Filomat 36:11 (2022), 3609–3624
https://doi.org/10.2298/FIL2211609L

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

On the Karush-Kuhn-Tucker Reformulation of the Bilevel Optimization Problems on Riemannian Manifolds

Jiagen Liao[a], Zhongping Wan[a]

[a]*School of Mathematics and Statistics, Wuhan University, Wuhan, 430072, China*

Abstract. Bilevel programming problems are often reformulated using the Karush-Kuhn-Tucker conditions for the lower level problem resulting in a mathematical program with complementarity constraints (MPCC). First, we present *KKT* reformulation of the bilevel optimization problems on Riemannian manifolds. Moreover, we show that global optimal solutions of the MPCC correspond to global optimal solutions of the bilevel problem on the Riemannian manifolds provided the lower level convex problem satisfies the Slater's constraint qualification. But the relationship between the local solutions of the bilevel problem and its corresponding MPCC is incomplete equivalent. We then also show by examples that these correspondences can fail if the Slater's constraint qualification fails to hold at lower-level convex problem. In addition, *M*- and *C*-type optimality conditions for the bilevel problem on Riemannian manifolds are given.

1. Introduction

The aim of this paper is to study the bilevel optimization problem on Riemannian manifolds. Bilevel programming problems are hierarchical optimization problems combining decisions of two decision makers, the so-called leader and the so-called follower.

An approach to investigate bilevel optimization problems on the Euclidean spaces is to replace the lower level problem by its (under certain assumptions necessary and sufficient) optimality conditions, the Karush-Kuhn-Tucker conditions. This replace the bilevel optimization problem by a so-called mathematical program with complementarity conditions (see e.g., [3, 13, 15, 24]). Clearly, both problems are closely related. Dempe et al [14] showed that global and local optimal solutions of the MPCC correspond to global and local optimal solutions of the bilevel problem provided the lower-level problem satisfies the Slater's constraint qualification and also showed by examples that this correspondence can fail if the Slater's constraint qualification fails to hold at lower level problem. Many theoretical results can be found in the monographs by Dempe et al [11, 16] on that topic.

Extending optimization problems from Euclidean spaces to Riemannian manifolds is natural and non-trivial. Some constrained optimization problems can be seen as unconstrained ones from the Riemannian geometry viewpoint. Moreover, some nonconvex optimization problems in the Euclidean setting may become convex introducing an appropriate Riemannian metric. For instance [2, 10].

Such extensions have different advantages. For example, Quiroz et al [26] extended the full convergence of the steepest descent method with a generalized Armijo search and a proximal regularization to solve

2020 *Mathematics Subject Classification*. 26B25; 90C26
Keywords. Bilevel optimization problem; KKT reformulation; Riemannian manifolds; optimality conditions;
Received: 15 September 2020; Revised: 16 August 2021; Accepted: 06 March 2022
Communicated by Predrag Stanimirović
Email address: JiagenLiao@163.com (Jiagen Liao)

minimization problems with quasiconvex objective functions. Previous convergence results are obtained as particular cases and some examples in non-Euclidian spaces are given. In particular, this approach can be used to solve constrained minimization problems with nonconvex objective functions in Euclidian spaces if the set of constraints is a Riemannian manifold and the objective function is quasiconvex in this manifold.

Bento and Nelo [5] presented a subgradient type algorithm for solving convex feasibility problem on Riemannian manifold. The sequence generated by the algorithm converges to a solution of the problem, provided the sectional curvature of the manifold is non-negative. Moreover, assuming a Slater type qualification condition, a variant of the algorithm presented by Bento and Nelo [5], which generates a sequence with finite convergence property, *i.e.*, a feasible point is obtained after a finite number of iterations has been analysed. Wang et al [30, 31] then studied the convergence issue of the subgradient algorithm for solving the convex feasibility problems in Riemannian manifolds, which was first proposed and analysed by Bento and Nelo [5]. The linear convergence property about the subgradient algorithm for solving the convex feasibility problems with the Slater condition in Riemannian manifolds are established. These results extend and/or improve the corresponding known ones in both the Euclidean space and Riemannian manifolds.

Furthermore, Huang et al [19, 20] developed and analysed a generalization of the Broyden class of quasi-Newton methods to the problem of minimizing a smooth objective function f on a Riemannian manifold. A condition on vector transport and retraction that guarantees convergence and facilitates efficient computation is derived. Experimental evidence is presented showing the value of the extension to the Riemannian Broyden class through superior performance for some problems compared to existing Riemannian BFGS methods, in particular those that depend on differentiated retraction.

For more results about algorithms of multicriteria or multiobjective optimization problem on the Riemanmian manifolds, see [4, 6, 7].

Whether similar results can be obtained by extending the bilevel optimization problem from Euclidean space to Riemannian manifold. Bonnel et al [9] dealt with the semivectorial bilevel problem in the Riemannian setting with the upper level is a scalar optimization problem and the lower level is a multiobjective optimization problem acting in a cooperative way inside the greatest coalition and choosing among Pareto solutions with respect to a given ordering cone. The optimality conditions was given for the so-called optimistic problem when the followers choice among their best responses is the most favorable for the leader. Also for the so-called pessimistic problem, when there is no cooperation between the leader and the followers, and the followers choice may be the worst for the leader, a existence result was presented.

Inspired and motivated by the works [3, 12, 14, 15, 18, 23], our interest in the present paper is to give a Karush-Kuhn-Tucker reformulation of the bilevel optimization on Riemannian manifolds. We first research global and local solution of the equivalence between bilevel programming problem with corresponding MPCC under the lower level convex problem satisfying Slater's constraint qualification on the Riemannian manifolds. In general case, we define the M- and C-stationarity conditions for the bilevel problem problem on Riemannian manifolds, and then presented M- and C-optimization conditions for the bilevel problem problem.

Our paper is organized as follows. In Section 2, we present first some basic facts in Riemanmian geometry and then some basic notions and results from the subdifferential calculation on Riemannian manifolds. Section 3 is concerned with the optimization problem on Riemannian manifolds with operator constraint and inequalities constraint, and optimality conditions for these problems are derived.

Section 4 and Section 5 contain the main results using the results from Section 3. In Section 4, we show that the global and local optimal solutions of the MPCC correspond to global and local optimal solutions of the bilevel problem on the Riemannian manifolds under the Slater's constraint qualification. We also show by examples that this correspondence can fail if the Slater's constraint qualification fails to hold at lower-level convex problem. M- and C-type optimality conditions for the bilevel problem problem are given in Section 5.

2. Preliminaries

An m-dimensional Riemannian manifold is a pair (M, g), where M stands for an m-dimensional smooth manifold and g stands for a smooth, symmetric, positive definite $(0, 2)$-tensor field on M, called a Riemannian metric on M. If (M, g) is a Riemannian manifold, then, for any point $p \in M$, the restriction $g_p : T_pM \times T_pM \to \mathbb{R}$ is an inner product on the tangent space T_pM. The tangent bundle TM over M is $TM := \bigcup_{x \in M} T_xM$.

If $\gamma : [a, b] \subset \mathbb{R} \to M$ is a piecewise smooth curve in M, then its length is defined by $L(\gamma) := \int_a^b \|\dot{\gamma}(t)\|_{\gamma(t)} dt$, where $\dot{\gamma}$ means the first derivative of γ with respect to (w.r.t.) t. Let p and q be two points in (M, g) and Γ_{pq} the set of all piecewise smooth curves joining p and q. The function

$$d : M \times M \to \mathbb{R}, \quad d(p, q) := \inf\{L(\gamma) : \gamma \in \Gamma_{pq}\}$$

is a distance on M, and the induced metric topology on M coincides with the topology of M as manifold.

A piecewise smooth curve $\gamma : [a, b] \to M$ is said to be parameterized by arc length if $\|\dot{\gamma}(t)\|_{\gamma(t)}$ is constant on $[a, b]$, and γ is called a geodesic joining the points $\gamma(a)$ and $\gamma(b)$ if for any $t \in [a, b]$, $\nabla_{\dot{\gamma}(t)}\dot{\gamma}(t) = 0$. If moreover the length $L(\gamma)$ is equal to the distance between the points $\gamma(a)$ and $\gamma(b)$, then γ is said to be a minimizing geodesic. A geodesic curve is always parameterized by arc length. Given a point $p \in M$ and a tangent vector $v \in T_pM$, there exists $\varepsilon > 0$ and precisely one geodesic $\gamma_v : [0, \varepsilon] \to M$, depending smoothly on p and v, such that $\gamma_v(0) = p$ and $\dot{\gamma}_v(0) = v$.

