J. Nonlinear Var. Anal. 7 (2023), No. 5, pp. 839-857
Available online at http://jnva.biemdas.com
https://doi.org/10.23952/jnva.7.2023.5.10

THE GENERALIZED CONDITIONAL GRADIENT METHOD FOR COMPOSITE MULTIOBJECTIVE OPTIMIZATION PROBLEMS ON RIEMANNIAN MANIFOLDS

XIAOBO LI[1], XIAOCHUN GE[2], KAI TU[3,*]

[1]*School of Sciences, Civil Aviation Flight University of China, Guanghan, 618300, China*
[2]*School of Sciences, Southwest Petroleum University, Chengdu 610500, China*
[3]*School of Mathematical Sciences, Shenzhen University, Shenzhen 518060, China*

Abstract. In this paper, we consider a class of composite multiobjective optimization problems, subject to a closed convex constraint set, defined on Riemannian manifolds. To tackle this problem, we propose the generalized conditional gradient method with two step size strategies, including Armijo step size and the nonmonotone line search step size. Under some reasonable conditions, the global convergence result is established, and the iteration-complexity bound for composite multiobjective optimization problems is presented on Riemannian manifolds.
Keywords. Composite multiobjective optimization problems; Generalized conditional gradient method; Global convergence; Iteration-complexity bound.

1. Introduction

Many practical problems in real-world problems with conflicting objectives, economics, engineering, human decision making, architecture, and machine learning can be modelled as multiobjective optimization problems; see, for example, [1, 2, 3, 4, 5] and the references therein. The most straightforward approach to multiobjective optimization problems is linear scalarization approach, where multiobjective optimization problems are converted into scalar-valued ones. Thus, aiming to solve multiobjective optimization problems, we can apply standard, single-objective optimization algorithms to solve the resulting relaxed problem; see, for example, [4, 5, 6, 7, 8].

In recent years, some algorithms for multiobjective optimization problems have been studied on Riemannian manifolds; see, for example, [9, 10, 11, 12, 13, 14, 15, 16, 17, 18]. These algorithms include steepest descent method [13, 19], Newton's method [13, 19], projected gradient method [20], and trust-region method [21]. Due to the practical applicability of multiobjective optimization problems, it is imperative to explore new convergent algorithms for this new and growing area of research.

In this paper, we study the composite multiobjective optimization problem defined on Riemannian manifolds, where the objective function $F : M \to (\mathbb{R}\bigcup\{+\infty\})^m$, given by $F(x) :=$

*Corresponding author.
E-mail addresses: lixiaobo.double@163.com (X. Li), gxcmath@163.com (X. Ge), kaitu_02@163.com (K. Tu).
Received April 1, 2023; Accepted July 30, 2023.

©2023 Journal of Nonlinear and Variational Analysis

$(f_1(x),\cdots,f_m(x))$, has the following special separable structure:

$$f_j(x) := g_j(x) + h_j(x), \quad \forall j = 1,\cdots,m,$$

where M is a Riemannian manifold, $g_j : M \to \mathbb{R}\bigcup\{+\infty\}$ is proper, convex, and lower semi-continuous, and $h_j : M \to \mathbb{R}$ is continuous differentiable. We define $G : M \to (\mathbb{R}\bigcup\{+\infty\})^m$ by $G(x) := (g_1(x),\cdots,g_m(x))$ and $H : M \to \mathbb{R}^m$ by $H(x) := (h_1(x),\cdots,h_m(x))$, and denote this problem by

$$\min_{x\in M} F(x) := G(x) + H(x). \tag{1.1}$$

The set $\operatorname{dom}(G) := \{x \in M | g_j(x) < +\infty, j = 1,\cdots,m\}$ is assumed to be convex and compact in this paper.

In 2018, Bot and Grad [7] proposed a forward-backward proximal point type algorithm for problem (1.1) in Euclidean space. Tanabe et al. [8] designed a proximal gradient method for solving composite multiobjective optimization problems. Assunc*ão* et al. [6] extended the conditional gradient method, known as Frank-Wolfe algorithm, for composite multiobjective optimization problems. The conditional gradient method was exhibited for the convex linear system in [22]. Moreover, they proved that the conditional gradient method applied to the equivalent minimization formulation of the convex linear system, converges to a solution at a linear rate. In [23], a generalized conditional gradient method and its connection to an iterative shrinkage method was investigated, and the convergence of this generalized method for general class of functions was proved. Based on the results of [6, 8, 22, 23], Assunc*ão* et al. [24] proposed the generalized version of the conditional gradient method for composite multiobjective optimization problems with three step size strategies, including Armijo type, adaptive and diminishing step sizes. Moreover, some asymptotic convergence properties and iteration-complexity bounds were established in [24].

Motivated by the results described above, in this paper, we design the generalized conditional gradient method to solve composite multiobjective optimization problems with Armijo and nomonotone line search step sizes on Riemannian manifolds. Under some reasonable conditions, we establish the global convergence of the generalized gradient method for composite multiobjective optimization problems. The iteration-complexity bound on the objective function is also established. Since a Riemannian manifold, in general, does not have a linear structure, usual techniques in the Euclidean space cannot be applied and new techniques have to be developed. Our results are distinguished from the following aspects: First, we generalize the generalized conditional gradient method in [24] from $\mathbb{R}^n$ to Riemannian manifolds; Second, our results can be regarded as a generalization of [20] from smooth objective functions to non-smooth objective functions on Riemannian manifolds; Third, exponential mappings and parallel transports in [20] are extended to retractions and isometric vector transports in this paper, respectively, which makes the method more efficient.

