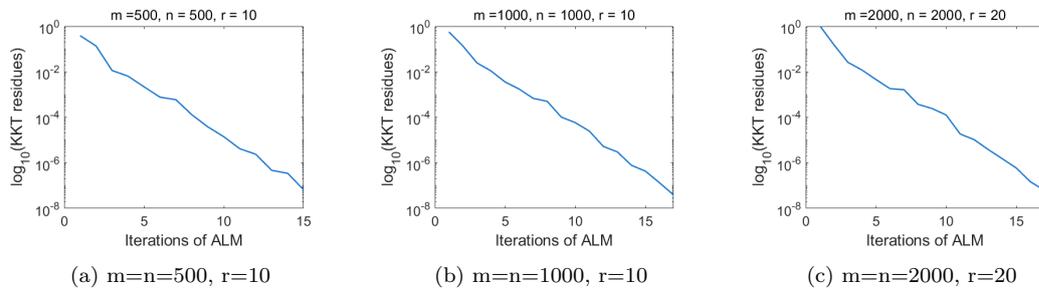(a) m=n=500, r=10   (b) m=n=1000, r=10   (c) m=n=2000, r=20

Fig. 4: The KKT residues of robust matrix completion problems generated by inexact ALM

numerical results are displayed in Table 1. The last column in Table 1 shows an exact recovery of A_{ex}, although we do not know whether it is the global solution of problem (5.6).

Table 1: The performance of Riemannian augmented Lagrangian method for the robust matrix completion problem.

m	n	r	iteration	time(sec)	maximum KKT residual	$\|X - A_{ex}\|$
500	500	10	15	5.24	4.8339e-08	1.6186e-08
1000	1000	10	17	14.87	3.5088e-08	8.3608e-08
2000	2000	20	17	33.47	2.5007e-08	2.4877e-08
5000	5000	20	30	430.02	6.6165e-09	3.4093e-09

6 Conclusion

This paper studies the strict Robinson constraint qualification and the second order condition for the nonsmooth optimization problems on manifolds. We show that the M-SRCQ and M-

SOSC conditions are equivalent to the robust isolated calmness of the KKT solution mapping. Under these two conditions, we show that the iterations generated by Riemannian augmented Lagrangian method converge to a KKT point, and the rate of convergence is linear. Numerical results on a class of nonsmooth optimizations over sphere and the robust matrix completion problems demonstrate the convergence rate, respectively.

The work done in this paper on the perturbation properties of nonsmooth manifold optimization problem (1.1) and the local convergence analysis of the Riemannian augmented Lagrangian method is by no means complete. Due to the rapid advances in the applications of nonsmooth optimization problems on manifolds in different fields, the study on manifold optimization problems will become even more important, and many other properties of perturbation analysis and algorithm design are waiting to be explored. For example, consider the sparse principal component analysis (SPCA) problem (cf. e.g., [25, 48])

$$
\begin{aligned}
&\min_{X \in \mathbb{R}^{n \times p}} \ -\mathrm{tr}\left(X^\top A^\top A X\right) + \mu\|X\|_1 \\
&\text{s.t.} \qquad X^\top X = I_p,
\end{aligned}
\tag{6.1}
$$

where $\mathrm{St}(n, p) = \{X \in \mathbb{R}^{n \times p} \mid X^\top X = I_p\}$ and $f : \mathrm{St}(n, p) \to \mathbb{R}$ is a smooth function. It is worth noting that unlike the nonsmooth optimization problem on sphere (5.1), the local optimal solutions of the SPCA problem (6.1) may not satisfy the M-SRCQ condition if $p > 1$. On the other hand, the numerical experiments given in [47] applying the Riemannian ALM to SPCA problems have shown that the Algorithm 1 performs well numerically and usually can be observed a local fast linear convergence rate (cf. [47]). Therefore, the weaker condition for ensuring the local linear convergence rate of the Riemannian ALM of nonsmooth optimization problems on manifolds is an interesting and important issue for our future work. Another critical and practical issue is designing the efficient algorithms for solving the Riemannian augmented Lagrangian subproblem (4.4) in Algorithm 1.