The gradient of a differentiable function $f : M \to \mathbb{R}$ w.r.t. the Riemannian metric g is the vector field $\mathrm{grad} f$ defined by $g(\mathrm{grad} f, X) = df(X), \forall X \in TM$, where df denotes the differential of the function f. In local coordinates $(x^1, \ldots, x^n)$ around $p \in M$,and the local components of df are denoted $f_i = \frac{\partial f}{\partial x_i}$, then the local components of $\mathrm{grad} f$ are $f^i = g^{ij} f_j$. Here, g^{ij} are the local components of g^{-1}.

$$\mathrm{grad} f(p) = g^{ij} \frac{\partial f}{\partial x^i} \frac{\partial}{\partial x^i} .$$

Let $p \in M$ be fixed. The inverse of the exponential map $\exp_p^{-1}$ maps diffeomorphically a neighborhood of p onto a neighborhood of the origin of T_pM. Considering an orthonormal basis $(\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n})$ in T_pM with respect to the scalar product $g_p(\cdot, \cdot)$, this diffeomorphism establishes a local coordinate system $(x^1, \ldots, x^n)$ around the point p called normal coordinate system. In this normal coordinates system, the geodesics through p are represented by lines passing through origin. Moreover, the matrix (g_{ij}) associated with the bilinear form g at the point p in this orthonormal basis reduces to the identity matrix and the Christoffel symbols vanish. Thus, for any smooth function $f : M \to \mathbb{R}$, in normal coordinates around p, we obtain

$$\mathrm{grad} f(p) = \sum_i \frac{\partial f}{\partial x^i}(p) \frac{\partial}{\partial x^i} .$$

A function $f : M \to \mathbb{R}$ is said to be convex if its restriction to any geodesic curve $\gamma : [a, b] \to M$ is convex in the classical sense; that the one real variable function $f \circ \gamma : [a, b] \to \mathbb{R}$ is convex.

We now recall some basic notions and results from the subdifferential calculation on Riemannian manifolds, developed by Azagra, Ferrera and López-Mesas [1], Ledyaev and Zhu [23]. Simultaneously, we introduce two subdifferential notions based on the cut-locus, and we establish an analytical characterization of the limiting/Fréchet normal cone on Riemannian manifolds.

Let $f : M \to \mathbb{R}$ be a lower semicontinuous function with $\mathrm{dom}(f) \neq \emptyset$. The Fréchet-subdifferential of f at $p \in \mathrm{dom}(f)$ is the set

$$\partial_F f(p) = \{dh(p) : h \in C^1(M) \text{ and } f - h \text{ attains a local minimum at } p\}.$$

Theorem 2.1. (See [22].) *Let (M, g) be an m-dimensional Riemannian manifold and let $f : M \to \mathbb{R}$ be a lower semicontinuous function, $p \in \mathrm{dom}(f) \neq \emptyset$ and $\xi \in T_p^*M$. The following statements are equivalent:*

(i) $\xi \in \partial_F f(p)$;

(ii) For every chart $\Psi : U_p \subset M \to \mathbb{R}^m$ with $p \in U_p$, if $\zeta = \xi \circ d\Psi^{-1}(\Psi(p))$, we have that

$$\liminf_{v\to 0} \frac{(f\circ\Psi^{-1})(\Psi(p)+v) - f(p) - g(\zeta, v)}{\|v\|} \geq 0;$$

(iii) There exists a chart $\Psi : U_p \subset M \to \mathbb{R}^m$ with $p \in U_p$. If $\zeta = \xi \circ d\Psi^{-1}(\Psi(p))$, then

$$\liminf_{v\to 0} \frac{(f\circ\Psi^{-1})(\Psi(p)+v) - f(p) - g(\zeta, v)}{\|v\|} \geq 0.$$

In addition, if f is locally bounded from below, i.e., for every $q \in M$ there exists a neighborhood U_q of q such that f is bounded from below on U_q, the above conditions are also equivalent to

(iv) There exists a function $h \in C^1(M)$ such that $f - h$ attains a global minimum at p and $\xi = dh(p)$.

The limiting subdifferential and singular subdifferential of f at $p \in M$ are the sets

$$\partial_L f(p) = \left\{\lim_{k\to\infty} \xi_k : \xi_k \in \partial_F f(p_k), (p_k, f(p_k)) \to (p, f(p))\right\},$$

and

$$\partial_\infty f(p) = \left\{\lim_{k\to\infty} t_k\xi_k : \xi_k \in \partial_F f(p_k), (p_k, f(p_k)) \to (p, f(p)), t_k \to 0^+\right\}.$$

Theorem 2.2. (See [22].) *Let (M, g) be an m-dimensional Riemannian manifold and let $f : M \to \mathbb{R}$ be a lower semicontinuous function. Then, we have*

(i) $\partial_F f(p) \subset \partial_L f(p), p \in dom(f)$;

(ii) $0 \in \partial_\infty f(p), p \in M$;

(iii) If $p \in dom(f)$ is a local minimum of f, then $0 \in \partial_F f(p) \subset \partial_L f(p)$.

Theorem 2.3. (See [22].) *Let (M, g) be an m-dimensional Riemannian manifold and let $f_1, \ldots, f_H : M \to \mathbb{R}$ be lower semicontinuous functions. Then, for every $p \in M$ we have either $\partial_L(\sum_{l=1}^H f_l)(p) \subset \sum_{l=1}^H \partial_L f_l(p)$, or there exist $\xi_l^\infty \in \partial_\infty f_l(p), l = 1, \ldots, H$, not all zero such that $\sum_{l=1}^H \xi_{l=0}^\infty$.*

The cut-locus subdifferential of f at $p \in \mathrm{dom}(f)$ is defined as

$$\partial_{cl} f(p) = \left\{\xi \in T_pM : f(q) - f(p) \geq g(\xi, \exp_p^{-1}(q)) \; for\ all\ q \in M\backslash C_p\right\},$$

where C_p is the cut-locus of the point $p \in M$. Note that $M\backslash C_p$ is the maximal open set in M such that every element from it can be joined to p by exactly one minimizing geodesic, see Klingenberg [8] Theorem 2.1.14. Therefore, the cut-locus subdifferential is well-defined, *i.e.*, $\exp^{-1} p(q)$ makes sense and is unique for every $q \in M\backslash C_p$. Let $f : M \to \mathbb{R}$ be a proper, lower semicontinuous function. Then, for every $p \in \mathrm{dom}(f)$ we have $\partial_{cl} f(p) \subset \partial_F f(p) \subset \partial_L f(p)$. Moreover, if f is convex, the above inclusions become equalities.

Let $K \subset M$ be a closed set. Following Ledyaev and Zhu [23], the Fréchet-normal cone and limiting normal cone of K at $p \in K$ are the sets

$$N_F(p; K) = \partial_F \delta_K(p) \quad \text{and} \quad N_L(p; K) = \partial_L \delta_K(p),$$

where δ_K is the indicator function of the set K, *i.e.*, $\delta_K(q) = 0$ if $q \in K$ and $\delta_K(q) = +\infty$ if $q \notin K$.

The following result is one of our key tool to study stationarity points on Riemannian manifolds. Let (M, g) be a Riemannian manifold, $K \subset M$ be a closed, geodesic convex set, and $p \in K$. Then,we have

$$N_F(p; K) = N_L(p; K) = \partial_{cl}\delta K(p) = \left\{\xi \in T_pM : g(\xi, exp_p^{-1}(q)) \leq 0 \; for\ all\ q \in K\right\}.$$

For more details and complete information on the fundamentals in Riemannian geometry, see [17, 21, 29]. For more results about calculus Riemannian manifolds, we refer to [22, 23].

3. Constrained Optimization Problem on Riemannian Manifolds

In this section, we first consider the optimization problem with operator constraint on the m-dimensional Riemannian manifolds (M, g)

$$\min\{f(p) : x \in \Omega \cap \Psi^{-1}(\Lambda)\}, \tag{3.1}$$

where $f : M \to \mathbb{R}$ and $\Psi : M \to \mathbb{R}^m$ are smooth functions, and the sets $\Omega \subseteq M$; $\Lambda \subseteq \mathbb{R}^m$ are closed. Nonetheless, similarly to the optimization problem with operator constraint in the Euclidean Space, a constraint qualification (CQ) is provided in order to derive detailed KKT type dual optimality conditions in terms of problem data. we define the basic CQ at a feasible point p^* of problem (3.1) by

$$\left.\begin{array}{r} 0 \in \lambda^T \partial_L f(p^*; \Lambda) + N_L(p^*; \Omega) \\ \lambda \in N_L(\Psi(p^*); \Lambda) \end{array}\right\} \Rightarrow \lambda = 0. \tag{3.2}$$

The optimization problem with operator constraint (3.2) can turn into an unconstrained optimization problem by using the indicator function of the set $\Omega \cap \Psi^{-1}(\Lambda)$:

$$\min f(p) + \delta_{\Omega\cap\Psi^{-1}(\Lambda)}(p).$$

Applying Theorem 2.2 (*iii*) leads to

$$0 \in \partial_L(f(p^*) + \delta_{\Omega\cap\Psi^{-1}(\Lambda)}(p^*)) = \partial_L f(p^*) + N_L(p^*; \Omega \cap \Psi^{-1}(\Lambda)). \tag{3.3}$$

In the other hand, according to the well-known result which can be found *e.g.* in [25, 27] under the basic CQ with the the Fréchet-normal cone and limiting normal cone of sets $\Omega \cap \Psi^{-1}(\Lambda)$ on the Riemannian manifolds, the following theorem is provided.

