

Article

Inexact Restoration Methods for Semivectorial Bilevel Programming Problem on Riemannian Manifolds

Jiagen Liao ^{*,†}  and Zhongping Wan [†]

School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China

* Correspondence: jiagenliao@163.com

† These authors contributed equally to this work.

Abstract: For a better understanding of the bilevel programming on Riemannian manifolds, a semivectorial bilevel programming scheme is proposed in this paper. The semivectorial bilevel programming is firstly transformed into a single-level programming problem by using the Karush–Kuhn–Tucker (KKT) conditions of the lower-level problem, which is convex and satisfies the Slater constraint qualification. Then, the single-level programming is divided into two stages: restoration and minimization, based on which an Inexact Restoration algorithm is developed. Under certain conditions, the stability and convergence of the algorithm are analyzed.

Keywords: Riemannian manifolds; semivectorial bilevel programming; Inexact Restoration algorithm

MSC: 58C05; 90C25; 90C29



Citation: Liao, J.; Wan, Z. Inexact Restoration Methods for Semivectorial Bilevel Programming Problem on Riemannian Manifolds. *Axioms* **2022**, *11*, 696. <https://doi.org/10.3390/axioms11120696>

Academic Editor: Clemente Cesarano

Received: 30 September 2022

Accepted: 2 December 2022

Published: 5 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The bilevel optimization problem on Euclidean spaces has been shown to be NP-hard, and even the verification of the local optimality for a feasible solution is in general NP-hard. Bilevel optimization problems are often nonconvex optimization problems, and this makes the computation of an optimal solution a challenging task. Thus, it is natural to consider the bilevel optimization problems on Riemannian manifolds. Actually, studying optimization problems on Riemannian manifolds has many advantages. Some constrained optimization problems on Euclidean spaces can be seen as unconstrained ones from the Riemannian geometry viewpoint. Moreover, some nonconvex optimization problems in the setting of Euclidean spaces may become convex optimization problems by introducing an appropriate Riemannian metric. See for instance [1,2]. The aim of this paper is to study the bilevel optimization problem on Riemannian manifolds.

In order to study the bilevel optimization problem on Riemannian manifolds, it is reasonable to have some idea of solving the bilevel optimization problem in Euclidean spaces. An approach to investigate bilevel optimization problems on Euclidean spaces is to replace the lower-level problem by its (under certain necessary and sufficient assumptions) KKT optimality conditions. In a recent article [3], the authors presented the KKT reformulation of the bilevel optimization problems on Riemannian manifolds. Moreover, it has been shown that global optimal solutions of the KKT reformulation correspond to global optimal solutions of the bilevel problem on the Riemannian manifolds provided the lower level convex problem satisfies Slater's constraint qualification. On this basis, we consider a semivectorial bilevel optimization problem on Riemannian manifolds with a multiobjective problem in the lower-level problem. Since the Inexact Restoration (IR) algorithm [4,5] was introduced to solve constrained optimization problems and if we transform the semivectorial bilevel optimization problem into a single-level problem, it also can be solved by using the IR algorithm as a constrained optimization problem.

For the convenience of the readers, let us review the IR algorithm on Euclidean spaces firstly. Each iteration of the IR algorithm consists of two phases: restoration and minimization. Consider the following nonlinear programming:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & C(x) \leq 0, \quad x \in \Omega, \end{aligned} \quad (1)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $C : \mathbb{R}^n \rightarrow \mathbb{R}^m$ are continuous differentiable functions and the set $\Omega \subset \mathbb{R}^n$ is closed convex. The algorithm generates feasible iterates with respect to Ω , $x^k \in \Omega$ (for all $k = 0, 1, 2, \dots$).

In the restoration step, which is executed once per iteration, an intermediate point $y^k \in \Omega$ is found such that the infeasibility at y^k is a fraction of the infeasibility at x^k . Immediately after restoration, we construct an approximation π_k of the feasible region using available information at y^k . In the minimization step, we compute a trial point $z^{ki} \in \pi_k$ such that $f(z^{ki}) \ll f(y^k)$. Here, the symbol \ll means sufficiently smaller than, and $\|z^{ki} - y^k\| \leq \delta_{ki}$, where δ_{ki} is a trust-region radius. The trial point z^{ki} is accepted as a new iteration one if the value of a nonsmooth (exact penalty) merit function at z^{ki} is sufficiently smaller than its value at x^k . If z^{ki} is not acceptable, the trust-region radius is reduced.

The IR algorithm is related to classical feasible methods for nonlinear programming, such as the generalized reduced gradient (GRG) and the family of sequential gradient restoration algorithms. There are several studies on the numerical characteristics of the IR algorithm. For example, this method was applied to the general constraint problem in [6], and good results were obtained. In addition, the IR algorithm using the regularization strategy was proposed in [7], in which the problem of derivative-free optimization was effectively solved. The IR algorithms are especially useful when there is some natural way to restore feasibility. One of the most successful applications of the IR algorithm is electronic structure calculation, as shown in [8]. Moreover, the IR algorithm has also been successfully applied to optimization problems with the box constraint in [9] and problems with multiobjective constraints under weighted-sum scalarization in [10]. For more applications, please see [11,12].

Since the IR algorithm is so important in applications, many researches have been trying to improve it from different angles. The restoration phase improves feasibility, and in the minimization step, optimality is improved as a linear tangent approximation of the constraints. When a sufficient descent criterion does not hold, the trial point is modified in such a way that, eventually, acceptance occurs at a point that may be close to the solution of the restoration (first) phase. The acceptance criterion may use merit functions [4,5] or filters [13]. The minimization step consists of an inexact (approximate) minimization of f with linear constraints. In this case, the restoration step represents also an inexact minimization of infeasibility with linear constraints. Therefore, the available algorithms for (large-scale) linearly constrained minimization can be fully exploited; see the published articles [14–16]. Furthermore, IR techniques for constrained optimization were improved, extended, and analyzed in [7,17–19], among others.

Inspired and motivated by the research works [4,10,20–25], we introduce a kind of bilevel programming with a multiobjective problem in the lower level on Riemannian manifolds, the so-called semivectorial bilevel programming. Then, we transform the semivectorial bilevel programming into a single-level programming by using the KKT optimality conditions of the lower-level problem, which is convex and satisfies the Slater constraint qualification. Finally, we divide the single-level programming into two stages: restoration and minimization, and give an IR algorithm for semivectorial bilevel programming. Under certain conditions, we analyze the well-definiteness and convergence of the presented algorithm.

The remainder of this paper is organized as follows: In Section 2, some basic concepts, notations, and important results of Riemannian geometry are presented. In Section 3, we propose the semivectorial bilevel programming on the Riemannian manifold and give

the KKT reformulation, and then, we present an algorithm by using the IR technique for solving the semivectorial bilevel programming on Riemannian manifolds. In Section 4, its convergence properties are studied. The conclusions 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 $x \in M$, the restriction $g_x : T_xM \times T_xM \rightarrow \mathbb{R}$ is an inner product on the tangent space T_xM . The tangent bundle TM over M is $TM := \bigcup_{x \in M} T_xM$, and a vector field on M is a section of the tangent bundle, which is a mapping $X : M \rightarrow TM$ such that, for any $x \in M$, $X(x) \equiv X_x \in T_xM$.