Our work is organized as follows: In Section 2, some necessary definitions and concepts are provided on Riemannian manifolds. In Section 3, we present the generalized conditional gradient method for composite multiobjective optimization problems on Riemannian manifolds. In Section 4, under some reasonable conditions, the global convergence result of the generalized conditional gradient method for composite multiobjective optimization problems is provided, and the iteration-complexity bound is established on Riemannian manifolds. The last section, Section 5, ends this paper with a conclusion.

2. PRELIMINARIES

In this section, we recall some definitions and results from Riemannian manifolds, which can be found in some introductory books on Riemannian geometry, see for example [11, 15].

Let M be a finite-dimensional differentiable manifold. Given $x \in M$, the tangent space of M at x is denoted by T_xM and the tangent bundle of M by $TM = \bigcup_{x\in M} T_xM$. We denote by $\langle , \rangle_x$ the inner product on T_xM with the associated norm $\|\cdot\|_x$. If there is no confusion, then we omit the subscript x. If M is endowed with a Riemannian metric g, then M is a Riemannian manifold. Given a piecewise smooth curve $\gamma : [t_0,t_1] \to M$ joining x to y, that is, $\gamma(t_0) = x$ and $\gamma(t_1) = y$, we can define the length of γ by $l(\gamma) = \int_a^b \|\gamma'(t)\|dt$. By minimizing $l(\gamma)$ over the set of all curves, we can obtain a Riemannian distance $d(x,y)$ which induces the original topology on M.

A Riemannian manifold is complete if for any $x \in M$, all geodesic emanating from x are defined for all $t \in \mathbb{R}$. By Hopf-Rinow theorem [17], any pair of points $x,y \in M$ can be joined by a minimal geodesic. The exponential mapping $\exp_x : T_xM \to M$ is defined by $\exp_x v = \gamma_v(1,x)$ for each $v \in T_xM$, where $\gamma(\cdot) = \gamma_v(\cdot,x)$ is the geodesic starting x with velocity v, i.e., $\gamma(0) = x$ and $\gamma'(0) = v$. It is easy to see that $\exp_x tv = \gamma_v(t,x)$ for each real number t.

The systematic use of the exponential mapping may not be desirable in all cases. Some local mappings to T_xM may reduce the computational cost while preserving the useful convergence property of the considered method.

Definition 2.1. [19] Given $x \in M$, a retraction is a smooth mapping $R_x : T_xM \to M$ such that

(i) $R_x(0_x) = x$ for all $x \in M$, where 0_x denotes the zero element of T_xM;
(ii) $DR_x(0_x) = \mathrm{id}_{T_xM}$, where DR_x denotes the derivative of R_x and id denotes the identity mapping.

It is well-known that the exponential mapping is a special retraction, and some retractions are approximations of the exponential mapping.

The parallel transport is often too expensive to compute in a practical method. The generalized vector transport, built upon the retraction R_x, has thus been proposed in [14, 16]. A vector transport $\mathscr{T} : TM \bigoplus TM \to TM$, $(\eta_x, \xi_x) \mapsto \mathscr{T}_{\eta_x}\xi_x$ with the associated retraction R_x is a smooth mapping such that, for all η_x in the domain of R_x and all $\xi_x, \zeta_x \in T_xM$, (i) $\mathscr{T}_{\eta_x}\xi_x \in T_{R_x(\eta_x)}M$; (ii) $\mathscr{T}_{0_x}\xi_x = \xi_x$; (iii) $\mathscr{T}_{\eta_x}$ is a linear mapping. Let $\mathscr{T}_S$ be the isometric vector transport [16] with R_x as the associated retraction. Then it satisfies (i), (ii), (iii), and

$$\text{(iv)} \qquad g(\mathscr{T}_{S(\eta_x)}\xi_x, \mathscr{T}_{S(\eta_x)}\zeta_x) = g(\xi_x, \zeta_x).$$

In most practical cases, $\mathscr{T}_{S(\eta_x)}$ exists for all $\eta_x \in T_xM$, and we make this assumption throughout the paper. Furthermore, let $\mathscr{T}_{\eta_x}$ be the derivative of the retraction, i.e.,

$$\mathscr{T}_{\eta_x}\xi_x = \mathrm{D}R_x(\eta_x)[\xi_x] = \frac{\mathrm{d}}{\mathrm{d}t}R_x(\eta_x + t\xi_x)|_{t=0}.$$