Theorem 3.1. *Let $p^* \in \Omega \cap \Psi^{-1}(\Lambda)$ and assume that the basic CQ (3.2) holds at p^*. Then the following inclusion is satisfied:*

$$N_L\Big(p^*; \Omega \cap \Psi^{-1}(\Lambda)\Big) \subset \cup\Big\{\lambda^T \partial_L \Psi(p^*) + N_L(p^*; \Omega) : \lambda \in N_L(\Psi(p^*); \Lambda)\Big\}.$$

Proof. Let $v \in N_F(p^*, \Omega \cap \Psi^{-1}(\Lambda))$. By Theorem 2.1 there's a smooth function h on Riemannian manifold such that $\delta_{\Omega\cap\Psi^{-1}(\Lambda)} - h$ attains a global minimum at p^*, that is $\operatorname{argmax}_{\Omega\cap\Psi^{-1}(\Lambda)} h = \{p^*\}$, $v = dh(p^*)$. We take any sequence of values $\tau^v \to 0$ and analyze for each v the problem of minimizing over $\Omega \times \Lambda$ the C^1-function

$$\varphi^v(p, u) = -h(p) + \frac{1}{2\tau^v}\|\Psi(x) - u\|^2.$$

Through our arrangement that Ω and Λ are compact, the minimum is attained at some point (p^v, u^v) (not necessarily unique). Moreover $(p^v, u^v) \to (p^v, \Psi(p^*))$. The optimality condition of (3.2) gives us

$$0 \in \partial_{L_p}\varphi^v(p^v, u^v) + N_L(p^v, \Omega), \quad 0 \in \partial_{L_u}\varphi^v(p^v, u^v) + N_L(u^v, \Lambda),$$

in as much as $\operatorname{argmin}_{p\in\Omega}\varphi^v(p, u^v) = \{p^v\}$ and $\operatorname{argmin}_{u\in\Lambda}\varphi^v(p^v, u) = \{u^v\}$. W differentiate φ^v in u and differentiate φ^v next in p, respectively. We see that

$$-\partial_{L_u}\varphi^v(p^v, u^v) = y^v = \frac{\Psi(p^v) - u^v}{\tau^v},$$

and

$$-\partial_{L_p}\varphi^v(p^v, u^v) = z^v = dh(p^v) - \frac{\mathcal{J}\Psi(p^v)\Psi(p^v) - u^v}{\tau^v} \text{ with } \mathcal{J}\Psi(p^v) \to \mathcal{J}\Psi(p^*),\ dh(p^v) \to v,$$

where $\mathcal{J}\Psi(p)$ is denoted by $\partial_L\Psi(p^*)$. By passing to subsequences to the limit, we can reduce to having the sequence of vectors $y^\upsilon \in N_L(u^\upsilon, \Lambda)$ either convergent to some y or such that $\lambda^\upsilon y^\upsilon \to y \neq 0$ for a choice of scalars $\lambda^\upsilon \to 0$. In both cases we have $y \in N_L(\Psi(p^*), \Lambda)$, because $N_L(u^\upsilon, \Lambda)$ is a cone and $u^\upsilon \to \Psi(p^*)$.

If $y^\upsilon \to y$, we have at the same time that $z^\upsilon \to z = v - y * \mathcal{J}\Psi(p^*)$ with $z \in N_L(p^\upsilon, \Omega)$. This yields the desired representation $v = y * \mathcal{J}\Psi(p^*) + z$. On the other hand, if $\lambda^\upsilon y^\upsilon \to y \neq 0$, $\lambda^\upsilon \to 0$, we obtain from $v^\upsilon = y^\upsilon * \mathcal{J}\Psi(p^\upsilon) + z^\upsilon$ that $\lambda^\upsilon z^\upsilon \to z = y * \mathcal{J}\Psi(p^*)$, $z \in N_L(p^*, \Omega)$, which produces a representation $0 = y * \mathcal{J}\Psi(p^*) + z$ of the sort forbidden by the constraint qualification. Therefore, only the first case is viable.

This proves that $N_F(p^*, \Omega \cap \Psi^{-1}(\Lambda)) \subset S(p^*)$, where we now denote by $S(p) = \{y * \mathcal{J}\Psi(p) + z | p \in \Omega \cap \Psi^{-1}(\Lambda), y \in N_L(\Psi(p), \Lambda), z \in N_L(p, \Omega)\}$.

Next we'll proved that $N_L(p^*, \Omega \cap \Psi^{-1}(\Lambda)) \subset S(p^*)$, we can therefore use the fact that $N_L(p^*, \Omega \cap \Psi^{-1}(\Lambda)) = \limsup_{p \to p^*} N_F(p^*, \Omega \cap \Psi^{-1}(\Lambda))$ by the definition of Fréchet-normal cone and limiting normal cone on the Riemannian manifolds.

Let $p^\upsilon \to p^*$ and $v^\upsilon \to v$ with $v^\upsilon \in S(p^\upsilon)$, so that $v^\upsilon = y^\upsilon * \mathcal{J}\Psi(p^\upsilon) + z^\upsilon$ with $y^\upsilon \in N_L(\Psi(p^\upsilon), \Lambda)$ and $z^\upsilon \in N_L(p^\upsilon, \Omega)$. We can revert once more to two cases: either $(y^\upsilon, z^\upsilon) \to (y, z)$ or $\lambda^\upsilon(y^\upsilon, z^\upsilon) \to (y, z) \neq (0, 0)$ for some sequence $\lambda^\upsilon \to 0$. In the first case we obtain in the limit that $v = y * \mathcal{J}\Psi(p^*) + z$ with $y \in N_L(\Psi(p^*), \Lambda)$ and $z \in N_L(p^*, \Omega)$, hence $v \in S(p^*)$ as desired. But the second case is impossible, because it would give us $\lambda^\upsilon v^\upsilon = \lambda^\upsilon y^\upsilon * \mathcal{J}\Psi(p^\upsilon) + \lambda^\upsilon z^\upsilon$ and in the limit $0 = y * \mathcal{J}\Psi(p^*) + z$ in contradiction to the constraint qualification. This confirms that $N_L(p^*, \Omega \cap \Psi^{-1}(\Lambda)) \subset S(p^*)$. This completes the proof.

Using the above Theorem 3.1, we are now ready to state a KKT-type optimality condition for optimization problem with operator constraint on the Riemannian manifolds under the basic CQ.

Theorem 3.2. *Let p^* be a local optimal solution of problem (3.1) and $\mathbb{B}$ be an unit ball in R^m, assume that the basic CQ (3.2) holds at p^*. Then, there exists $\theta > 0$ such that for any $\varepsilon \geq \theta$, and one can find $\lambda \in \varepsilon\mathbb{B} \cap N_L(\Psi(p^*); \Lambda)$ such that*

$$0 \in \partial_L f(p^*) + \lambda^T \partial_L \Psi(p^*) + N_L(p^*; \Omega).$$

Next we refer to the convex program on the Riemannian manifolds (M, g)

$$\min \{f(p) : \psi_i(p) \leq 0, i \in I = \{1, 2, \cdots, r\},\ p \in M\}, \tag{3.4}$$

where $f, \psi_i : M \to \mathbb{R}, i \in I$ are convex functions on the Riemannian manifolds (M, g). We denote by $I(p^*)$ the set of indices i having the property that the inequalities which describe active set at p^*,

$$I(p^*) = \{i \in I : \psi_i(x^*) = 0\}.$$

The convex program (3.4) is called primal problem. The Lagrange function attached to the primal problem is defined by

$$L(p, u) = f(p) + u^T\psi(p), p \in M, u \geq 0, \psi(x) = \{\psi_1(p), \psi_2(p), \cdots, \psi_r(p)\}.$$

The dual problem is given by

$$\begin{aligned} \max\ & L(p, u) \\ s.t.\ & p \in M, u \geq 0, \\ & \mathrm{grad} f(p) + u^T \mathrm{grad}\psi(p) = 0. \end{aligned} \tag{3.5}$$

Theorem 3.3. (see [29].) *Suppose that the convex program (3.4) is super-consistent and the functions f and $\psi_i, i \in I$ are of class C^1. If p^* is the optimal solution of the primal problem (3.4), then there exists $u^* \geq 0$ such that (p^*, u^*) is the optimal solution of dual problem (3.5) and $f(p^*) = L(p^*, v^*)$.*

Theorem 3.4. (see [29].) *Suppose that the convex program (3.4) is superconsistent and the functions f and $\psi_i, i \in I$ are of class C^1. If p^* is the optimal solution of the primal problem (3.4), then there exists $\bar{u} \geq 0$ such that*

$$\begin{aligned} \mathrm{grad} f(p^*) + \bar{u}^T \mathrm{grad}\psi(p^*) &= 0, \\ \bar{u}^T\psi(p^*) &= 0. \end{aligned}$$

4. Bilevel Programming on Riemannian Manifolds

Let (M_1, g_1) (the leader decision variables set) and (M_2, g_2) (the follower decision variables set) be two connected Riemannian manifolds of dimension n and m, respectively. Moreover, (M_2, g_2) is supposed to be complete. The corresponding Riemannian metrics will be denoted by $g_1(\cdot,\cdot)$ and $g_2(\cdot,\cdot)$, respectively. Let also $F : M_1 \times M_2 \to \mathbb{R}$ be the leader objective function, $f : M_1 \times M_2 \to \mathbb{R}$ be the follower objective function, and $\psi_i : M_1 \times M_2 \to \mathbb{R}, i \in I = \{1, 2, \cdots, r\}$ be the lower level constraint functions.