References

 1. Absil, P.A., Hosseini, S.: A collection of nonsmooth riemannian optimization problems. In: Nonsmooth Optimization and Its Applications, pp. 1–15. Springer (2019)
 2. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization algorithms on matrix manifolds. Princeton University Press (2009)
 3. Bergmann, R., Herzog, R.: Intrinsic formulation of kkt conditions and constraint qualifications on smooth manifolds. SIAM Journal on Optimization **29**(4), 2423–2444 (2019)
 4. Bertsekas, D.P.: Constrained optimization and lagrange multiplier methods. Computer Science and Applied Mathematics (1982)
 5. Bonnans, J.F.: Local analysis of newton-type methods for variational inequalities and nonlinear programming. Applied Mathematics and Optimization **29**(2), 161–186 (1994)
 6. Bonnans, J.F., Cominetti, R., Shapiro, A.: Second order optimality conditions based on parabolic second order tangent sets. SIAM Journal on Optimization **9**(2), 466–492 (1999)
 7. Bonnans, J.F., Shapiro, A.: Perturbation analysis of optimization problems. Springer Science & Business Media (2013)
 8. Boumal, N.: An introduction to optimization on smooth manifolds. Available online, May **3** (2020)
 9. Cambier, L., Absil, P.A.: Robust low-rank matrix completion by riemannian optimization. SIAM Journal on Scientific Computing **38**(5), S440–S460 (2016)
10. Chen, S., Deng, Z., Ma, S., So, A.M.C.: Manifold proximal point algorithms for dual principal component pursuit and orthogonal dictionary learning. IEEE Transactions on Signal Processing **69**, 4759–4773 (2021)
11. Chen, S., Ma, S., Man-Cho So, A., Zhang, T.: Proximal gradient method for nonsmooth optimization over the stiefel manifold. SIAM Journal on Optimization **30**(1), 210–239 (2020)
12. Clarke, F.H.: Optimization and nonsmooth analysis. SIAM (1990)
13. Conn, A.R., Gould, N.I., Toint, P.L.: Trust region methods. SIAM (2000)
14. Cui, Y., Sun, D.: A complete characterization on the robust isolated calmness of the nuclear norm regularized convex optimization problems. arXiv preprint arXiv:1702.05914 (2017)