Let $\mathfrak{L}(TM,TM)$ be the fiber bundle with base space $M \times M$ such that the fiber over $(x,y) \in M \times M$ is $\mathfrak{L}(T_xM, T_yM)$. We recall from [5, Section 4] that a transporter $\mathfrak{L}$ on M is a smooth section of the bundle $\mathfrak{L}(TM,TM)$. Furthermore, $\mathfrak{L}^{-1}(x,y) = \mathfrak{L}(y,x)$ and $\mathfrak{L}(x,z) = \mathfrak{L}(y,z)\mathfrak{L}(x,y)$. Given a retraction R_x, for any $\eta_x, \xi_x \in T_xM$, the isometric vector transport $\mathscr{T}_S$ is defined by

$$\mathscr{T}_{S(\eta_x)}\xi_x = \mathfrak{L}(x, R_x(\eta_x))(\xi_x).$$

In this paper, from the locking condition proposed by Huang [13], we require

$$\mathscr{T}_{\eta_x}\xi_x = \mathscr{T}_{S(\eta_x)}\xi_x.$$

In some manifolds, there exist retractions such that the above equality holds, e.g., the Stiefel manifold and the Grassman manifold [13]. Furthermore, from the above results, we have

$$\|\xi_x\| = \|\mathscr{T}_{S(\eta_x)}\xi_x\| = \|\mathfrak{L}(x, R_x(\eta_x))(\xi_x)\| = \|\mathscr{T}_{\eta_x}\xi_x\| = \|\mathrm{D}R_x(\eta_x)[\xi_x]\|.$$

3. The generalized conditional gradient method

Let $J = \{1,2,\cdots,m\}$, $\mathbb{R}^m_+ = \{x \in \mathbb{R}^m : x_j \geq 0, j \in J\}$ and $\mathbb{R}^m_{++} = \{x \in \mathbb{R}^m : x_j > 0, j \in J\}$. For $x, y \in \mathbb{R}^m_+$, $y \succeq x$ (or $x \preceq y$) means that $y - x \in \mathbb{R}^m_+$ and $y \succ x$ (or $x \prec y$) means that $y - x \in \mathbb{R}^m_{++}$. Let $\Psi : M \to (\mathbb{R}\bigcup\{+\infty\})^m$ be a vector-valued function with $\Psi := (\psi_1(x), \psi_2(x), \cdots, \psi_m(x))$ and consider the problem

$$\min_{x\in M} \Psi(x). \tag{3.1}$$

We study the problem of finding a Pareto optimal point of Ψ, i.e., a point $x^* \in M$ such that there exists no other $x \in M$ with $\Psi(x) \preceq \Psi(x^*)$ and $\Psi(x) \neq \Psi(x^*)$. In turn, $x^* \in M$ is called a weakly Pareto optimal point of (3.1), if there exists no other $x \in M$ such that $\Psi(x) \prec \Psi(x^*)$. A necessary optimality condition for problem (3.1) at a point $x^* \in M$ is given by

$$\max_{j\in J} \psi_j'(x^*; d) \geq 0, \quad \forall d \in T_xM. \tag{3.2}$$

A point $x^* \in \mathrm{dom}(\Psi)$ satisfying (3.2) is called a Pareto critical point of (3.1). From [13, 19], Ψ is said to be convex on $\mathrm{dom}(\Psi)$ if $\Psi(R_y t(R_y^{-1}x)) \preceq t\Psi(x) + (1-t)\Psi(y)$, for all $x, y \in \mathrm{dom}(\Psi)$ and $t \in [0,1]$. If Ψ is differentiable and convex on $\mathrm{dom}(\Psi)$, we have $\psi_j(y) - \psi_j(x) \geq \langle \mathrm{grad}\psi_j(x),\ R_x^{-1}y\rangle$, for all $x, y \in \mathrm{dom}(\Psi)$ and $j \in J$. Moreover, if ψ_j is differentiable, we have $\psi_j'(x; d) = \langle \mathrm{grad}\psi_j(x),\ d\rangle$, for all $d \in T_xM$ and $j \in J$.

By the similar proof of Section 4.1 in [25], we have the following result.

Lemma 3.1. *Let $\psi : M \to \mathbb{R}\bigcup\{+\infty\}$ be a convex function. Then, the function $\lambda \mapsto \frac{\psi(R_x\lambda d) - \psi(x)}{\lambda}$ is non-decreasing in $(0, +\infty)$. In particular, for any $\lambda \in (0,1]$, we have*

$$\frac{\psi(R_x\lambda d) - \psi(x)}{\lambda} \leq \psi(R_xd) - \psi(x), \quad \forall d \in T_xM. \tag{3.3}$$

Consequently, $\psi'(x; d) \leq \psi(R_xd) - \psi(x)$.

In this paper, the following assumptions will be considered only when explicitly stated.

(A1) For all $j \in J$, the function g_j is proper, convex and Lipschitz continuous on $\mathrm{dom}(G)$ with constant $L_{g_j} > 0$;

(A2) For all $j \in J$, the gradient function $\mathrm{grad}h_j$ is Lipschitz continuous with constant $L_{h_j} > 0$.