Consider the bilevel programming problem (P) on the Riemannian manifolds (M_1, g_1) and (M_2, g_2)

$$\text{minimize} \quad F(x, y) \;\; \text{subject to} \quad x \in K \subset M_1, \;\; y \in S(x), \tag{4.1}$$

where $S(x)$ is the solution set of the following lower level Problem (LLP):

$$\text{minimize} \quad f(x, y) \;\; \text{subject to} \;\; y \in M_2, \psi_i(x, y) \leq 0. \tag{4.2}$$

We consider for each x the function $f(x,\cdot)$ and $\psi_i(x,\cdot), i \in I$ are convex on the complete Riemanmian manifolds (M_2, g_2).

From now on we will assume that the problem (LLP) always has a solution, and consider here the optimistic solution. To introduce the optimistic case, consider the function

$$\varphi_0(x) = \inf\{F(x, y) : y \in S(x), x \in M_1\}.$$

Then the optimistic bilevel problem reads as

$$\min_x \;\; \varphi_0(x).$$

We next assume that the upper and lower level feasible set are given as

$$K := \{x \in M_1 : G(x) \leq 0\} \quad \text{and} \quad K(x) := \{y \in M_2 : \psi(x, y) \leq 0, x \in X\},$$

where $\psi(x, y) := \big(\psi_1(x, y), \ldots, \psi_r(x, y)\big)$, the functions $G : R^n \to \mathbb{R}^k$ and $\psi : R^n \times \mathbb{R}^m \to R^r$ being class of C^1, respectively.

Consider the convex lower level problem, combining Theorem 3.3 and Theorem 3.4, then the Karush-Kuhn-Tucker conditions are satisfied:

$$\text{grad}_y L(x, y, u) = 0, \; u \geq 0, \; u^T\psi(x, y) = 0.$$

Here the function $L(x, y, u) = f(x, y) + u^T\psi(x, y)$ is the Lagrange function, and grad_y denotes the gradient with respect to the variables y only with the Riemannian metric g_2.

This leads to a reformulation of the bilevel optimization problem on the Riemannian manifolds (M_1, g_1) and (M_2, g_2) (MPCC):

$$\begin{aligned} \min \;\; & F(x, y) \\ s.t. \;\; & G(x) \leq 0 \\ & \text{grad}_y L(x, y, u) = 0 \\ & \psi(x, y) \leq 0 \\ & u \geq 0, u^T\psi(x, y) = 0. \end{aligned} \tag{4.3}$$

Next, we will research the equivalency with the Bilevel Programming and MPCC on the Slater's constraint qualification (Slater's CQ).

Slater's CQ : There exists $y^*(x)$ such that $\psi_i(x, y^*(x)) < 0, i \in I$.

Theorem 4.1. *Let $(\bar{x}, \bar{y})$ be a global optimal solution of the bilevel problem on the Riemannian manifolds (M_1, g_1) and (M_2, g_2) and assume that the lower level problem is convex for which Slater's constraint qualification is satisfied at $x = \bar{x}$. Then, for each*

$$\bar{u} = \Lambda(\bar{x}, \bar{y}) := \{u \geq 0 : \mathrm{grad}_y L(\bar{x}, \bar{y}, u) = 0, u^T \psi(\bar{x}, \bar{y}) = 0\},$$

the point $(\bar{x}, \bar{y}, \bar{u})$ ia a global optimal solution of problem (4.3).

Proof. Since the $(\bar{x}, \bar{y})$ is a global optimal solution of the bilevel problem, then $(\bar{x}, \bar{y})$ is the optimal solution of the lower level problem. Moreover, Slater's constraint qualification is satisfied at $x = \bar{x}$. So, there exist $u \geq 0$ such that $\mathrm{grad} f(\bar{x}, y^*(\bar{x})) + u^T \mathrm{grad}\psi(\bar{x}, y^*(\bar{x})$ and $u^T \psi(\bar{x}, y^*(\bar{x}) = 0$. Thus, point $(\bar{x}, \bar{y}, \bar{u})$ ia a global optimal solution of MPCC with $\bar{u} = \Lambda(\bar{x}, \bar{y})$.

The following small example shows that, without regularity of the lower level problem, the Theorem 4.1 is in general not correct. Then, the bilevel programming problem can have a global optimal solution while the corresponding MPCC has no solution.

Example 4.2. *We consider in this subsection the particular case $C = \mathbb{R}^2_+$, $M_1 = \mathbb{R}_{++} :=]0, +\infty[$ with the Euclidean metric, and $M_2 = \mathbb{R}^2_{++} := \{(y_1, y_2) \in \mathbb{R}^2 | y_1 > 0, y_2 > 0\}$ with the metric g_2 given in Cartesian coordinates (y_1, y_2) around the point $y \in M_2$ by the matrix*

$$M_2 \ni y \mapsto (g_{ij})_y = \Big(g_2\Big(\frac{\partial}{\partial y_i}, \frac{\partial}{\partial y_j}\Big)\Big) := \mathrm{diag}\Big(y_1^{-2}, y_2^{-2}\Big).$$

In other words, for any vectors $u = (u_1, u_2)$ and $v = (v_1, v_2)$ in the tangent plane at $y \in M_2$, denoted $T_y M_2$, which coincides with $\mathbb{R}^2$, we have

$$g_2(u, v) = \frac{u_1 v_1}{y_1^2} + \frac{u_2 v_2}{y_1^2}.$$

Let $a = (a_1, a_2) \in M_2$ and $v = (v_1, v_2) \in T_a M_2$. It is easy to see that the (minimizing) geodesic curve $t \mapsto \gamma(t)$ verifying $\gamma(0) = a$, $\gamma(0) = v$ is given by

$$\mathbb{R} \ni t \mapsto (a_1 e^{\frac{v_1}{a_1} t}, a_2 e^{\frac{v_2}{a_2} t}).$$

Hence, M_2 is a complete Riemannian manifold. Also, the (minimizing) geodesic segment $\gamma : [0, 1] \to M_2$ joining the points $a = (a_1, a_2)$ and $b = (b_1, b_2)$, i.e., $\gamma(0) = a$, $\gamma(1) = b$ is given by $\gamma_i(t) = a_1^{1-t} b_i^t$, $i = 1, 2$. Thus, the distance d on the metric space (M_2, g_2) is given by

$$d(a, b) = \int_0^1 \|\dot{\gamma}(t)\|_{\gamma(t)} dt = \int_0^1 \sqrt{(\frac{\dot{\gamma}_1(t)}{\gamma_1(t)})^2 + (\frac{\dot{\gamma}_2(t)}{\gamma_2(t)})^2} dt$$
$$= \sqrt{(\ln \frac{a_1}{b_1})^2 + (\ln \frac{a_2}{b_2})^2}.$$

It follows easily that the closed ball $\mathbb{B}(a; R)$ centered in $a \in M_2$ of radius $R \geq 0$ verifies

$$\Big[a_1 e^{-\frac{R}{\sqrt{2}}}, a_1 e^{-\frac{R}{\sqrt{2}}}\Big] \times \Big[a_2 e^{-\frac{R}{\sqrt{2}}}, a_2 e^{-\frac{R}{\sqrt{2}}}\Big] \subset \mathbb{B}(a; R),$$

thus every closed rectangle $[\rho_1, \eta_1] \times [\rho_2, \eta_2]$ $(\rho_1 > 0, \rho_2 > 0)$ is bounded in the metric space (M_2, g_2) with the distance d.

Consider now the functions $F : M_1 \times M_2 \to \mathbb{R}$, $f : M_1 \times M_2 \to \mathbb{R}$ and $g : M_1 \times M_2 \to \mathbb{R}$ given for any $(x, y) \in M_1 \times M_2$ by

$$F(x,y)=\frac{1}{x}\ln(y_1)+(x-1)^2-\frac{1}{\sqrt{y_2}}(1+y_2)\ln(y_1),$$
$$f(x,y)=\ln^2(y_1)+x\sqrt{y_2},$$
$$\psi(x,y)=x\ln(y_1)-\ln(y_2).$$

It is easy to see that for each fixed $x\in M_1$, non of the functions $f(x,\cdot)$ and $g(x,\cdot)$ is convex on M_2 with the Euclidean metric. And for any geodesic segment $\gamma:[0,1]\to M_2$ with $\gamma(0)=a,\gamma(1)=b$, the functions $f(x,\cdot)\circ\gamma:[0,1]\to\mathbb{R}$ and $g(x,\cdot)\circ\gamma:[0,1]\to\mathbb{R}$ are convex. Hence, the functions $f(x,\cdot)$ and $g(x,\cdot)$ are convex on the Riemannian manifold (M_2,g_2).