We denote $\langle \cdot, \cdot \rangle_x$ by the scalar product on T_xM with the associated norm $\|\cdot\|_x$. The length of a tangent vector $v \in T_xM$ is defined by $\|v\|_x = \langle v, v \rangle_x^{\frac{1}{2}}$. Given a piecewise smooth curve $\gamma : [a, b] \subset \mathbb{R} \rightarrow M$ joining x to y , i.e., $\gamma(a) = x$ and $\gamma(b) = y$, 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 t . Let x and y be two points in Riemannian manifold (M, g) and $\Gamma_{x,y}$ the set of all piecewise smooth curves joining x and y . The function:

$$d : M \times M \rightarrow \mathbb{R}, d(x, y) := \inf\{L(\gamma) : \gamma \in \Gamma_{x,y}\}$$

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

Let ∇ be the Levi-Civita connection associated with the Riemannian metric and γ be a smooth curve in M . A vector field X is said to be parallel along $\gamma : [0, 1] \rightarrow M$ if $\nabla_{\dot{\gamma}}X = 0$. If $\dot{\gamma}$ itself is parallel along γ joining x to y ,

$$\gamma(0) = x, \gamma(1) = y \text{ and } \nabla_{\dot{\gamma}}\dot{\gamma} = 0 \text{ on } [0, 1],$$

then we say that γ is a geodesic, and in this case, $\|\dot{\gamma}\|$ is constant. When $\|\dot{\gamma}\| = 1$, γ is said to be normalized. A geodesic joining x to y in M is said to be minimal if its length equals $d(x, y)$.

By the Hopf–Rinow theorem, we know that, if M is complete, then any pair of points in M can be joined by a minimal geodesic. Moreover, (M, d) is a complete metric space, and the bounded closed subsets are compact. Furthermore, for the exponential mapping at x , $\exp_x : T_xM \rightarrow M$ is well defined on T_xM . Clearly, a curve $\gamma : [0, 1] \rightarrow M$ is a minimal geodesic joining x to y if and only if there exists a vector $v \in T_xM$ such that $\|v\| = d(x, y)$ and $\gamma(t) = \exp_x(tv)$ for each $t \in [0, 1]$.

Set $p \in M$ and $V_p := \{v \in T_pM : \gamma_v \text{ defined in } [0, 1]\}$. The exponential mapping $\exp_p : V_p \rightarrow M$ is defined by $\exp_p(v) = \gamma_v(1), \forall v \in V_p$. The exponential mapping $\exp_p : T_pM \rightarrow M$ at $p \in M$ is well posed on the tangent space T_pM . Obviously, a curve $\gamma : [0, 1] \rightarrow M$ joining p and q is a minimum geodesic, if and only if there is a vector $v \in T_pM$ such that $\|v\| = d(p, q)$ and $\gamma(t) = \exp_p(tv)$ hold for every $t \in [0, 1]$.

The gradient of a differentiable function $f : M \rightarrow \mathbb{R}$ with respect to the Riemannian metric g is the vector field $\text{grad}f$ defined by $g(\text{grad}f, X) = df(X), \forall X \in TM$, where df denotes the differential of the function f .

In this normal coordinate system, the geodesics through p are represented by lines passing through the 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 $f : M \rightarrow \mathbb{R}$, in normal coordinates around p , we obtain

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

Now, consider a smooth function $f : M \rightarrow \mathbb{R}$ and the real-valued function $T_p M \ni v \mapsto f_p(v) := f(\exp_p v)$ defined around 0 in $T_p M$.

It is easy to see that

$$\frac{\partial f_p}{\partial x^i}(0) = \frac{\partial f}{\partial x^i}(p).$$

The Taylor–Young formula (for Euclidean spaces) applied to f_p around the origin can be written using matrices as

$$f_p(v) = f_p(0) + J_{f_p}(0)v + \frac{1}{2}v^T \text{Hess}_{f_p}(0)v + o(\|v\|^2),$$

where

$$\begin{aligned} v &= [v^1 \dots v^n]^T, \\ J_{f_p}(0) &= \left[\frac{\partial f}{\partial x^1}(p) \dots \frac{\partial f}{\partial x^n}(p) \right], \\ \text{Hess}_{f_p}(0) &= \left(\frac{\partial^2 f}{\partial x^i \partial x^j}(p) \right) = \text{Hess}_p f(v, v). \end{aligned}$$

In other words, we have the following Taylor–Young expansion for f around p :

$$f(\exp_p v) = f(p) + g_p(\text{grad} f, v) + \frac{1}{2} \text{Hess}_p f(v, v) + o(\|v\|_p^2)$$

which holds in any coordinate system.

The set $A \subset M$ is said to be convex if it contains a geodesic segment γ whenever it contains the end points of γ , that is $\gamma((1 - t)a + tb)$ is in A whenever $x = \gamma(a)$ and $y = \gamma(b)$ are in A , and $t \in [0, 1]$. A function $f : M \rightarrow \mathbb{R}$ is said to be convex if its restriction to any geodesic curve $\gamma : [a, b] \rightarrow M$ is convex in the classical sense, such that the one real variable function $f \circ \gamma : [a, b] \rightarrow \mathbb{R}$ is convex. Let P_A denote the projection on $A \subset M$, that is, for each $x \in M$,

$$P_A x = \left\{ \bar{x} \in A : d(x, \bar{x}) = \inf_{z \in A} d(x, z) \right\}. \tag{2}$$

For more details and complete information on the fundamentals in Riemannian geometry, see [1,26–28].

3. Inexact Restoration Algorithm

We study an optimistic bilevel programming on an m -dimensional Riemannian manifold (M, g) , where the lower-level problem is a multi-objective problem, the so-called semivectorial bilevel programming. The problem is formulated below:

$$\begin{aligned} \min \quad & F(x) \\ \text{s.t.} \quad & x \in \text{Sol}(\text{MOP}), \end{aligned} \tag{3}$$

where $F : M \rightarrow \mathbb{R}$ and $\text{Sol}(\text{MOP})$ is the effective solution set of the following multi-objective problem (MOP):

$$\begin{aligned} \min \quad & \{f_1(x), \dots, f_p(x)\} \\ \text{s.t.} \quad & h(x) = 0, \\ & x \in M, \end{aligned} \tag{4}$$

where $f = \{f_1(x), \dots, f_p(x)\} : M \rightarrow \mathbb{R}^p$, $I := \{1, \dots, p\}$, $h : M \rightarrow \mathbb{R}^n$, and $D = \{x \in M : h(x) = 0\}$ denote the feasible solution of the MOP.

Definition 1. Let $f : M \rightarrow \mathbb{R}^p$ be a vectorial function on Riemannian manifold M . Then, f is said to be convex on M if, for every $x, y \in M$ and every geodesic segment $\gamma : [0, 1] \rightarrow M$ joining x to y , i.e., $\gamma(0) = x$ and $\gamma(1) = y$, it holds that

$$f(\gamma(t)) \preceq (1 - t)f(x) + tf(y), \quad t \in [0, 1].$$

The above definition is a natural extension of the definition of convexity in Euclidean space to the Riemannian context; see [29].

Definition 2. A point $x \in M$ is said to be Pareto critical of f on Riemannian manifold M if, for any $v \in T_x M$, there are an index $i \in I$ and $u \in \text{grad} f_i(x)$, such that

$$\langle u, v \rangle \geq 0.$$

Definition 3. (a) A point $x^* \in M$ is a Pareto-optimal point of f on Riemannian manifold M if there is no $x \in M$ with $f(x) \preceq f(x^*)$. (b) A point $x^* \in M$ is a weak Pareto-optimal point of f on Riemannian manifold M if there is no $x \in M$ with $f(x) \prec f(x^*)$.