Let

$$L_G := \max\{L_{g_j} \mid j \in J\}, \quad L_H := \max\{L_{h_j} \mid j \in J\}. \tag{3.4}$$

The gap function $\theta : \mathrm{dom}(G) \to \mathbb{R}$ associated to the problem (1.1) is defined by

$$\theta(x) = \min_{u\in M}\max_{j\in J}\{g_j(u) - g_j(x) + \langle \mathrm{grad}h_j(x),\ R_x^{-1}u\rangle\}. \tag{3.5}$$

Under (A1) and (A2), it is easy to check that the problem (3.5) has a solution and it belongs to $\mathrm{dom}(G)$. Thus, we use $p(x) \in \mathrm{dom}(G)$ to denote a solution of (3.5), that is,

$$p(x) \in \arg\min_{u\in M}\max_{j\in J}\{g_j(u) - g_j(x) + \langle \mathrm{grad} h_j(x),\, R_x^{-1}u\rangle\}. \tag{3.6}$$

Therefore, combining (3.5) and (3.6), we conclude that

$$\theta(x) = \max_{j\in J}\{g_j(p(x)) - g_j(x) + \langle \mathrm{grad} h_j(x),\, R_x^{-1}p(x)\rangle\}.$$

Lemma 3.2. *Let $\theta : \mathrm{dom}(G) \to \mathbb{R}$ be defined as (3.5). Then,*
(i) $\theta(x) \leq 0$;
(ii) $\theta(x) = 0$ if and only if x is a Pareto critical point of (1.1);
(iii) $\theta(x)$ is upper semicontinuous.

Proof. (i) From the definition of $\theta(x)$, it follows that

$$\theta(x) \leq \max_{j\in J}\{g_j(u) - g_j(x) + \langle \mathrm{grad} h_j(x),\, R_x^{-1}u\rangle\}, \quad \forall u \in M. \tag{3.7}$$

Thus, letting $u = x$ in the previous inequality, we conclude that $\theta(x) = 0$, which proves (i).

(ii) We first assume that x is a Pareto critical point of (1.1). By (3.2), we obtain

$$\max_{j\in J} f_j'(x;d) \geq 0, \quad \forall d \in T_xM. \tag{3.8}$$

Note that, for all $d \in T_xM$,

$$f_j'(x;d) = g_j'(x;d) + \langle \mathrm{grad} h_j(x),\, d\rangle. \tag{3.9}$$

This, together with Lemma 3.1 and (3.8), gives that

$$\max_{j\in J}\{g_j(R_xd) - g_j(x) + \langle \mathrm{grad} h_j(x),\, d\rangle\} \geq 0. \tag{3.10}$$

Setting $d = R_x^{-1}p(x)$ in (3.10) yields that

$$\theta(x) = \max_{j\in J}\{g_j(p(x)) - g_j(x) + \langle \mathrm{grad} h_j(x),\, R_x^{-1}p(x)\rangle\} \geq 0, \tag{3.11}$$

which, together with item (i), implies that $\theta(x) = 0$.

Conversely, assuming $\theta(x) = 0$, we have

$$\max_{j\in J}\{g_j(u) - g_j(x) + \langle \mathrm{grad} h_j(x),\, R_x^{-1}u\rangle\} \geq 0, \quad \forall u \in M. \tag{3.12}$$

In particular, given any $\alpha > 0$, letting $u = R_x\alpha d$ in (3.12), we conclude that

$$\max_{j\in J}\{\frac{g_j(R_x\alpha d) - g_j(x)}{\alpha} + \langle \mathrm{grad} h_j(x),\, d\rangle\} \geq 0, \quad \forall d \in T_xM. \tag{3.13}$$

Taking $\alpha \to 0$ in (3.13), it holds

$$\max_{j\in J} f_j'(x;d) = \max_{j\in J}\{g_j'(x;d) + \langle \mathrm{grad} h_j(x),\, d\rangle\} \geq 0, \quad \forall d \in T_xM. \tag{3.14}$$

Therefore, x is a Pareto critical point of problem (1.1).

(iii) Let $\{x^k\}$ be the sequence such that $\lim_{k\to\infty} x^k = x$. Since $p(x) \in \mathrm{dom}(G)$, we obtain

$$\theta(x^k) \leq \max_{j\in J}\{g_j(p(x)) - g_j(x^k) + \langle \mathrm{grad} h_j(x^k),\, R_{x^k}^{-1}p(x)\rangle\}. \tag{3.15}$$

From the property of the retraction R (see Section 4 in [19]), we have that $R_{(\cdot)}^{-1}p(x)$ is continuous, and

$$\lim_{k\to\infty}\sup\theta(x^k) \leq \max_{j\in J}\{g_j(p(x)) + \langle \mathrm{grad}h_j(x),\, R_x^{-1}p(x)\rangle + \lim_{k\to\infty}\sup(-g_j(x^k))\}.$$

Since g_j is lower semicontinuous, we obtain

$$\lim_{k\to\infty}\sup[-g_j(x^k)] \leq -g_j(x).$$

Thus,

$$\lim_{k\to\infty}\sup\theta(x^k) \leq \max_{j\in J}\{g_j(p(x)) - g_j(x) + \langle \mathrm{grad}h_j,\, R_x^{-1}p(x)\rangle\} = \theta(x), \tag{3.16}$$

which concludes the proof. □

Next, we denote $e := (1,\cdots,1)^T \in \mathbb{R}^m$. Then we have the following result.