We now consider the convex lower level problem on the Riemannian manifold (M_2,g_2)

$$\min_{y_1,y_2}\{\ln^2(y_1)+x\sqrt{y_2}:x\ln(y_1)-\ln(y_2)\le 0\}.$$

Then, for $x=1$, Slater's condition is violated. Consider the bilevel programming problem

$$\min_{x,y}\{\frac{1}{x}\ln(y_1)+(x-1)^2-\frac{1}{\sqrt{y_2}}(1+y_2)\ln(y_1):x>0,y\in\Psi(x)\},$$

where $\Psi(x)$ is the solution set mapping of convex lower level problem. Then the unique (global) optimal solution of the bilevel programming problem is $x=1,y=(1,1)$ and there does not exist local optimal solutions.

We then consider the corresponding MPCC

$$\begin{array}{ll}\min\limits_{x,y} & F(x,y)=\frac{1}{x}\ln(y_1)+(x-1)^2-\frac{1}{\sqrt{y_2}}(1+y_2)\ln(y_1)\\ s.t. & x>0,y>0\\ & \mathrm{grad}_y f(x,y)+u\,\mathrm{grad}_y\psi(x,y)=0\\ & \psi(x,y)\le 0\\ & u\psi(x,y)=0\\ & u\ge 0.\end{array}$$

By the definition of the gradient of a differentiable function w.r.t. the Riemannian metric g_2, we so have

$$\begin{aligned}\mathrm{grad}_y f(x,y)+u\,\mathrm{grad}_y g(x,y)&=y_1^2\frac{\partial f}{\partial y_1}(x,y)\frac{\partial}{\partial y_1}+y_2^2\frac{\partial f}{\partial y_2}(x,y)\frac{\partial}{\partial y_2}+u\left(y_1^2\frac{\partial g}{\partial y_1}(x,y)\frac{\partial}{\partial y_1}+y_2^2\frac{\partial g}{\partial y_2}(x,y)\frac{\partial}{\partial y_2}\right)\\&=(2y_1\ln(y_1)+uxy_1)\frac{\partial}{\partial y_1}+\left(\frac{x}{2}y_2^{\frac{3}{2}}-uy_2\right)\frac{\partial}{\partial y_2}\\&=0.\end{aligned}$$

Then, we need solve the following systems of equations

$$\begin{aligned}2\ln(y_1)+ux&=0,\\ \frac{x}{2}y_2^{\frac{1}{2}}-u&=0,\\ u\left(x\ln(y_1)-\ln(y_2)\right)&=0.\end{aligned}\tag{4.4}$$

In the fact that $x>0$ and $y>0$, so the above systems (4.4) has none solution. Therefore, there does not exist a corresponding global optimal solution of the MPCC.

The opposite implication of Theorem 3.3 is also true under a very mild assumption. We also give the proof of this result.

Theorem 4.3. *Let $(\bar{x}, \bar{y}, \bar{u})$ be a global optimal solution of problem (4.3), let the lower level problem (4.2) be convex on the Riemannian manifolds and assume that Slater's constraint qualification is satisfied for the lower level problem for each $x \in X$. Then, $(\bar{x}, \bar{y})$ is a global optimal solution of the bilevel programming problem on the Riemannian manifolds.*

Proof. If $(\bar{x}, \bar{y}, \bar{u})$ is a global optimal solution of (4.3) then $\Lambda(\bar{x}, \bar{y}) \neq \emptyset$ and, since the objective function value of problem (4.3) is independent of $u \in \Lambda(\bar{x}, \bar{y})$, each solution $(\bar{x}, \bar{y}, \bar{u}), u \in \Lambda(\bar{x}, \bar{y})$ is a global optimal solution, too. Assume now, that $(\bar{x}, \bar{y})$ is not a global optimal solution of the bilevel programming problem. Hence there exists (x, y) with $x \in X$ and $y \in \Psi(x)$ such that $F(x, y) < F(\bar{x}, \bar{y})$. Since $y \in \Psi(x)$ and the Slater's constraint qualification holds at x the *KKT* conditions hold and thus there exists $u \in \mathbb{R}^r_+$ such that

$$\begin{aligned} \operatorname{grad}_y f(x, y) + \sum_{i=1}^{r} \operatorname{grad}_y \psi_i(x, y) &= 0, \\ u^T \psi(x, y) &= 0, \\ \psi_i(x, y) &\leq 0. \end{aligned}$$

This clearly shows that (x, y, u) is a feasible solution of the problem (4.3). This fact combined with the fact that $F(x, y) < F(\bar{x}, \bar{y})$ shows that $(\bar{x}, \bar{y}, \bar{u})$ is not a global optimal point of (4.3). This is a contradiction. Hence the result.

Example 4.4. *Consider the following particular case $C = \mathbb{R}^2_+$, $M_1 = \mathbb{R}_+ := [0, +\infty)$ with the Euclidean metric, and $M_2 = \mathbb{R}^2_+ := \{(y_1, y_2) \in \mathbb{R}^2 | y_1 \geq 0, y_2 \geq 0\}$ with the metric g_2 given in Cartesian coordinates (y_1, y_2) around the point $y \in M_2$ by the matrix*

$$M_2 \ni y \mapsto (g_{ij})_y = \left(g_2\Big(\frac{\partial}{\partial y_i}, \frac{\partial}{\partial y_j}\Big)\right) := \operatorname{diag}(2y_1, 2y_2).$$

Let $a = (a_1, a_2) \in M_2$ and $v = (v_1, v_2) \in T_a M_2$. It is easy to see that the (minimizing) geodesic curve $t \mapsto \gamma(t)$ verifying $\gamma(0) = a$, $\gamma(0) = v$ is given by

$$\mathbb{R} \ni t \mapsto \left(a_1 e^{\frac{v_1}{a_1} t}, a_2 e^{\frac{v_2}{a_2} t} \right).$$

Hence, M_2 is a complete Riemannian manifold. Also, the (minimizing) geodesic segment $\gamma : [0, 1] \to M_2$ joining the points $a = (a_1, a_2)$ and $b = (b_1, b_2)$, i.e., $\gamma(0) = a$, $\gamma(1) = b$ is given by $\gamma_i(t) = a_1^{1-t} b_i^t$, $i = 1, 2$.

Consider now the functions $F : M_1 \times M_2 \to \mathbb{R}$, $f : M_1 \times M_2 \to \mathbb{R}$ and $g : M_1 \times M_2 \to \mathbb{R}$ given for any $(x, y) \in M_1 \times M_2$ by

$$\begin{aligned} &F(x, y) = (x - 1)^2 + y_1^2 + y_2^2, \\ &f(x, y) = x^2 y_1^2 + \sqrt{x} y_2^2, \\ &\psi(x, y) = y_1^4 + 2y_2^3. \end{aligned}$$

We now consider the bilevel programming problem on the Riemannian manifold (M_2, g_2)

$$\min_{x,y} \left\{ (x - 1)^2 + y_1^2 + y_2^2 : x \in M_1, y = (y_1, y_2) \in \Psi(x) \right\}.$$

where $\Psi(x)$ is the solution set mapping of the following lower level problem

$$\min_{x,y} \left\{ x^2 y_1^2 + \sqrt{x} y_2^2 : y_1^4 + 2y_2^3 \leq 0 \right\}.$$

It is easily show that the unique global optimal solution of the bilevel programming problem is $x = 1, y = (0, 0)$. Further it is simple to get that $x = 0, y = (0, 0)$ is the only point where the KKT condition are satisfied. And the corresponding MPCC is present as follows

$$\begin{aligned} \min_{x,y} \quad & F(x, y) = (x-1)^2 + y_1^2 + y_2^2 \\ s.t. \quad & x \geq 0, y \geq 0, \\ & \mathrm{grad}_y f(x, y) + u\,\mathrm{grad}_y g(x, y) = 0, \\ & g(x, y) \leq 0, \\ & ug(x, y) = 0, \\ & u \geq 0. \end{aligned}$$

According to the gradient of the functions $f(x, \cdot)$ and $g(x, \cdot)$ on the Riemannian manifold (M_2, g_2), we so have

$$\mathrm{grad}_y f(x, y) + u\,\mathrm{grad}_y g(x, y) = (x^2 + 2uy_1^2)\frac{\partial}{\partial y_1} + (\sqrt{x} + 3y_2)\frac{\partial}{\partial y_2} = 0.$$

Then, we need solve the following systems

$$\begin{aligned} x^2 + 2uy_1^2 &= 0, \\ \sqrt{x} + 3y_2 &= 0, \\ y_1^4 + 2y_2^3 &\leq 0, \\ u\left(y_1^4 + 2y_2^3\right) &= 0, \\ u &\geq 0. \end{aligned} \tag{4.5}$$

It is clear that the only solution of the above systems (4.5) are of the form $x = 0, y = (0, 0), u \geq 0$. That is the point $(0, (0, 0), u)$ is the fesasible point of the MPCC problem and is the global solution of MPCC. However, we have already shown that $x = 0, y = (0, 0)$ is a not a global solution of the bilevel problem. It is easily to see that no global minimum of the MPCC is a global minimum of the bilevel programming problem.