We know that criticality is a necessary, but not a sufficient condition for optimality. Under the convexity of the vectorial function f , the following proposition shows that criticality is equivalent to weak optimality.

Proposition 1 ([29]). Let $f : M \rightarrow \mathbb{R}^p$ be a convex function given by $f = \{f_1(x), \dots, f_p(x)\}$. A point $x \in M$ is a critical Pareto-optimal point of the function f if and only if it is a weak Pareto-optimal point of the function f .

We assume that the functions $f = \{f_1(x), \dots, f_p(x)\} : M \rightarrow \mathbb{R}^p$ and $h : M \rightarrow \mathbb{R}^n$ are twice continuously differentiable and consider the weighted sun scaling problem related to the MOP, as follows.

Let $\omega_i \geq 0, i = 1, \dots, p$ such that $\sum_{i=1}^p \omega_i = 1$:

$$\begin{aligned} \min_x \quad & \sum_{i=1}^p \omega_i f_i(x) \\ \text{s.t.} \quad & h(x) = 0, \\ & x \in M. \end{aligned} \tag{5}$$

Note that, if $\omega_i \geq 0, i = 1, \dots, p$ such that $\sum_{i=1}^p \omega_i = 1$, then the weak Pareto-optimal solution sets of Problem (4) are equivalent to the union of the optimal solution sets of Problem (5). Meanwhile, if $f_i : M \rightarrow \mathbb{R}, i = 1, \dots, p$ is the convex function on the Riemannian manifold, then the function $\sum_{i=1}^p \omega_i f_i(x)$ is also convex. Thus, the bilevel programming (3)–(4) can be transformed into the following problem:

$$\begin{aligned} \min_{x, \omega} \quad & F(x) \\ \text{s.t.} \quad & \sum_{i=1}^p \omega_i = 1, \\ & \omega_i \geq 0, i \in I, \\ & x \in \arg \min \left\{ \begin{array}{l} \min \sum_{i=1}^p \omega_i f_i(x) \\ \text{s.t. } h(x) = 0, \\ x \in M. \end{array} \right\}. \end{aligned} \tag{6}$$

A strategy to solve the bilevel problem (6) on the Riemannian manifolds is to replace the lower-level problem with the KKT conditions. When the lower-level problem is convex and satisfies the Slater constraint qualification, the global optimal solutions of the KKT reformulation correspond to the global optimal solutions of the bilevel problem on the Riemannian manifolds. See Theorems 4.1 and 4.2 in [3].

In the following, we give the KKT reformulation of the semivectorial bilevel programming on Riemannian manifolds.

$$\begin{aligned}
 & \min_{x, \omega} F(x) \\
 & \text{s.t. } \omega \in W, \\
 & \sum_{i=1}^p \omega_i \text{grad}_x f_i(x) + \text{grad}_x h(x) \mu = 0, \\
 & h(x) = 0, \\
 & x \in M,
 \end{aligned} \tag{7}$$

where

$$W = \left\{ \omega \in \mathbb{R}^p : \sum_{i=1}^p \omega_i = 1, \omega_i \geq 0, i = 1, \dots, p \right\}$$

is a convex and compact set, $\mu \in \mathbb{R}^n$, and M is a complete m -dimensional Riemannian manifold.

We will adopt an IR method to solve the optimization problem in two stages, first pursuing feasibility and optimality, keeping a certain control over the feasibility that has been realized. Consequently, the approach exploits the inherent minimization structure of the problem, especially in the feasibility phase, so that it can obtain better solutions. Moreover, in the feasibility phase of the IR strategy, the user is free to choose the method of his/her choice, as long as the recovered iteration satisfies some mild assumptions [4,5].

For simplicity, we introduce the following notations:

$$C(x, \omega, \mu) = \begin{pmatrix} \sum_{i=1}^p \omega_i \text{grad}_x f_i(x) + \text{grad}_x h(x) \mu \\ h(x) \end{pmatrix} \in \mathbb{R}^{m+n} \tag{8}$$

and

$$L(x, \omega, \mu, \lambda) = F(x) + C(x, \omega, \mu)^T \lambda, \quad \lambda \in \mathbb{R}^{m+n}. \tag{9}$$

We write shortly $s = (x, \omega, \mu) \in M \times W \times \mathbb{R}^n$ and give the Jacobian of C as follows:

$$C'(s) = \begin{pmatrix} \sum_{i=1}^p \omega_i \text{Hess}_x f_i + \sum_{j=1}^n \mu_j \text{Hess}_x h_j & \text{grad}_x f_1 & \cdots & \text{grad}_x f_p & \text{grad}_x h \\ \text{grad}_x h^T & 0 & \cdots & 0 & 0 \end{pmatrix}. \tag{10}$$

Thus, the semivectorial bilevel programming can be reduced:

$$\begin{aligned}
 & \min F(s) \\
 & \text{s.t. } C(s) = 0, \\
 & s \in M \times W \times \mathbb{R}^n.
 \end{aligned} \tag{11}$$

Before giving a rigorous description of the algorithm, let us start with an overview of each step.

Restoration step: We apply any globally convergent optimization algorithm to solve the lower-level minimization problem parameterized by $z^k = (\bar{x}, \omega^k, \bar{\mu})$. Once an approximate minimizer \bar{x} and a pair of corresponding estimated Lagrange multiplier vectors are obtained, then we compute the current set π_k and the direction d_{tan}^k .

Approximate linearized feasible region: The set π_k is a linear approximation of the region described by KKT(\bar{x}) containing $z^k = (\bar{x}, \omega^k, \bar{\mu})$. This auxiliary region is given by

$$\pi_k = \{s \in M \times W \times \mathbb{R}^n : \langle C'(z^k), \hat{\gamma}_{s,z^k}(0) \rangle = 0\}.$$

Descent direction: Using the projection on Riemannian manifolds, the projection defined on π_k is represented as follows:

$$P_{\pi_k}(z^k) = P_k\left(\exp_{z^k}\left(-\eta \text{grad}_s L(z^k, \lambda^k)\right)\right),$$

where $\eta > 0$ is an arbitrary scaling parameter independent of k . It turns out that

$$d_{\text{tan}}^k = P_k\left(\exp_{z^k}\left(-\eta \text{grad}_s L(z^k, \lambda^k)\right)\right) - z^k$$

which is a feasible descent direction on π_k .

Minimization step: The objective of the minimization step is to obtain $v^{k,i} \in \pi_k$ such that $L(v^{k,i}, \lambda^k) < L(z^k, \lambda^k)$ and $v^{k,i} \in B_{k,i} = \{v : d(v, z^k) \leq \delta_{k,i}\}$, where $\delta_{k,i}$ is a trust-region radius. The first trial point at each iteration is obtained using a trust-region radius $\delta_{k,0}$. A successive trust-region radius is tried until a point $v^{k,i}$ is found such that the merit function at this point is sufficiently smaller than the merit function at s^k .

Merit function and penalty parameter: We decided to use a variant of the sharp Lagrangian merit function, given by

$$\Psi(s, \lambda, \theta) = \theta L(s, \lambda) + (1 - \theta)|C(s)|,$$

where $\theta \in (0, 1]$ is a penalty parameter used to give different weights to the objective function and the feasibility objective. The choice of the parameter θ at each iteration depends on practical and theoretical considerations. Roughly speaking, we wish the merit function at the new point to be less than the merit function at the current point s^k .