Lemma 3.3. *Assume that F satisfies (A2). Let $\lambda \in [0,1]$. Then,*

$$F(R_x\lambda R_x^{-1}p(x)) \preceq F(x) + (\lambda\theta(x) + \frac{L_H}{2}\|R_x^{-1}p(x)\|^2\lambda^2)e. \tag{3.17}$$

Proof. Since g_j is convex, and $\mathrm{grad}h_j$ is Lipschitz continuous with constant L_{h_j}, it follows that for any $x \in \mathrm{dom}(G)$ and $\lambda \in [0,1]$,

$$\begin{aligned} f_j(R_x\lambda(R_x^{-1}p(x))) \leq{}& (1-\lambda)g_j(x) + \lambda g_j(p(x)) + h_j(x) + \lambda\langle \mathrm{grad}h_j(x),\, R_x^{-1}p(x)\rangle \\ &+ \frac{L_{h_j}}{2}\|R_x^{-1}p(x)\|^2\lambda^2, \quad \forall j \in J, \end{aligned} \tag{3.18}$$

which yields

$$\begin{aligned} f_j(R_x\lambda(R_x^{-1}p(x))) \leq{}& f_j(x) + \lambda[\langle \mathrm{grad}h_j(x),\, R_x^{-1}p(x)\rangle - g_j(x) + g_j(p(x))] \\ &+ \frac{L_{h_j}}{2}\|R_x^{-1}p(x)\|^2\lambda^2, \quad \forall j \in J. \end{aligned} \tag{3.19}$$

Therefore, by the definition of L_H, we have

$$F(R_x\lambda(R_x^{-1}p(x))) \preceq F(x) + (\lambda\theta(x) + \frac{L_H}{2}\|R_x^{-1}p(x)\|^2\lambda^2)e, \tag{3.20}$$

which concludes the proof. □

Based on the above results, we introduce the generalized conditional gradient method for composite multiobjective optimization problems on Riemannian manifolds, which generates the conditional gradient method in [6, 22, 23, 24] from $\mathbb{R}^n$ to Riemannian manifolds.

Algorithm 1 The generalized conditional gradient method

Step 0. Choose $x^0 \in \mathrm{dom}(G)$ and set $k := 0$.
Step 1. Compute $p(x^k)$ and $\theta(x^k)$ as follows:

$$p(x^k) \in \arg\min_{u\in M}\max_{j\in J}\{g_j(u) - g_j(x) + \langle \mathrm{grad}h(x^k),\ R_{x^k}^{-1}u\rangle\}, \tag{3.21}$$

$$\theta(x^k) = \max_{j\in J}\{g_j(p(x^k)) - g_j(x^k) + \langle \mathrm{grad}h_j(x^k),\ R_{x^k}^{-1}p(x^k)\rangle\}. \tag{3.22}$$

Step 2. If $\theta(x^k) = 0$, then stop.
Step 3. Compute $\alpha_k \in (0,1]$ by **Algorithm 2** or **Algorithm 3** and set

$$x^{k+1} := R_{x^k}\alpha_k(R_{x^k}^{-1}p(x^k)). \tag{3.23}$$

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

Next, we compute the step size α_k of Algorithm 1 with two step sizes. We begin by presenting the Armijo step size.

Algorithm 2 Armijo step size

Step 0. Take $\zeta \in (0,1)$, $0 < \mu_1 < \mu_2 < 1$, $\alpha_{k_0} = 1$ and set $l := 0$.
Step 1. If $F(R_{x^k}\alpha_{k_l}(R_{x^k}^{-1}p(x^k))) \preceq F(x^k) - \zeta\alpha_{k_l}|\theta(x^k)|e$, then set $\alpha_k := \alpha_{k_l}$ and Stop.
Step 2. Find $\alpha_{k_{l+1}} \in [\mu_1\alpha_{k_l}, \mu_2\alpha_{k_l}]$.
Step 3. Set $l := l+1$ and go to Step 1.

The second step size is the nonmonotone line search, see for example [18, 20, 21].

Algorithm 3 Nonmonotone line search rule

Step 0. Take $\sigma \in (0,1), \beta \in (0,1)$, $0 \le \eta_{\min} \le \eta_{\max} \le 1$, and set $l := 0$.
Step 1. Set $\alpha_k = \beta^l$ and compute $x^{k+1} = R_{x^k}\alpha_k(R_{x^k}^{-1}p(x^k))$.
Step 2. If

$$F(R_{x^k}\alpha_k(R_{x^k}^{-1}p(x^k))) \preceq C^k + \sigma\alpha_k\theta(x^k)e, \tag{3.24}$$

then stop. Here $C^0 = F(x^0)$ and $C^k = (C_1^k, C_2^k, \cdots, C_m^k)$ for $k \ge 1$. Furthermore, C_j^k is computed by the following rule:

$$C_j^{k+1} := \frac{\eta_k Q_k C_j^k + f_j(x^{k+1})}{Q_{k+1}}, \quad \forall k \ge 0, j \in J,$$

with $Q_0 = 1$, $\eta_0 \in [\eta_{\min}, \eta_{\max}]$, $Q_{k+1} := \eta_k Q_k + 1$, and η_{k+1} being generated by an adaptive formular [18].
Step 3. Set $l := l+1$ and go to Step 1.