Theorem 4.5. *Let the lower level problem (4.2) be convex, Slater¡¯s constraint qualification be satisfied at the point $\bar{x}$ and $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution for problem (4.3) for all $\bar{u} \in \Lambda(\bar{x}, \bar{y})$. Then, the point $(\bar{x}, \bar{y})$ is a local optimal solution of problem (4.1), (4.2), too.*

Proof. Let $(\bar{x}, \bar{y})$ not be a local optimal solution, i.e. let there exists a sequence $\{(x^k, y^k)\} \subset gph\Psi$ converging to $(\bar{x}, \bar{y})$ with $x^k \in X$ and $F(x^k, y^k) < F(\bar{x}, \bar{y})$ for all k. Since Slater's constraint qualification is satisfied at $\bar{x}$ and persistent in some open neighborhood of $\bar{x}$, there exists $u^k \in \Lambda(x^k, y^k)$ having an accumulation point $\hat{u} \in \Lambda(\bar{x}, \bar{y})$. This means that there is a sequence (x^k, y^k, u^k) of feasible solutions to problem (4.3) converging to a feasible point $(\bar{x}, \bar{y}, \hat{u})$ of (4.3) with $F(x^k, y^k) < F(\bar{x}, \bar{y})$. This violates local optimality of $(\bar{x}, \bar{y}, \bar{u})$ for all $\bar{u} \in \Lambda(\bar{x}, \bar{y})$ for (4.3). Hence, $(\bar{x}, \bar{y})$ is local optimal for the bilevel programming problem.

Example 4.6. *Consider $M_1 = \mathbb{R}^2$ with the Euclidean metric g_1, and $M_2 = \mathbb{R}^2$ with the metric g_2 given in Cartesian coordinates (y_1, y_2) around the point $y \in M_2$ by the matrix*

$$M_2 \ni y \mapsto (g_{ij})_y = \left(g_2\left(\frac{\partial}{\partial y_i}, \frac{\partial}{\partial y_j}\right)\right) := \mathrm{diag}(1, 1).$$

It is easily to see that manifold (M_1, g_1) and (M_2, g_2) are complete Riemannian manifold. Moreover, the functions $f, \psi_1, \psi_2 : M_1 \times M_2 \to \mathbb{R}$ defined by $f(x, y) = y_1^2 + (y_2 + 1)^2$, $\psi_1(x, y) = (y_1 - x_1)^2 + (y_2 - 1 - x_1)^2$, $\psi_2(x, y) = (y_1 + x_2)^2 + (y_2 - 1 - x_2)^2$ are convex w.r.t y on the Riemannian manifold (M_2, g_2). So we can get the similar result that the local optimal solution of the MPCC need not be a local optimal solution to the bilevel programming problem.

In fact, the above Example 4.6 of bilevel program on the Riemannian manifolds show that where Slater's constraint qualification fails to the lower level problem, whereas a local solution of (4.1) is not solved MPCC.

5. *M*- and *C*-type Optimality Conditions

The above results in Section 4 motivate the following definition for the notion of optimality conditions for the bilevel optimization problem from the perspective of the *KKT* reformulation. We partition the set of indices of the functions involved in the complementarity slackness as follows

$$\eta = \eta(x^*, y^*, u^*) = \{i \in I : u_i^* = 0, \psi_i(x^*, y^*) < 0\},$$
$$\mu = \mu(x^*, y^*, u^*) = \{i \in I : u_i^* = 0, \psi_i(x^*, y^*) = 0\},$$
$$\nu = \nu(x^*, y^*, u^*) = \{i \in I : u_i^* > 0, \psi_i(x^*, y^*) = 0\}.$$

Definition 5.1. *A point (x^*, y^*) will be said to be M-stationary for the bilevel optimization problem (4.1) if there exists $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+r+m}$ such that $\forall u^* \in \Lambda(x, y)$:*

$$\mathrm{grad}_x F(x^*, y^*) + \mathrm{grad}_x G(x^*)^T \alpha + \mathrm{grad}_x \psi(x^*, y^*)^T \beta + \mathrm{grad}_{xy} L(x^*, y^*, u^*)^T \gamma = 0, \tag{5.1}$$

$$\mathrm{grad}_y F(x^*, y^*) + \mathrm{grad}_y \psi(x^*, y^*)^T \beta + \mathrm{grad}_{yy} L(x^*, y^*, u^*)^T \gamma = 0, \tag{5.2}$$

$$\alpha \geq 0, \alpha^T G(x^*) = 0, \tag{5.3}$$

$$\mathrm{grad}_y \psi_\nu(x^*, y^*)^T \gamma = 0, \beta_\eta = 0, \tag{5.4}$$

$$\forall i \in \mu, (\beta_i > 0 \wedge \mathrm{grad}_y \psi_i(x^*, y^*)^T \gamma > 0) \vee \beta_i(\mathrm{grad}_y \psi_i(x^*, y^*)^T \gamma) = 0. \tag{5.5}$$

Condition (5.1)-(5.5) are called the M-stationarity conditions for problem (4.3).

Definition 5.2. *A point (x^*, y^*) will be said to be C-stationary for the bilevel optimization problem (4.1) if there exists $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+r+m}$ such that $\forall u^* \in \Lambda(x, y)$, we have the relationships (5.1)-(5.4) to be satisfied together with the following condition:*

$$\forall i \in \mu, \beta_i \gamma^T \mathrm{grad}_y \psi_i(x^*, y^*) \geq 0. \tag{5.6}$$

Condition (5.1)-(5.4) and (5.6) are called the C-stationarity conditions for problem (4.3).

Theorem 5.3. *Let (x^*, y^*, u^*) be a local optimal solution of problem (4.3) and assume that the following CQ holds at (x^*, y^*, u^*):*

$$\left.\begin{array}{l} \mathrm{grad}_x G(x^*)^T \alpha + \mathrm{grad}_x \psi(x^*, y^*)^T \beta + \mathrm{grad}_{xy} L(x^*, y^*, u^*)^T \gamma = 0, \\ \mathrm{grad}_y \psi(x^*, y^*)^T \beta + \mathrm{grad}_{yy} L(x^*, y^*, u^*)^T \gamma = 0, \\ \alpha \geq 0, \alpha^T G(x^*) = 0, \\ \mathrm{grad}_y \psi_\nu(x^*, y^*)^T \gamma = 0, \beta_\nu = 0, \\ \forall i \in \mu, (\beta_i > 0 \wedge \mathrm{grad}_y \psi_i(x^*, y^*)^T \gamma > 0) \vee \beta_i(\mathrm{grad}_y \psi_i(x^*, y^*)^T \gamma) = 0. \end{array}\right\} \Rightarrow \left\{\begin{array}{l} \alpha = 0, \\ \beta = 0, \\ \gamma = 0. \end{array}\right.$$

Then, there exists $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+r+m}$, with $\|(\alpha, \beta, \gamma)\| \leq \varepsilon$ (for some $\varepsilon > 0$) such that the M-stationarity conditions are satisfied.

Proof. Let us set $\Psi(x, y, u, v) = (G(x), \psi(x, y) + v, \mathrm{grad}_y L(x, y, u))$, $\Lambda = \mathbb{R}^k_- \times \{0_{r+m}\}$ ($\mathbb{R}^k_-$ denote the nonpositive orthant in $\mathbb{R}^k$ and $\{0_{r+m}\}$ denote origin of the space $\mathbb{R}^{r+m}$) and $\Omega = M_1 \times M_2 \times \Theta$, $\Theta = \{(u, v) \in \mathbb{R}^{2r} | u \geq 0, v \geq 0, u^T v = 0\}$. Let (x^*, y^*, u^*) be a local optimal solution of problem (4.3). One can easily verify that there is a vector v^* such that (x^*, y^*, u^*, v^*) is a local optimal solution of the problem to

$$\begin{aligned} \min\ & F(x, y) \\ s.t.\ & (x, y, u, v) \in W = \Omega \cap \Psi^{-1}(\Lambda). \end{aligned} \tag{5.7}$$

According to the definition of normal cone on the Riemannian manifold, we have

$$N_L(x^*,y^*,u^*,v^*;\Omega)=\{0_{n+m}\}\times N_L(u^*,v^*;\Theta),$$
$$N_L(\Psi(x^*,y^*,u^*,v^*);\Lambda)=\{(\alpha,\beta,\gamma)|\alpha\geq 0,\alpha^TG(x^*)=0\},$$
$$\mathrm{grad}\Psi(x,y,u,v)^T(\alpha,\beta,\gamma)=\begin{bmatrix} A(\alpha,\beta,\gamma)\\ \beta\end{bmatrix}.$$