That is, we want $\text{Ared}_{k,i} > 0$, where $\text{Ared}_{k,i}$ is the actual reduction of the merit function, defined by

$$\text{Ared}_{k,i} = \Psi(s^k, \lambda^k, \theta_{k,i}) - \Psi(v^{k,i}, \lambda^k, \theta_{k,i}).$$

So,

$$\text{Ared}_{k,i} = \theta_{k,i} [L(s^k, \lambda^k) - L(v^{k,i}, \lambda^k)] + (1 - \theta_{k,i}) [|C(s^k)| - |C(v^{k,i})|].$$

However, merely a reduction of the merit function is not sufficient to guarantee convergence. In fact, we need a sufficient reduction of the merit function, which will be defined by the satisfaction of the following test:

$$\text{Ared}_{k,i} \geq 0.1 \text{Pred}_{k,i},$$

where $\text{Pred}_{k,i}$ is a positive predicted reduction of the merit function $\Psi(s, \lambda, \theta)$ between s^k and $v^{k,i}$. It is defined by

$$\text{Pred}_{k,i} = \theta_{k,i} [L(s^k, \lambda^k) - L(v^{k,i}, \lambda^k) - C(z^k)^T(\lambda^{k,i} - \lambda^k)] + (1 - \theta_{k,i}) [|C(s^k)| - |C(z^k)|].$$

The quantity $\text{Pred}_{k,i}$ defined above can be nonpositive depending on the value of the penalty parameter. Fortunately, if $\theta_{k,i}$ is small enough, $\text{Pred}_{k,i}$ is arbitrarily close to

$[|C(s^k)| - |C(z^k)|]$, which is necessarily nonnegative. Therefore, we will always be able to choose $\theta_{k,i} \in (0, 1]$ such that

$$\text{Pred}_{k,i} \geq \frac{1}{2} [|C(s^k)| - |C(z^k)|].$$

When the criterion $\text{Ared}_{k,i} \geq 0.1\text{Pred}_{k,i}$ is satisfied, we accept $v^{k,i} = z^k$. Otherwise, we reduce the trust-region radius.

To establish IR methods for semivectorial bilevel programming on Riemannian manifolds, we adapt the IR method presented in [4]. In the presented algorithm, the parameters $\eta > 0, N > 0, \theta_{-1} \in (0, 1), \delta_{\min} > 0, \tau_1 > 0$, and $\tau_2 > 0$ are given. The initial approximations $s^0 \in W \times M \times \mathbb{R}^n, \lambda^0 \in \mathbb{R}^{m+n}$, as well as a sequence $\{\omega^k\}$ such that $\sum_{k=0}^{\infty} \omega^k < +\infty$ are also given.

4. Convergence Results

Using the method for studying the convergence of the IR algorithm in Euclidean spaces [20,22], the convergence results of IR algorithms for semivectorial bilevel programming on Riemannian manifolds are given under the following assumptions. From now on, we assume that the semivectorial bilevel optimization problems on Riemannian manifolds satisfy assumptions H_1 – H_3 stated below:

H_1 There exists L_1 such that, for all $(x, \omega), (\bar{x}, \bar{\omega}) \in M \times W, \mu, \bar{\mu} \in \mathbb{R}^n$, and $\zeta \in [0, \zeta_{\max}]$,

$$|C'(x, \omega, \mu) - C'(\bar{x}, \bar{\omega}, \bar{\mu})| \leq L_1 d((x, \omega, \mu), (\bar{x}, \bar{\omega}, \bar{\mu})).$$

H_2 There exists L_2 such that, for all $x, \bar{x} \in M$,

$$|\text{gard}_x F(x) - \text{gard}_x F(\bar{x})| \leq L_2 d((x, \bar{x})).$$

H_3 There exists $r \in [0, 1)$, independently of k , such that the point $z^k = (\bar{x}, \bar{\omega}, \bar{\mu})$ obtained at the restoration phase satisfies

$$|C(z^k)| \leq r|C(s^k)|,$$

where $s^k = (x^k, \omega^k, \mu^k)$. Moreover, if $C(s^k) = 0$, then $z^k = s^k$.

Theorem 1 (Well-definiteness). Under assumptions H_1 – H_3 , IR Algorithm 1 for bilevel programming is well defined.

Algorithm 1: Inexact Restoration algorithm

- 1 Define $\theta_k^{\min} = \min\{1, \theta_{k-1}, \dots, \theta_{-1}\}, \theta_k^{\text{large}} = \min\{1, \theta_k^{\min} + \omega^k\}$, and $\theta_{k,-1} = \theta_k^{\text{large}}$.
- 2 **(Restoration phase)** Find an approximate minimizer \bar{x} and multipliers $\bar{\mu} \in \mathbb{R}^n$ for the problem:

$$\begin{aligned} \min_x \quad & \sum_{i=1}^p \omega_i^k f_i(x) \\ \text{s.t.} \quad & h(x) = 0, \\ & x \in M, \end{aligned}$$

and define $z^k = (\bar{x}, \omega^k, \bar{\mu})$.

Algorithm 1: Cont.

3 (Direction) Compute

$$d_{\text{tan}}^k = P_k \left(\exp_{z^k} \left(-\eta \text{grad}_s L(z^k, \lambda^k) \right) \right) - z^k,$$

where P_k is the projection on

$$\pi_k = \{s \in M \times W \times \mathbb{R}^n : \langle C'(z^k), \hat{\gamma}_{s,z^k}(0) \rangle = 0\},$$

and $P_k \left(\exp_{z^k} \left(-\eta \text{grad}_s L(z^k, \lambda^k) \right) \right)$ is a solution of the following problem:

$$\begin{aligned} \min_{y \in M \times W \times \mathbb{R}^n} \quad & \frac{1}{2} \left\| y - \exp_{z^k} \left(-\eta \text{grad}_s L(z^k, \lambda^k) \right) \right\|^2 \\ \text{s.t.} \quad & \langle C'(z^k), \hat{\gamma}_{y,z^k}(0) \rangle = 0. \end{aligned}$$

If $z^k = s^k$, $d_{\text{tan}}^k = 0$, then stop and return x^k as a solution of Problem (7). Otherwise, we set $i \leftarrow 0$ and choose $\delta_{k,0} \geq \delta_{\text{min}}$.

4 (Minimization phase) If $d_{\text{tan}}^k = 0$, then we take $v^{k,i} = z^k$. Otherwise, we take $t_{\text{break}}^{k,i} = \min \left\{ 1, \frac{\delta_{k,i}}{d_{\text{tan}}^k} \right\}$ and find $v^{k,i} \in \pi_k$ such that, for some $0 < t < t_{\text{break}}^{k,i}$, we have

$$L(v^{k,i}, \lambda^k) \leq \max \left\{ L(z^k + td_{\text{tan}}^k, \lambda^k), L(z^k, \lambda^k) - \tau_1 \delta_{k,i}, L(z^k, \lambda^k) - \tau_2 \right\}$$

and $d(v^{k,i}, z^k) \leq \delta_{k,i}$.

5 If $d_{\text{tan}}^k = 0$, define $\lambda^{k,i} = \lambda^k$. Otherwise, we take $\lambda^{k,i} \in \mathbb{R}^{n+m}$ such that $|\lambda^{k,i}| \leq N$.