If $M = \mathbb{R}^n$ and $R_x\eta = x+\eta$, the components of function G are the indicator function of a set C, then Algorithm 1 can be regarded as a generalization of Algorithm 3 of [6] from $\mathbb{R}^n$ to Riemannian manifolds..

4. Convergence Analysis

In this section, under some assumptions, we prove that every accumulation point of the sequence produced by Algorithm 1 is a Pareto critical point of F, and the iteration-complexity bound is studied on Riemannian manifolds.

4.1. **Convergence analysis using the Armijo step size.**

Proposition 4.1. *Let $\zeta \in (0,1)$ and $x^k \in \mathrm{dom}(G)$. Then, there exists $0 < \bar{\eta} \leq 1$ such that*

$$F(R_{x^k}\eta(R_{x^k}^{-1}p(x^k))) \preceq F(x^k) - \zeta\eta|\theta(x^k)|e, \quad \forall \eta \in (0,\bar{\eta}]. \tag{4.1}$$

Proof. Since H is differentiable and G is convex, for all $\eta \in (0,1)$ and $j \in J$, we have

$$\begin{aligned}
& f_j(R_{x^k}\eta(R_{x^k}^{-1}p(x^k))) \\
&= g_j(R_{x^k}\eta(R_{x^k}^{-1}p(x^k))) + h_j(R_{x^k}\eta(R_{x^k}^{-1}p(x^k))) \\
&\leq (1-\eta)g_j(x^k) + \eta g_j(p(x^k)) + h_j(x^k) + \eta\langle \mathrm{grad} h_j(x^k),\, R_{x^k}^{-1}(p(x^k)\rangle + \frac{o(\eta)}{\eta},
\end{aligned} \tag{4.2}$$

which yields

$$\begin{aligned}
& f_j(R_{x^k}\eta(R_{x^k}^{-1}p(x^k))) \\
&\leq f_j(x^k) + \eta(\langle \mathrm{grad} h_j(x^k),\, R_{x^k}^{-1}p(x^k)\rangle + g_j(p(x^k)) - g_j(x^k)) + \frac{o(\eta)}{\eta}.
\end{aligned} \tag{4.3}$$

For all $j \in J$, since $\langle \mathrm{grad} h_j(x^k),\, R_{x^k}^{-1}p(x^k)\rangle \leq \max_{j\in J}\langle \mathrm{grad} h_j(x^k),\, R_{x^k}^{-1}(p(x^k)\rangle$, we obtain

$$F(R_{x^k}\eta(R_{x^k}^{-1}p(x^k))) \preceq F(x^k) + \eta\zeta\theta(x^k)e + \eta((1-\zeta)\theta(x^k) + \frac{o(\eta)}{\eta})e.$$

Since $\lim_{\eta\to 0}\frac{o(\eta)}{\eta} = 0, \theta(x^k) < 0$ and $\zeta \in (0,1)$, there exists $\bar{\eta} \in (0,1]$ such that (4.1) holds for all $\eta \in (0,\bar{\eta}]$. ☐

Theorem 4.1. *Assume that F satisfies (A1). Let $\{x^k\}$ be the sequence generated by Algorithm 1 with Armijo step size. Then, every limit point $\bar{x}$ of $\{x^k\}$ is a Pareto critical point of problem (1.1).*

Proof. Let $\bar{x} \in \mathrm{dom}(G)$ be a limit point of $\{x^k\}$ generated by Algorithm 1 with Armijo step size, and $\lim_{k\in\mathbb{K}} x^k = \bar{x}$. It follows from Step 1 of Algorithm 2 that

$$0 \prec -\zeta\alpha_k\theta(x^k)e \preceq F(x^k) - F(x^{k+1}), \quad \forall k \in \mathbb{K}. \tag{4.4}$$

Consequently, the sequence $\{F(x^k)\}_{k\in\mathbb{K}}$ is monotone decreasing.

On the other hand, since $\{x^k\} \subseteq \mathrm{dom}(G)$ and $\mathrm{dom}(G)$ is compact, there exists $x^* \in \mathrm{dom}(G)$ such that $\lim_{k\in\mathbb{K}} x^k = x^*$. Then, we have

$$\begin{aligned}
\|F(x^k) - F(x^*)\| &= \|G(x^k) - G(x^*) + H(x^k) - H(x^*)\| \\
&\leq \|G(x^k) - G(x^*)\| + \|H(x^k) - H(x^*)\| \\
&\leq L_G d(x^k, x^*) + \|H(x^k) - H(x^*)\|, \quad \forall k \in \mathbb{K}.
\end{aligned}$$

Since H is continuous and $\lim_{k\in\mathbb{K}} x^k = x^*$, it holds

$$\lim_{k\in\mathbb{K}} F(x^k) = F(x^*).$$

Thus, due to the monotonicity of $\{F(x^k)\}_{k\in\mathbb{K}}$, by (4.4), we have

$$\lim_{k\in\mathbb{K}} \alpha_k \theta(x^k) = 0. \tag{4.5}$$

Therefore, there exists $\mathbb{K}_1 \subseteq \mathbb{K}$ such that at least one of the following conclusions holds:

$$\lim_{k\in\mathbb{K}_1} \theta(x^k) = 0 \text{ or } \lim_{k\in\mathbb{K}_1} \alpha_k = 0.$$

Consider the case that $\lim_{k\in\mathbb{K}_1} \theta(x^k) = 0$. Using Lemma 3.2, we obtain $\theta(\bar{x}) = 0$, which implies that $\bar{x}$ is a Pareto critical point.