Combining the gradient of a differentiable function $\Psi(x,y,u,v)$ w.r.t. the Riemannian metric g_1 and g_2. We denote $N_F(u^*,v^*;\Theta)$ and $A(\alpha,\beta,\gamma)$ by

$$N_L(u^*,v^*;\Theta)=\left\{ (\bar{u},\bar{v})\in\mathbb{R}^{2r}: \begin{array}{ll} \bar{u}_i=0 & \forall i: u_i^*>0=v_i^* \\ \bar{v}_i=0 & \forall i: u_i^*=0<v_i^* \\ (\bar{u}_i<0\wedge\bar{v}_i<0)\vee\bar{u}_i\bar{v}_i=0 & \forall i: u_i^*=0=v_i^* \end{array}\right\},$$

and

$$A(\alpha,\beta,\gamma)=\begin{bmatrix} \mathrm{grad}_xG(x^*)^T\alpha+\mathrm{grad}_x\psi(x^*,y^*)^T\beta+\mathrm{grad}_{xy}L(x^*,y^*,u^*)^T\gamma \\ \mathrm{grad}_y\psi(x^*,y^*)^T\beta+\mathrm{grad}_{yy}L(x^*,y^*,u^*)^T\gamma \\ \mathrm{grad}_y\psi(x^*,y^*)^T\gamma \end{bmatrix}.$$

Using (3.3), we derive

$$0\in\partial_LF(x^*,y^*,v^*,u^*)+N_L(x^*,y^*,v^*,u^*;W).$$

Assumption (3.2) implies that the basic constraint qualification

$$\left.\begin{array}{l} 0\in(\lambda)^T\partial_L\Psi(x,y,v,u)+N_L(x,y,v,u;\Omega) \\ \lambda\in N_L(\Psi(x,y,u,v);\Lambda) \end{array}\right\}\Rightarrow\lambda=0$$

is satisfied. Hence, $N_L(x^*,y^*,v^*,u^*;W)=\{z^*:\exists\lambda\in N_L(\Psi(x,y,u,v);\Lambda)$ with $z^*=(\lambda)^T\partial_L\Psi(x,y,v,u;\Lambda)+N_L(x,y,v,u;\Omega)\}$. Thus, by using Theorem 3.2 we have that there exists $\theta>0$ such that, for all $\varepsilon\geq\theta$, we can get that $\lambda\in N_L(\Psi(x,y,u,v);\Lambda)$ and $\|\lambda\|\leq\varepsilon$ with

$$0\in\partial_LF(x^*,y^*,v^*,u^*;W)+\lambda^T\partial_L\Psi(x^*,y^*,v^*,u^*;\Lambda)+N_L(x^*,y^*,v^*,u^*;\Omega).$$

This implies that the basic CQ applied to problem (5.7) at (x^*,y^*,u^*,v^*) can equivalently be formulated as follows: there is no nonzero vector $(\alpha,\beta,\gamma)\in\mathbb{R}_{k+r+m}$ such that

$$\mathrm{grad}_xG(x^*)^T\alpha+\mathrm{grad}_x\psi(x^*,y^*)^T\beta+\mathrm{grad}_{xy}L(x^*,y^*,u^*)^T\gamma=0,$$
$$\mathrm{grad}_y\psi(x^*,y^*)^T\beta+\mathrm{grad}_{yy}L(x^*,y^*,u^*)^T\gamma=0,$$
$$\alpha\geq 0,\alpha^TG(x^*)=0,$$
$$(-\mathrm{grad}_y\psi(x^*,y^*)^T\gamma,-\beta)\in N_L(u^*,v^*;\Theta).$$

By noting that $v_i^*=-\psi_i(x^*,y^*)$, for $i\in I$. Hence, there exists $(\alpha,\beta,\gamma)\in\mathbb{R}^{k+r+m}$, with $\|(\alpha,\beta,\gamma)\|\leq\varepsilon$ (for some $\varepsilon>0$) such that (5.1)-(5.5) are satisfied.

Corollary 5.4. *Let (x^*,y^*) be a local optimal solution of Bilevel problem (4.1) and (4.2), where the lower-level problem (4.2) is convex. Assume that the Slater's CQ holds at x^* while for all $u^*\in\Lambda(x^*,y^*)$, CQ (5.1)-(5.2) and (5.4)-(5.5) hold at (x^*,y^*,u^*). Then (x^*,y^*) is M-stationary.*

Theorem 5.5. *Let (x^*, y^*, u^*) be a local optimal solution of problem (4.3) and assume that the following CQ holds at (x^*, y^*, u^*):*

$$\left.\begin{array}{l} \operatorname{grad}_x G(x^*)^T\alpha + \operatorname{grad}_x \psi(x^*, y^*)^T\beta + \operatorname{grad}_{xy} L(x^*, y^*, u^*)^T\gamma = 0, \\ \operatorname{grad}_y \psi(x^*, y^*)^T\beta + \operatorname{grad}_{yy} L(x^*, y^*, u^*)^T\gamma = 0, \\ \alpha \geq 0, \alpha^T G(x^*) = 0, \\ \operatorname{grad}_y \psi_\nu(x^*, y^*)^T\gamma = 0, \beta_\nu = 0, \\ \forall i \in \mu, \beta_i(\operatorname{grad}_y \psi_i(x^*, y^*)^T\gamma = 0. \end{array}\right\} \Rightarrow \left\{\begin{array}{l} \alpha = 0, \\ \beta = 0, \\ \gamma = 0. \end{array}\right. \tag{5.8}$$

Then, the C-stationarity conditions (5.1)-(5.4) and (5.6) hold with $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+r+m}$, with $\|(\alpha, \beta, \gamma)\| \leq \varepsilon$ for some $\varepsilon > 0$.

Proof. Following the work by Scheel and Scholtes [28], the *KKT* reformulation (4.3) can take the operator constraint form (3.1) with $\Omega = M_1 \times M_2 \times \mathbb{R}^r$, $\Psi(x, y, u) = \{G(x), V(x, y, u), \operatorname{grad}_y L(x, y, u)\}$ and $L := \mathbb{R}^k_- \times \{0_{m+r}\}$. Here,

$$V_i(x, y, u) = \min\{u_i, -\psi_i(x, y)\} \; for \; i \in I.$$

Applying Theorem 3.2 to the corresponding operator constraint reformulation of (3.1), there exists (α, β, γ) with $\|(\alpha, \beta, \gamma)\| \leq \varepsilon$ for some $\varepsilon > 0$ such that we have:

$$0 \in \begin{pmatrix} \partial_L F(x^*, y^*) \\ 0 \end{pmatrix} + (\alpha, \beta, \gamma)^T \partial_L \Psi(x^*, y^*, u^*) + N_L(x^*, y^*, u^*; \Omega)$$

with

$$(\alpha, \beta, \gamma) \in N_L(\Psi(x^*, y^*, u^*), \Lambda) = \{(\alpha, \beta, \gamma) | \alpha \geq 0, \alpha^T G(x^*) = 0\},$$

provided the following implication is satisfied at the point (x^*, y^*, u^*):

$$\left.\begin{array}{l} 0 \in (\alpha, \beta, \gamma)^T \partial_L \Psi(x^*, y^*, u^*) + N_L(x^*, y^*, u^*; \Omega) \\ (\alpha, \beta, \gamma) \in N_L(\Psi(x^*, y^*, u^*); \Lambda) \end{array}\right\} \Rightarrow (\alpha, \beta, \gamma) = 0. \tag{5.9}$$

Over here $N_L(x^*, y^*, u^*; \Omega) = 0$, and observe that $(\alpha, \beta, \gamma)^T \partial_L \Psi(x^*, y^*, u^*)$ can be written as follows

$$(\alpha, \beta, \gamma)^T \partial_L \Psi(x^*, y^*, u^*) = \begin{pmatrix} \alpha^T \operatorname{grad}_x G(x^*) + \beta^T \operatorname{grad}_x V(x^*, y^*, u^*) + \gamma^T \operatorname{grad}_x y L(x^*, y^*, u^*) \\ \beta^T \operatorname{grad}_y V(x^*, y^*, u^*) + \gamma^T \operatorname{grad}_y y L(x^*, y^*, u^*) \\ \beta^T \operatorname{grad}_u V(x^*, y^*, u^*) + \gamma^T \operatorname{grad}_y \psi(x^*, y^*) \end{pmatrix},$$

Moreover,

$$\partial_L V(x^*, y^*, u^*) = \begin{cases} -(\partial_L \psi_i(x^*, y^*), 0), & i \in \upsilon; \\ (0, e^i), & i \in \eta; \\ \operatorname{conv}(-(\partial_L \psi_i(x^*, y^*), 0), (0, e^i)), & i \in \mu. \end{cases}$$

For $i \in \eta$, we have $\beta_i \operatorname{grad}_x V_i(x^*, y^*, u^*) = \beta_i \operatorname{grad}_y V_i(x^*, y^*, u^*) = 0$, and $\beta_i \operatorname{grad}_u V_i(^*x, y^*, u^*) = \beta_i$. For $i \in \upsilon$, we have $\beta_i \operatorname{grad}_x V_i(x^*, y^*, u^*) = -\beta_i \operatorname{grad}_x \psi_i(x^*, y^*, u^*)$, $\beta_i \operatorname{grad}_y V_i(x^*, y^*, u^*) = -\beta_i \operatorname{grad}_y \psi_i(x^*, y^*, u^*)$, and $\beta^T \operatorname{grad}_u V(^*x, y^*, u^*) = 0$.