6 For all $\theta \in [0, 1]$, we define

$$\begin{aligned} \text{Pred}_{k,i}(\theta) = & \theta \left[L(s^k, \lambda^k) - L(v^{k,i}, \lambda^k) - C(z^k)^T (\lambda^{k,i} - \lambda^k) \right] + \\ & (1 - \theta) \left[|C(s^k)| - |C(z^k)| \right]. \end{aligned}$$

We take $\theta_{k,i}$ as the maximum $\theta \in [0, \theta_{k,i-1}]$ that it satisfies:

$$\text{Pred}_{k,i}(\theta) \geq \frac{1}{2} \left[|C(s^k)| - |C(z^k)| \right], \tag{12}$$

and define $\text{Pred}_{k,i} = \text{Pred}_{k,i}(\theta_{k,i})$.

Algorithm 1: *Cont.*

7 Compute

$$\text{Ared}_{k,i} = \theta_{k,i} [L(s^k, \lambda^k) - L(v^{k,i}, \lambda^{k,i})] + (1 - \theta_{k,i}) [|C(s^k)| - |C(v^{k,i})|].$$

If

$$\text{Ared}_{k,i} \geq 0.1 \text{Pred}_{k,i},$$

then we take

$$s^{k+1} = v^{k,i}, \quad \lambda^{k+1} = \lambda^{k,i}, \quad \theta_k = \theta_{k,i}, \quad \delta_k = \delta_{k,i},$$

$$\text{Ared}_k = \text{Ared}_{k,i}, \quad \text{Pred}_k = \text{Pred}_{k,i}.$$

and finish the current k th iteration. Otherwise, we choose $\delta_{k,i+1} \in [0.1\delta_{k,i}, 0.9\delta_{k,i}]$, set $i \leftarrow i + 1$, and go to Step 4.

Proof. According to Step 6 and Step 7 of Algorithm 1, it can be calculated that

$$\begin{aligned} \text{Ared}_{k,i} - 0.1 \text{Pred}_{k,i} &= 0.9 \text{Pred}_{k,i} + (1 - \theta_{k,i}) [|C(z^k)| - |C(v^{k,i})|] \\ &\quad + \theta_{k,i} [L(v^{k,i}, \lambda^k) - L(v^{k,i}, \lambda^{k,i}) + C(z^k)^T (\lambda^{k,i} - \lambda^k)] \\ &= 0.9 \text{Pred}_{k,i} + (1 - \theta_{k,i}) [|C(z^k)| - |C(v^{k,i})|] \\ &\quad + \theta_{k,i} [C(v^{k,i})^T \lambda^k - C(v^{k,i})^T \lambda^{k,i} + C(z^k)^T (\lambda^{k,i} - \lambda^k)] \\ &= 0.9 \text{Pred}_{k,i} + (1 - \theta_{k,i}) [|C(z^k)| - |C(v^{k,i})|] \\ &\quad + \theta_{k,i} (C(z^k) - C(v^{k,i}))^T (\lambda^{k,i} - \lambda^k). \end{aligned}$$

Through the condition (12), we have

$$\begin{aligned} \text{Ared}_{k,i} - 0.1 \text{Pred}_{k,i} &\geq 0.45 [|C(s^k)| - |C(z^k)|] + (1 - \theta_{k,i}) [|C(z^k)| - |C(v^{k,i})|] \\ &\quad + \theta_{k,i} (C(z^k) - C(v^{k,i}))^T (\lambda^{k,i} - \lambda^k). \end{aligned} \tag{13}$$

Then, from the assumption H_3 ,

$$\begin{aligned} \text{Ared}_{k,i} - 0.1 \text{Pred}_{k,i} &= 0.45(1 - r) |C(s^k)| + (1 - \theta_{k,i}) [|C(z^k)| - |C(v^{k,i})|] \\ &\quad + \theta_{k,i} (C(z^k) - C(v^{k,i}))^T (\lambda^{k,i} - \lambda^k). \end{aligned}$$

If $C(s^k) \neq 0$, due to the continuity of C and $\delta_{k,i} \rightarrow 0$, we have $|C(z^k)| - |C(v^{k,i})| \rightarrow 0$. Thus, there exists a positive constant $\delta_{k,i}$ such that

$$\text{Ared}_{k,i} - 0.1 \text{Pred}_{k,i} \geq 0.$$

This means that the algorithm is well defined when $C(s^k) \neq 0$.

If $C(s^k) = 0$, then s^k is feasible. Since the algorithm does not terminate at the k th iteration, we know that $d_{\text{tan}}^k \neq 0$. Therefore, we have

$$z^k = s^k \text{ and } C(z^k) = C(s^k) = 0.$$

Combining the condition (12), it follows that

$$\text{Pred}_{k,i}(\theta) = \theta \left[L(s^k, \lambda^k) - L(\vartheta^{k,i}, \lambda^k) \right] \geq 0,$$

and independent of θ , for all i , $\theta_{k,i} = \theta_{k,-1}$. In terms of the inequality (13), when $\delta_{k,i}$ is sufficiently small, we obtain

$$\text{Ared}_{k,i} - 0.1\text{Pred}_{k,i} \geq 0.$$

Therefore, Algorithm 1 is well defined. \square

The next theorem is an important tool for proving the convergence of Algorithm 1. We prove that the actual reduction Ared_{k,i^*} , with i^* the accepted value of i , achieved at each iteration necessarily tends to 0.

Theorem 2. Under the assumptions H_1 – H_3 , if Algorithm 1 generates an infinite sequence, then

$$\lim_{k \rightarrow +\infty} \text{Ared}_k = 0, \quad \lim_{k \rightarrow +\infty} |C(s^k)| = 0.$$

The same results above occur when $\lambda^k = 0$, for all k .

Proof. Let us prove that $\lim_{k \rightarrow +\infty} \text{Ared}_k = 0$, i.e., we need to prove

$$\lim_{k \rightarrow +\infty} \left[\theta_k \left[L(s^k, \lambda^k) - L(s^{k+1}, \lambda^{k+1}) \right] + (1 - \theta_k) \left[|C(s^k)| - |C(s^{k+1})| \right] \right] = 0,$$

that is

$$\lim_{k \rightarrow +\infty} \left[\theta_k L(s^k, \lambda^k) + (1 - \theta_k) |C(s^k)| - \left[\theta_k L(s^{k+1}, \lambda^{k+1}) + (1 - \theta_k) |C(s^{k+1})| \right] \right] = 0,$$

namely

$$\lim_{k \rightarrow +\infty} \left[\Psi(s^k, \theta_k) - \Psi(s^{k+1}, \theta_k) \right] = 0,$$

where $\Psi(s^k, \theta_k) = \theta_k L(s^k, \lambda^k) + (1 - \theta_k) |C(s^k)|$.

By contradiction, suppose that there is an infinite indicator set $T_1 \subset \{0, 1, 2, \dots\}$ and a positive constant $\zeta > 0$ such that, for any $k \in T_1$, we have

$$\Psi(s^{k+1}, \theta_k) \leq \Psi(s^k, \theta_k) - \zeta.$$

Let $\Psi_k = \Psi(s^k, \theta_k)$, then

$$\begin{aligned} \Psi_{k+1} &= \theta_{k+1} L(s^{k+1}, \lambda^{k+1}) + (1 - \theta_{k+1}) |C(s^{k+1})| \\ &= \theta_{k+1} L(s^{k+1}, \lambda^{k+1}) + (1 - \theta_{k+1}) |C(s^{k+1})| \\ &\quad - \theta_k L(s^{k+1}, \lambda^{k+1}) + (1 - \theta_k) |C(s^{k+1})| \\ &\quad + \theta_k L(s^{k+1}, \lambda^{k+1}) + (1 - \theta_k) |C(s^{k+1})| \\ &= (\theta_{k+1} - \theta_k) L(s^{k+1}, \lambda^{k+1}) + (\theta_k - \theta_{k+1}) |C(s^{k+1})| \\ &\quad + \theta_k L(s^{k+1}, \lambda^{k+1}) + (1 - \theta_k) |C(s^{k+1})| \\ &\leq (\theta_k - \theta_{k+1}) \left[|C(s^{k+1})| - L(s^{k+1}, \lambda^{k+1}) \right] \\ &\quad + \theta_k L(s^k, \lambda^k) + (1 - \theta_k) |C(s^k)| - \zeta_k. \end{aligned}$$

Equivalently,

$$\Psi_{k+1} \leq (\theta_k - \theta_{k+1}) \left[|C(s^{k+1})| - L(s^{k+1}, \lambda^{k+1}) \right] + \Psi_k - \zeta_k, \tag{14}$$

where $\zeta_k > 0$ and $\zeta_k > \zeta > 0, k \in T_1$.