Now, consider the case that $\lim_{k\in\mathbb{K}_1} \alpha_k = 0$. Suppose by contradiction that $\theta(\bar{x}) < 0$. Since $\theta(\cdot)$ is upper semi-continuous, $\theta(\bar{x}) < 0$ and $\lim_{k\in\mathbb{K}_1} \alpha_k = 0$, there exist $\delta > 0$ and $\mathbb{K}_2 \subseteq \mathbb{K}_1$ such that

$$\theta(x^k) < -\delta, \quad \forall k \in \mathbb{K}_2. \tag{4.6}$$

Recall that $\{p(x^k)\} \subseteq \mathrm{dom}(G)$ and $\mathrm{dom}(G)$ is compact. Without loss of generality, we assume that there exists $\bar{p} \in \mathrm{dom}(G)$ such that

$$\lim_{k\in\mathbb{K}_2} p(x^k) = \bar{p}. \tag{4.7}$$

Since $\alpha_k < 1$ for all $k \in \mathbb{K}_2$, by Armijo step size, there exists $\bar{\alpha}_k \in (0, \alpha_k/\mu_1)$ such that

$$F(R_{x^k}\bar{\alpha}_k(R_{x^k}^{-1}p(x^k))) \npreceq F(x^k) + \zeta\bar{\alpha}_k\theta(x^k)e, \quad \forall k \in \mathbb{K}_2, \tag{4.8}$$

which implies that

$$f_{j_k}(R_{x^k}\bar{\alpha}_k(R_{x^k}^{-1}p(x^k))) > f_{j_k}(x^k) + \zeta\bar{\alpha}_k\theta(x^k), \quad \forall k \in \mathbb{K}_2, \tag{4.9}$$

for at least one $j_k \in J$. Then, we have there exist $j^* \in J$ and $\mathbb{K}_3 \subseteq \mathbb{K}_2$ such that

$$\frac{f_{j^*}(R_{x^k}\bar{\alpha}_k(R_{x^k}^{-1}p(x^k))) - f_{j^*}(x^k)}{\bar{\alpha}_k} > \zeta\theta(x^k), \quad \forall k \in \mathbb{K}_3. \tag{4.10}$$

On the other hand, since $0 < \bar{\alpha}_k \leq 1$, we can apply Lemma 3.1 to obtain

$$\frac{g_{j^*}(R_{x^k}\bar{\alpha}_k(R_{x^k}^{-1}p(x^k))) - g_{j^*}(x^k)}{\bar{\alpha}_k} \leq g_{j^*}(p(x^k)) - g_{j^*}(x^k), \quad \forall k \in \mathbb{K}_3. \tag{4.11}$$

Since h_{j^*} is differentiable and $\lim_{k\in\mathbb{K}_3} \bar{\alpha}_k = 0$, for all $k \in \mathbb{K}_3$, we have

$$\bar{\alpha}_k\langle \mathrm{grad} h_{j^*}(x^k),\, R_{x^k}^{-1}p(x^k)\rangle = h_{j^*}(R_{x^k}\bar{\alpha}_k(R_{x^k}^{-1}p(x^k))) - h_{j^*}(x^k) - o(\bar{\alpha}_k\|R_{x^k}^{-1}p(x^k)\|), \tag{4.12}$$

which, together with (4.11), implies that

$$\begin{aligned} \theta(x^k) &\geq g_{j^*}(p(x^k)) - g_{j^*}(x^k) + \langle \mathrm{grad} h_{j^*}(x^k),\, R_{x^k}^{-1}p(x^k)\rangle \\ &\geq \frac{f_{j^*}(R_{x^k}\bar{\alpha}_k(R_{x^k}^{-1}p(x^k))) - f_{j^*}(x^k)}{\bar{\alpha}_k} - \frac{o(\bar{\alpha}_k\|R_{x^k}^{-1}p(x^k)\|)}{\bar{\alpha}_k}, \quad \forall k \in \mathbb{K}_3. \end{aligned} \tag{4.13}$$

From (4.13) and (4.10), we know

$$\frac{f_{j^*}(R_{x^k}\bar{\alpha}_k(R_{x^k}^{-1}p(x^k))) - f_{j^*}(x^k)}{\bar{\alpha}_k} > (\frac{-\zeta}{1-\zeta})\frac{o(\bar{\alpha}_k\|R_{x^k}^{-1}p(x^k)\|)}{\bar{\alpha}_k}, \quad \forall k \in \mathbb{K}_3. \tag{4.14}$$

On the other hand, it follows from (4.6) and (4.13) that

$$-\delta+\frac{o(\bar{\alpha}_k\|R_{x^k}^{-1}p(x^k)\|)}{\bar{\alpha}_k}>\frac{f_{j^*}(R_{x^k}\bar{\alpha}_kR_{x^k}^{-1}p(x^k))-f_{j^*}(x^k)}{\bar{\alpha}_k},\quad \forall k\in\mathbb{K}_3. \tag{4.15}$$