For $i \in \mu$, $\beta_i \operatorname{grad}_u V_i(^*x, y^*, u^*) = \beta_i(1 - \theta), \theta \in [0, 1]$. We so have that

$$\begin{array}{l} \operatorname{grad}_x G(x^*)^T\alpha + \operatorname{grad}_x \psi(x^*, y^*)^T\beta + \operatorname{grad}_{xy} L(x^*, y^*, u^*)^T\gamma = 0, \\ \operatorname{grad}_y \psi(x^*, y^*)^T\beta + \operatorname{grad}_{yy} L(x^*, y^*, u^*)^T\gamma = 0, \\ \alpha \geq 0, \alpha^T G(x^*) = 0, \\ \operatorname{grad}_y \psi_\nu(x^*, y^*)^T\gamma = 0, \beta_\nu = 0, \\ \forall i \in \mu, \beta_i(1 - \theta) + \gamma^T \operatorname{grad}_y \psi_i(x^*, y^*) = 0, \theta \in [0, 1]. \end{array} \tag{5.10}$$

By noting that, the above last equation $\beta_i(1-\theta)+\gamma^T \mathrm{grad}_y \psi_i(x^*, y^*) = 0$ can change to $\beta_i \gamma^T \mathrm{grad}_y \psi(x^*, y^*) > 0, i \in \mu$. Hence, one can easily check that CQ (5.8) implies the fulfilment of condition (5.9). The C-stationarity conditions (5.1)-(5.4) and (5.6) are obtained by successively inserting (5.10) with $(\alpha, \beta, \gamma) \in R^{k+r+m}$, $\|(\alpha, \beta, \gamma)\| \leq \varepsilon$ hold for some $\varepsilon > 0$.

Corollary 5.6. *Let (x^*, y^*) be a local optimal solution of the bilevel program, where the lower-level problem (4.2) is convex. Assume that the Slater's CQ holds at x^* while for all lower-level multipliers $u^* \in \Lambda(x^*, y^*)$, CQ (5.1)-(5.2), (5.4) and (5.6) holds at (x^*, y^*, u^*). Then (x^*, y^*) is C-stationary.*

6. Conclusion

In this paper, we first present *KKT* reformulation of the bilevel optimization on Riemannian manifolds and show that global or local optimal solutions of the MPCC correspond to global or local optimal solutions of the bilevel problem on the Riemannian manifolds provided the lower level convex problem satisfies the Slater's constraint qualification. Furthermore, we have given some examples to prove that this correspondence can fail if the Slater's constraint qualification fails to hold at lower-level convex problem. In the end, we have studied *M*- and *C*-type optimality conditions for the bilevel problem on Riemannian manifolds.

Funding

This work is supported by Natural Science Foundation of China (Grant No. 11871383).

Acknowledgments

The authors appreciate the worthy comments from reviewers.

References

[1] D. Azagra, J. Ferrera, F. López-Mesas, Nonsmooth analysis and HamiltonšCJacobi equations on Riemannian manifolds, Journal of Functional Analysis, 2 (2005) 304–361, https://doi.org/10.1016/j.jfa.2004.10.008.
[2] P.A. Absil, R. Mahony, Sepulchre, Rodolphe, Optimization Algorithms on Matrix Manifolds, Princeton: Princeton University Press, 2009. https://doi.org/10.1515/9781400830244.
[3] G. B. Allende, G. Still, Solving bilevel programs with the KKT-approach, Mathematical programming, 138 (2013) 309-332.
[4] G. C. Bento, O. P. Ferreira, P. R. Oliveira, Unconstrained steepest descent method for multicriteria optimization on Riemannian manifolds, Journal of Optimization Theory and Applications, 154 (2012) 88–107.
[5] G. C. Bento, J. C. Neto, Subgradient method for convex feasibility on Riemannian manifolds, Journal of Optimization Theory and Applications, 152 (2012) 773–785.
[6] G. C. Bento, J. X. da Cruz Neto, P. Santos, An inexact steepest descent method for multicriteria optimization on Riemannian manifolds, Journal of Optimization Theory and Applications, 159 (2013) 108–124.
[7] G. C. Bento, J. C. Neto, A subgradient method for multiobjective optimization on Riemannian manifolds, Journal of Optimization Theory and Applications, 159 (2013) 125–137.
[8] W. P. Klingenberg, Riemannian geometry. Walter de Gruyter, 2011.
[9] H. Bonnel, L. Todjihoundé, Udrişte, Constantin, Semivectorial bilevel optimization on riemannian manifolds, Journal of Optimization Theory and Applications, 167 (2015) 464–486.
[10] N. Boumal, An introduction to optimization on smooth manifolds. Available online, Aug 2020.
[11] S. Dempe, Foundations of bilevel programming, Springer Science & Business Media, 2002.
[12] J. Dutta, S. Dempe, Bilevel programming with convex lower level problems, In: Optimization with multivalued mappings. Springer, Boston, MA, (2006) 51–71.
[13] S. Dempe, A. B. Zemkoho, The generalized Mangasarian-Fromowitz constraint qualification and optimality conditions for bilevel programs, Journal of optimization theory and applications, 148 (2011) 46–68.
[14] S. Dempe, J. Dutta, Is bilevel programming a special case of a mathematical program with complementarity constraints?, Mathematical programming, 131 (2012) 37–48.
[15] S. Dempe, A. B. Zemkoho, On the Karush–Kuhn–Tucker reformulation of the bilevel optimization problem, Nonlinear Analysis: Theory, Methods & Applications, 75 (2012) 1202–1218.
[16] S. Dempe, V. Kalashnikov, G. A. Pérez-Valdés, N. Kalashnykova, Bilevel Programming Problems, Energy Systems. Springer, Berlin, 2015.
[17] L. P. Eisenhart, Riemannian geometry, Princeton university press, 2016.

[18] O. P. Ferreira, Dini derivative and a characterization for Lipschitz and convex functions on Riemannian manifolds, Nonlinear Analysis: Theory, Methods & Applications, 68 (2008) 1517–1528.
[19] W. Huang, P. A. Absil, K. A. Gallivan, A Riemannian BFGS method for nonconvex optimization problems, Numerical Mathematics and Advanced Applications ENUMATH, 2015 (2016) 627–634.
[20] W. Huang, K. A. Gallivan, P. A. Absil, A Broyden class of quasi-Newton methods for Riemannian optimization, SIAM Journal on Optimization, 25 (2015) 1660–1685.
[21] J. Jost, Riemannian geometry and geometric analysis, Vol. 42005, Springer Science & Business Media 2008.
[22] A. Kristály, Nash-type equilibria on Riemannian manifolds: a variational approach, Journal de Mathématiques Pures et Appliquées, 101 (1996) 660–688.
[23] Y. Ledyaev, Q. J. Zhu, Nonsmooth analysis on smooth manifolds, Transactions of the American Mathematical Society, 359 (2007) 3687–3732.
[24] Z. Q. Luo, J. S. Pang, D. Ralph, Mathematical programs with equilibrium constraints, Cambridge University Press, 1996.
[25] B. S. Mordukhovich, Variational analysis and generalized differentiation I: Basic theory. Vol. 330, Springer Science & Business Media, 2006.
[26] E. P. Quiroz, E. Quispe, P. R. Oliveira, Steepest descent method with a generalized Armijo search for quasiconvex functions on Riemannian manifolds, Journal of Optimization Theory and Applications, 341 (2008) 467–477.
[27] R. T. Rockafellar, R. J. Wets, Variational analysis, Vol. 317, Springer Science & Business Media, 2009.
[28] H. Scheel, S. Scholtes, Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity. Mathematics of Operations Research 25 (2000) 1–22.
[29] C. Udriste, Convex functions and optimization methods on Riemannian manifolds. Vol. 297, Springer Science & Business Media, 2013.
[30] X. M. Wang, C. Li, J. H. Wang, J. C. Yao, Linear convergence of subgradient algorithm for convex feasibility on Riemannian manifolds, SIAM Journal on Optimization, 25 (2015) 2334–2358.
[31] X. M. Wang, C. Li, J. C. Yao, Subgradient projection algorithms for convex feasibility on Riemannian manifolds with lower bounded curvatures, Journal of Optimization Theory and Applications, 164 (2015) 202–217.