According to the definition of $\theta_{k,-1}$,

$$\theta_k - \theta_{k+1} + \omega_k \geq 0, \quad k \in T_1.$$

There is an upper bound $c > 0$, such that

$$|C(s^k)| - \|L(s^{k+1}, \lambda^{k+1})\| \leq c. \tag{15}$$

Combining the inequalities (14) and (15), it follows that

$$\begin{aligned} \Psi_{j+1} &\leq (\theta_j - \theta_{j+1} + \omega_j) [|C(s^{j+1})| - L(s^{j+1}, \lambda^{j+1})] \\ &\quad + \Psi_j - \zeta_j - \omega_j [|C(s^{j+1})| - L(s^{j+1}, \lambda^{j+1})] \\ &\leq (\theta_j - \theta_{j+1} + \omega_j)c + \Psi_j - \zeta_j + \omega_jc \\ &\leq (\theta_j - \theta_{j+1})c + \Psi_j - \zeta_j + 2\omega_jc. \end{aligned}$$

Then, for all $k \geq 1$, we have

$$\begin{aligned} \Psi_k &\leq \Psi_0 + (\theta_0 - \theta_{k+1})c + - \sum_{j=0}^{k-1} \zeta_j + \sum_{j=0}^{k-1} 2\omega_jc \\ &\leq \Psi_0 + 2c + - \sum_{j=0}^{k-1} \zeta_j + \sum_{j=0}^{k-1} 2\omega_jc. \end{aligned}$$

Since $\sum_{j=0}^{k-1} 2\omega_j$ is the convergence and ζ_j is bounded away from zero, this implies that Ψ_k is unbounded. This is a contradiction. Thus, we have that $\lim_{k \rightarrow +\infty} \text{Ared}_k = 0$. In addition, in a similar way, we can prove $\lim_{k \rightarrow +\infty} |C(s^k)| = 0$. \square

According to Theorem 2, it means that the point generated by the IR algorithm for the KKT transformation (7) will converge to a feasible point eventually. Then, we prove that d_{tan}^k cannot be bounded away from zero under the following assumption H_4 . This means that the point generated by the IR algorithm will converge to a weak Pareto solution of Problem (7):

H_4 There exists $\beta > 0$, independently of k , such that

$$d(s^k, z^k) \leq \beta |C(s^k)|.$$

Theorem 3. Suppose that the assumptions H_1, H_2, H_3 , and H_4 hold. If $\{s^k\}$ is an infinite sequence generated by Algorithm 1, $\{z^k\}$ is the sequence defined at the restoration phase in Algorithm 1, then:

- 1 $|C(s^k)| \rightarrow 0$.
- 2 There exists a limit point s^* of $\{s^k\}$.
- 3 Every limit point of $\{s^k\}$ is a feasible point of the KKT reformulation (7).
- 4 If, for all ω , a global solution of the lower-level problem is found, then any limit point (x^*, ω^*) is feasible for the weighted semivectorial bilevel programming (6).
- 5 If s^* is a limit point of $\{s^k\}$, there exists an infinite set $K \subset \mathbb{N}$ such that

$$\lim_{k \in K} s^k = \lim_{k \in K} z^k = s^*, \quad C(s^*) = 0, \quad \lim_{k \in K} d_{tan}^k = 0.$$

Proof. We can prove the first two items from Theorem 2 and the assumption H_1-H_3 . Based on the conclusions of the first two terms, the third and fourth items are valid. The fifth item follows from the assumption H_4 and the first item. \square

The above conclusions give the well-definiteness and convergence of the algorithm proposed for semivectorial bilevel programming on Riemannian manifolds. From the point of view of the assumption put forward in this paper, the assumptions H_3 and H_4 are related to the sequences generated by the IR algorithm. Therefore, it is worth studying establishing sufficient conditions to ensure their effectiveness. Two assumptions about the lower-level problem are given below to verify the hypotheses H_3 and H_4 :

H_5 For every solution $s = (x, \omega, \mu)$ of $C(x, \omega, \mu) = 0$, such that the gradients $\text{grad}h_i(x)$, $i = 1, \dots, n$ of the active lower level constraints are linearly independent.

H_6 For every solution $s = (x, \omega, \mu)$ of $C(x, \omega, \mu) = 0$ such that the matrix:

$$H(x, \omega, \mu) = \sum_{i=1}^p \omega_i \text{Hess}_x f_i(x) + \sum_{i=1}^n \mu_i \text{Hess}_x h_i(x),$$

is positive definite in the following set:

$$Z(x) = \{d \in \mathbb{R}^n \mid \text{grad}h(x)^T d = 0, d_j = 0 \text{ for all } j\}.$$

For convenience, to verify H_3 and H_4 , we define the following matrix:

$$D'(s) = \begin{pmatrix} \sum_{i=1}^p \omega_i \text{Hess}_x f_i + \sum_{i=1}^n \mu_i \text{Hess}_x h_i & \text{grad}_x h \\ \text{grad}_x h^T & 0 \end{pmatrix}.$$

Lemma 1. The matrix $D'(s)$ is non-singular for any solution $s = (x, \omega, \mu)$ of $C(x, \omega, \mu) = 0$.

Proof. Assuming that there exist $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^p$ such that

$$D'(s) \begin{pmatrix} u \\ v \end{pmatrix} = 0,$$

then we have

$$\left(\sum_{i=1}^p \omega_i \text{Hess}_x f_i + \sum_{i=1}^n \mu_i \text{Hess}_x h_i \right) u + \text{grad}_x h v = 0, \tag{16}$$

$$\text{grad}_x h u = 0. \tag{17}$$

According to the assumptions H_5 – H_6 and Equalities (16) and (17), it follows that $u = 0$ and $v = 0$. This means that the matrix $D'(s)$ is non-singular for any solution $s = (x, \omega, \mu)$ of $C(x, \omega, \mu) = 0$. \square

Let $D(s)$ be defined on $M \times W \times \mathbb{R}^n$, for each $\omega \in W$, a solution $u(\omega) = (x(\omega), \mu(\omega))$ of $C(x, \omega, \mu) = 0$ such that the function $v(\omega) = u(\omega)$ is continuous on W . Now, we fix the function $v(\omega)$, by Lemma 1, and we can define a function $Y(\omega) = D'(\omega, v(\omega))^{-1}$ over the set W . Let $V(v(\omega), \alpha) = \{v \in M \times \mathbb{R}^n : d(v, v(\omega)) \leq \alpha\}$. Furthermore, the following lemma can be obtained.