15. Ding, C., Sun, D., Zhang, L.: Characterization of the robust isolated calmness for a class of conic programming problems. SIAM Journal on Optimization **27**(1), 67–90 (2017)
16. Dontchev, A.L., Rockafellar, R.T.: Implicit functions and solution mappings, vol. 543. Springer (2009)
17. Ferreira, O., Oliveira, P.: Subgradient algorithm on riemannian manifolds. Journal of Optimization Theory and Applications **97**(1), 93–104 (1998)
18. Ferreira, O., Oliveira, P.: Proximal point algorithm on riemannian manifolds. Optimization **51**(2), 257–270 (2002)
19. Grohs, P., Hosseini, S.: ε-subgradient algorithms for locally lipschitz functions on riemannian manifolds. Advances in Computational Mathematics **42**(2), 333–360 (2016)
20. Hestenes, M.R.: Multiplier and gradient methods. Journal of optimization theory and applications **4**(5), 303–320 (1969)
21. Hosseini, S., Pouryayevali, M.: Generalized gradients and characterization of epi-lipschitz sets in riemannian manifolds. Nonlinear Analysis: Theory, Methods & Applications **74**(12), 3884–3895 (2011)
22. Hu, J., Liu, X., Wen, Z.W., Yuan, Y.X.: A brief introduction to manifold optimization. Journal of the Operations Research Society of China **8**(2), 199–248 (2020)
23. Huang, W., Wei, K.: An extension of fast iterative shrinkage-thresholding to riemannian optimization for sparse principal component analysis. arXiv preprint arXiv:1909.05485 (2019)
24. Huang, W., Wei, K.: Riemannian proximal gradient methods. Mathematical Programming pp. 1–43 (2021)
25. Jolliffe, I.T., Trendafilov, N.T., Uddin, M.: A modified principal component technique based on the lasso. Journal of computational and Graphical Statistics **12**(3), 531–547 (2003)
26. Kangkang, D., Zheng, P.: An inexact augmented lagrangian method for nonsmooth optimization on riemannian manifold. arXiv preprint arXiv:1911.09900 (2019)
27. Kanzow, C., Steck, D.: Improved local convergence results for augmented lagrangian methods in $\mathcal{C}^2$-cone reducible constrained optimization. Mathematical Programming **177**(1), 425–438 (2019)
28. Klingenberg, W.P.: Riemannian geometry, vol. 1. Walter de Gruyter (2011)
29. Kovnatsky, A., Glashoff, K., Bronstein, M.M.: Madmm: a generic algorithm for non-smooth optimization on manifolds. In: European Conference on Computer Vision, pp. 680–696. Springer (2016)
30. Lai, R., Osher, S.: A splitting method for orthogonality constrained problems. Journal of Scientific Computing **58**(2), 431–449 (2014)
31. Lee, J.M.: Smooth manifolds. In: Introduction to Smooth Manifolds, pp. 1–31. Springer (2013)
32. Lee, P.Y., et al.: Geometric optimization for computer vision. Citeseer (2005)
33. Liu, Y.J., Zhang, L.W.: Convergence of the augmented lagrangian method for nonlinear optimization problems over second-order cones. Journal of optimization theory and applications **139**(3), 557–575 (2008)
34. Lu, Z., Zhang, Y.: An augmented lagrangian approach for sparse principal component analysis. Mathematical Programming **135**(1), 149–193 (2012)
35. Moreau, J.J.: Proximité et dualité dans un espace hilbertien. Bulletin de la Société mathématique de France **93**, 273–299 (1965)
36. Ozoliņš, V., Lai, R., Caflisch, R., Osher, S.: Compressed modes for variational problems in mathematics and physics. Proceedings of the National Academy of Sciences **110**(46), 18368–18373 (2013)
37. Powell, M.J.: A method for nonlinear constraints in minimization problems. Optimization pp. 283–298 (1969)
38. Riddell, R.: Minimax problems on grassmann manifolds. sums of eigenvalues. Advances in Mathematics **54**(2), 107–199 (1984)
39. Robinson, S.M.: Generalized equations and their solutions, part i: Basic theory. In: Point-to-Set Maps and Mathematical Programming, pp. 128–141. Springer (1979)
40. Rockafellar, R.T.: A dual approach to solving nonlinear programming problems by unconstrained optimization. Mathematical programming **5**(1), 354–373 (1973)
41. Rockafellar, R.T., Wets, R.J.B.: Variational analysis, vol. 317. Springer Science & Business Media (2009)
42. Shapiro, A.: Sensitivity Analysis of Generalized Equations. Journal of Mathematical Sciences **115**(4), 2554–2565 (2003)
43. Sun, D., Sun, J., Zhang, L.: The rate of convergence of the augmented lagrangian method for nonlinear semidefinite programming. Mathematical Programming **114**(2), 349–391 (2008)
44. Vandereycken, B.: Low-rank matrix completion by riemannian optimization. SIAM Journal on Optimization **23**(2), 1214–1236 (2013)
45. Wright, S., Nocedal, J., et al.: Numerical optimization. Springer Science **35**(67-68), 7 (1999)
46. Yang, W.H., Zhang, L.H., Song, R.: Optimality conditions for the nonlinear programming problems on riemannian manifolds. Pacific Journal of Optimization **10**(2), 415–434 (2014)
47. Zhou, Y., Bao, C., Ding, C., Zhu, J.: A semi-smooth newton based augmented lagrangian method for nonsmooth optimization on matrix manifolds. arXiv preprint arXiv:2103.02855 (2021)
48. Zou, H., Hastie, T., Tibshirani, R.: Sparse principal component analysis. Journal of computational and graphical statistics **15**(2), 265–286 (2006)
49. Zowe, J., Kurcyusz, S.: Regularity and stability for the mathematical programming problem in Banach spaces **5**(1), 49–62 (1979)