Lemma 2. There exist $\alpha > 0$ and $\beta > 0$, such that, for all $\omega \in W$, it holds $|Y(\omega)| < \beta$, and for all $v \in V(v(\omega), \alpha)$, $Y(\omega)$ coincides with the local inverse operator of $D'(\omega, \cdot)$.

Proof. Since $D'(\omega, v)$ is continuous on (ω, v) , $v(\omega)$ is continuous on W , and $Y(\omega)$ is continuous with respect to $\omega \in W$, there exists $\beta > 0$, such that, for all $\omega \in W$, $|Y(\omega)| < \beta$.

For each fixed value of $\omega \in W$, associated with each v , the continuously differentiable operator of the vector $C(\omega, v)$ verifies the assumption of the inverse function theorem at $v(\omega)$. Hence, there exists $\alpha > 0$ such that $C(\omega, \cdot)$ has a continuously differentiable local

inverse operator $G(\omega) : C(\omega, V(v(\omega), \alpha)) \mapsto V(v(\omega), \alpha)$, and the Jacobian matrix $[G(\omega)]'$ is consistent with $Y(\omega)$. This ends the proof. \square

Finally, we state that H_3 and H_4 hold under the assumptions H_5 to H_6 . The next theorem summarizes this fact, and it can be proven as follows.

Theorem 4. *Let $r \in [0, 1)$, $(\omega, u) \in W \times M \times \mathbb{R}^n$ be such that $C(\omega, u) \neq 0$. If the assumptions H_5 - H_6 hold, then there exist $\beta > 0$, $\omega \in W$, and $\bar{u} = (\bar{x}, \bar{\mu}) \in M \times \mathbb{R}^n$ such that*

$$|C(\omega, \bar{u})| \leq r|C(\omega, u)|,$$

and

$$d((\omega, u), (\omega, \bar{u})) \leq \beta|C(\omega, u)|.$$

Proof. According to Lemmas 1 and 2, combining the assumptions H_5 and H_6 , by using Taylor expansions of the functions on Riemannian manifolds, the statement follows from the results of [20]. This ends the proof. \square

Example 1. *We consider the particular case $M = \mathbb{R}_+^2 := \{(x_1, x_2) \in \mathbb{R}^2 | x_1 > 0, x_2 > 0\}$ with the metric g given in Cartesian coordinates (x_1, x_2) around the point $x \in M$ by the matrix:*

$$M \ni y \mapsto (g_{ij})_y = \left(g\left(\frac{\partial}{\partial y_i}, \frac{\partial}{\partial y_j}\right) \right) := \text{diag}(x_1^{-1}, x_2^{-1}).$$

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

$$g(u, v) = \frac{u_1 v_1}{x_1} + \frac{u_2 v_2}{x_2}.$$

Let $a = (a_1, a_2) \in M$ and $v = (v_1, v_2) \in T_a M$. It is easy to see that the (minimizing) geodesic curve $t \mapsto \gamma(t)$ verifying $\gamma(0) = a$, $\dot{\gamma}(0) = v$ 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 is a complete Riemannian manifold. Furthermore, the (minimizing) geodesic segment $\gamma : [0, 1] \rightarrow M_2$ 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 $\gamma_i(t) = a_i^{1-t} b_i^t$, $i = 1, 2$. Thus, the distance d on the metric space (M_2, g_2) is given by

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

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

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

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, g) with the distance d .

Next, we consider the functions $F : M \rightarrow \mathbb{R}$, $f : M \rightarrow \mathbb{R}^2$ and $h : M \rightarrow \mathbb{R}$ given for any $x \in M$ by

$$\begin{aligned} F(x) &= -x_1, \\ f_1(x) &= \frac{1}{2}(x_1 - 1)^2 - \frac{3}{4} \ln x_1 + \frac{3}{8}(x_2 - 1)^2, \\ f_2(x) &= \frac{1}{4}(x_1 - 1)^2 - \frac{3}{8} \ln x_1 + \frac{3}{16}(x_2 - 1)^2, \\ h(x) &= \frac{1}{3}(x_1 - 1)^2 + \frac{1}{3}(x_2 - 1)^2 - \frac{1}{3}. \end{aligned}$$

It is easy to see that, for $x \in M$ and any geodesic segment $\gamma : [0, 1] \rightarrow M$ with $\gamma(0) = a$, $\gamma(1) = b$, the functions $f_i(x), i = 1, 2$, and $h(x)$ are all convex on M with the Riemannian metric g . Moreover, the function $h(x)$ satisfies the Slater constraint qualification.

We then consider the corresponding KKT reformulation of the semivectorial bilevel programming on Riemannian manifolds:

$$\begin{aligned} \min_{x, \omega} \quad & F(x) = -x_2 \\ \text{s.t.} \quad & \omega \in W, \\ & \sum_{i=1}^2 \omega_i \text{grad}_x f_i(x) + \text{grad}_x h(x) \mu = 0, \\ & h(x) = 0, \\ & x \in M. \end{aligned}$$

By the definition of the gradient of a differentiable function with respect to the Riemannian metric g , let $\omega_1 = \frac{1}{3}$, $\omega_2 = \frac{2}{3}$, $\omega_1 + \omega_2 = 1$, and $\mu = (\frac{1}{2}, \frac{3}{4})^T \in \mathbb{R}^2$; we have

$$\begin{aligned} \min_{x, \omega} \quad & F(x) = -x_1 \\ \text{s.t.} \quad & (x_1 - \frac{1}{2})^2 + (x_2 - \frac{1}{2})^2 - 1 = 0, \\ & \frac{1}{3}(x_1 - 1)^2 + \frac{1}{3}(x_2 - 1)^2 - \frac{1}{3} = 0, \\ & x \in M. \end{aligned}$$

It is easy to see that the unique optimal solution of the KKT reformulation is $x = (\frac{3-\sqrt{7}}{4}, \frac{3+\sqrt{7}}{4})$.

According to Algorithm 1, we first give the initial approximations $s^0 \in W \times M \times \mathbb{R}^2$, $\lambda^0 \in \mathbb{R}^2$, and a sequence $\{\omega^k\}$. In the restoration phase, find an approximate minimizer $\bar{x} = (\bar{x}_1, \bar{x}_2) \in M$ and multiplier $\bar{\mu} = (\bar{\mu}_1, \bar{\mu}_2) \in \mathbb{R}^2$ for the problem:

$$\begin{aligned} \min_x \quad & \omega_1^k f_1(x) + \omega_2^k f_2(x) \\ \text{s.t.} \quad & h(x) = 0, \\ & x \in M, \end{aligned}$$

and define $z^k = (\bar{x}, \omega^k, \bar{\mu})$.

We then compute the direction by using the exponential mapping and the projection defined on Riemannian manifold M .

$$\begin{aligned} d_{\text{tan}}^k &= P_k \left(\exp_{z^k} \left(-\eta \text{grad}_s L(z^k, \lambda^k) \right) \right) - z^k, \\ &= P_k \left(z_1^k e^{-\eta \frac{\text{grad}_s L(z^k, \lambda^k)}{z_1^k}}, z_2^k e^{-\eta \frac{\text{grad}_s L(z^k, \lambda^k)}{z_2^k}} \right) - z^k, \end{aligned}$$

where $L(z^k, \lambda^k) = -x_1 + \lambda_1^k \left(\sum_{i=1}^2 \omega_i^k \text{grad}_{s f_i}(\bar{x}) + \text{grad}_s h(\bar{x}) \bar{\mu} \right) + \lambda_2^k h(\bar{x})$.

In the minimization phase, we first find $v^{k,i}$ such that $L(v^{k,i}, \lambda^k) < L(z^k, \lambda^k)$ and $v^{k,i} \in B_{k,i} = \{v : d(v, z^k) \leq \delta_{k,i}\}$. Then, by calculating the actual reduction $\text{Ared}_{k,i}$ and positive predicted reduction $\text{Pred}_{k,i}$ of the merit function $\Psi(s, \lambda, \theta)$ such that $\text{Ared}_{k,i} \geq 0.1 \text{Pred}_{k,i}$, we obtain a sequence $\{s^k\}$.

According to Theorems 3 and 4, the sequence $\{s^k\}$ generated by the IR method established in the present paper converges to a solution of the semivectorial bilevel programming on Riemannian manifolds.

5. Conclusions

In this paper, a new algorithm for solving the semivectorial bilevel programming based on the IR technique was proposed, which preserves the two-stage structure of the problem. In the feasibility phase, lower-level problems can be solved imprecisely using their properties, and users are free to use special-purpose solvers. In the optimal stage, a minimization algorithm with linear constraints was used. Moreover, it was also proven that the algorithm is well-defined and converges to the feasible point under mild conditions. Under more stringent assumptions, the convergence of sequences generated by the presented algorithm was proven. Furthermore, the validity of some conditions generated by the algorithm was given as well.

Author Contributions: Conceptualization, J.L. and Z.W.; methodology, J.L.; writing—original draft preparation, J.L.; writing—review and editing, J.L.; supervision, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of China Grant Number 11871383.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Udriște, C. *Convex Functions and Optimization Methods on Riemannian Manifolds*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013; Volume 297.
2. Boumal, N. An Introduction to Optimization on Smooth Manifolds. 2020. Available online: <https://www.nicolasboumal.net/book/> (accessed on 11 September 2020).
3. Liao, L.; Wan, Z. On the Karush–Kuhn–Tucker reformulation of the bilevel optimization problems on Riemannian manifolds. *Filomat* **2022**, *36*, accepted and in press.
4. Martínez, J.M.; Pilotta, E.A. Inexact-restoration algorithms for constrained optimization. *J. Optim. Theory Appl.* **2000**, *104*, 135–163. [[CrossRef](#)]
5. Martínez, J.M. Inexact-restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming. *J. Optim. Theory Appl.* **2001**, *111*, 39–58. [[CrossRef](#)]
6. Birgin, E.G.; Bueno, L.F.; Martínez, J.M. Assessing the reliability of general-purpose inexact restoration methods. *J. Comput. Appl. Math.* **2015**, *282*, 1–16. [[CrossRef](#)]
7. Bueno, L.F.; Friedlander, A.; Martínez, J.M.; Sobral, F.N. Inexact restoration method for derivative-free optimization with smooth constraints. *SIAM J. Optim.* **2013**, *23*, 1189–1213. [[CrossRef](#)]
8. Francisco, J.B.; Martínez, J.M.; Martínez, L.; Pisnitchenko, F. Inexact restoration method for minimization problems arising in electronic structure calculations. *Comput. Optim. Appl.* **2011**, *50*, 555–590. [[CrossRef](#)]
9. Banihashemi, N.; Kaya, C.Y. Inexact restoration for Euler discretization of box-constrained optimal control problems. *J. Optim. Theory Appl.* **2013**, *156*, 726–760. [[CrossRef](#)]
10. Bueno, L.F.; Haeser, G.; Martínez, J.M. An inexact restoration approach to optimization problems with multiobjective constraints under weighted-sum scalarization. *Optim. Lett.* **2016**, *10*, 1315–1325. [[CrossRef](#)]
11. Krejić, N.; Jerinkić, N.K.; Ostojić, T. An inexact restoration-nonsmooth algorithm with variable accuracy for stochastic nonsmooth convex optimization problems in machine learning and stochastic linear complementarity problems. *J. Comput. Appl. Math.* **2022**, *423*, 114943. [[CrossRef](#)]
12. Ma, Y.; Pan, B.; Yan, R. Feasible Sequential Convex Programming With Inexact Restoration for Multistage Ascent Trajectory Optimization. *IEEE T. Aero. Elec. Sys.* **2022**, 1–14. [[CrossRef](#)]
13. Gabay, D. Minimizing a differentiable function over a differential manifold. *J. Optim. Theory Appl.* **1982**, *37*, 177–219. [[CrossRef](#)]

14. Murtagh, B.A.; Saunders, M.A. Large-scale linearly constrained optimization. *Math. Program.* **1978**, *14*, 41–72. [[CrossRef](#)]
15. Gay, D.M. A trust-region approach to linearly constrained optimization. In *Numerical Analysis*; Springer: Berlin/Heidelberg, Germany, 1984; Volume 1066, pp. 72–105.
16. Gill, P.E.; Murray, W.; Wright, M.H. *Practical Optimization*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2019.
17. Fischer, A.; Friedlander, A. A new line search inexact restoration approach for nonlinear programming. *Comput. Optim. Appl.* **2010**, *46*, 333–346. [[CrossRef](#)]
18. Bueno, L.F.; Martínez, J.M. On the complexity of an inexact restoration method for constrained optimization. *SIAM J. Optim.* **2020**, *30*, 80–101. [[CrossRef](#)]
19. Andreani, R.; Ramos, A.; Secchin, L.D. Improving the Global Convergence of Inexact Restoration Methods for Constrained Optimization Problems. *Optimization-Online*, 2022. Available online: <https://optimization-online.org/?p=18821> (accessed on 28 March 2022).
20. Andreani, R.; Castro, S.L.C.; Chela, J.L.; Friedlander, A.; Santos, S.A. An inexact-restoration method for nonlinear bilevel programming problems. *Comput. Optim. Appl.* **2009**, *43*, 307–328. [[CrossRef](#)]
21. Friedlander, A.; Gomes, F.A.M. Solution of a truss topology bilevel programming problem by means of an inexact restoration method. *Comput. Appl. Math.* **2011**, *30*, 109–125.
22. Andreani, R.; Ramirez, V.A.; Santos, S.A.; Secchin, L.D. Bilevel optimization with a multiobjective problem in the lower level. *Numer. Algorithms* **2019**, *81*, 915–946. [[CrossRef](#)]
23. Martínez, J.M.; Pilotta, E.A. Inexact restoration methods for nonlinear programming: Advances and Perspectives. In *Optimization and Control with Applications*; Springer: Boston, MA, USA, 2005; pp. 271–291.
24. Fernández, D.; Pilotta, E.A.; Torres, G.A. An inexact restoration strategy for the globalization of the sSQP method. *Comput. Optim. Appl.* **2013**, *54*, 595–617. [[CrossRef](#)]
25. Dempe, S.; Kalashnikov, V.; Pérez-Valdés, G.A.; Kalashnykova, N. Bilevel programming problems. In *Energy Systems*; Springer: Berlin/Heidelberg, Germany, 2015.
26. Eisenhart, L.P. *Riemannian Geometry*; Princeton University Press: Princeton, NJ, USA, 1997; Volume 51.
27. Jost, J. *Riemannian Geometry and Geometric Analysis*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2008; Volume 42005.
28. Absil, P.A.; Mahony, R.; Sepulchre, R. *Optimization Algorithms on Matrix Manifolds*; Princeton University Press: Princeton, NJ, USA, 2009.
29. Bento, G.C.; Cruz Neto, J.X. A subgradient method for multiobjective optimization on Riemannian manifolds. *J. Optim. Theory Appl.* **2013**, *159*, 125–137. [[CrossRef](#)]