Together with (4.14), it holds

$$-\delta+\frac{o(\bar{\alpha}_k\|R_{x^k}^{-1}p(x^k)\|)}{\bar{\alpha}_k}>(\frac{-\zeta}{1-\zeta})\frac{o(\bar{\alpha}_k\|R_{x^k}^{-1}p(x^k)\|)}{\bar{\alpha}_k},\quad \forall k\in\mathbb{K}_3. \tag{4.16}$$

Considering (4.7) and the smoothness of the retarction mapping R, we obtain $\delta<0$, which contradicts to $\delta>0$. Thus, $\theta(\bar{x})=0$. From Lemma 3.2, we have $\bar{x}$ is a Pareto critical point. □

Let $x^*\in\operatorname{dom}(G)$ and $\hat{\mathscr{U}}$ be a neighborhood of 0_{x^*} in $T_{x^*}M$ such that $\mathscr{U}:=\exp_{x^*}\hat{\mathscr{U}}$. Then, we have

$$d(x,x^*)=\|\exp_{x^*}^{-1}x-\exp_{x^*}^{-1}x^*\|=\|\exp_{x^*}^{-1}x\|,\quad \forall x\in\mathscr{U}.$$

Since exp is a retraction, we obtain $D(R_{x^*}^{-1}\circ\exp_{x^*})(0_{x^*})=\mathrm{id}$. Hence,

$$\begin{aligned}\|R_{x^*}^{-1}x\|&=\|R_{x^*}^{-1}x-R_{x^*}^{-1}x^*\|=\|\exp_{x^*}^{-1}x-\exp_{x^*}^{-1}x^*\|+o(\|\exp_{x^*}^{-1}x-\exp_{x^*}^{-1}x^*\|)\\&=d(x,x^*)+o(d(x,x^*)).\end{aligned}$$

Since $\operatorname{dom}(G)$ is a compact set, from the above equality, there exists $\Omega>0$ such that $\Omega\geq\max\{\|R_x^{-1}y\|:x,y\in\operatorname{dom}(G)\}>\operatorname{diam}(\operatorname{dom}(G))$. Moreover, we define

$$\rho:=\sup\{\|\mathrm{grad}h_j(x)\|\mid x\in\operatorname{dom}(G),j\in J\}, \tag{4.17}$$

$$\gamma:=\min\{\frac{1}{(L_G+\rho)\Omega},\frac{2\mu_1(1-\zeta)}{L_H\Omega^2}\}. \tag{4.18}$$

Lemma 4.1. *Assume that F satisfies (A1)-(A2). Let $\{x^k\}$ be the sequence generated by Algorithm 1 with Armijo step size. Then, $\alpha_k\geq\gamma|\theta(x^k)|>0$.*

Proof. Since $\alpha_k\in(0,1]$, let us consider two cases: $\alpha_k=1$ and $0<\alpha_k<1$.
First, we assume that $\alpha_k=1$. From (3.22), we know

$$\theta(x^k)=\max_{j\in J}\{g_j(p(x^k))-g_j(x^k)+\langle\mathrm{grad}h_j(x^k),\,R_{x^k}^{-1}p(x^k)\rangle\}<0, \tag{4.19}$$

which implies that

$$0<-\theta(x^k)\leq g_j(x^k)-g_j(p(x^k))-\langle\mathrm{grad}h_j(x^k),\,R_{x^k}^{-1}p(x^k)\rangle, \tag{4.20}$$

for all $j\in J$. Thus, it follows from Cauchy-Schwarz inequality that

$$0<-\theta(x^k)\leq L_Gd(x^k,p(x^k))+\|\mathrm{grad}h_j(x^k)\|\|R_{x^k}^{-1}p(x^k)\|. \tag{4.21}$$

Using (4.17), we have $0<-\theta(x^k)\leq(L_G+\rho)\Omega$. Hence, from (4.18), we obtain

$$0<-\gamma\theta(x^k)\leq\frac{-\theta(x^k)}{(L_G+\rho)\Omega}\leq1, \tag{4.22}$$

which shows that the desired equality holds.

Now, we assume that $0 < \alpha_k < 1$. From Armijo step size, we conclude that there exist $0 < \bar{\alpha}_k \leq \min\{1, \frac{\alpha_k}{\mu_1}\}$ and $j_k \in J$ such that

$$f_{j_k}(R_{x^k}\bar{\alpha}_k(R_{x^k}^{-1}p(x^k))) > f_{j_k}(x^k) + \zeta\bar{\alpha}_k\theta(x^k). \tag{4.23}$$

By using Lemma 3.3, we have

$$f_j(R_{x^k}\bar{\alpha}_k(R_{x^k}^{-1}p(x^k))) \leq f_j(x^k) + \bar{\alpha}_k\theta(x^k) + \frac{L_H}{2}\|R_{x^k}^{-1}p(x^k)\|^2\bar{\alpha}_k^2, \quad \forall j \in J. \tag{4.24}$$

From the above inequality, we conclude that

$$-\theta(x^k)(1-\zeta) < \frac{L_H}{2}\|R_{x^k}^{-1}p(x^k)\|^2\bar{\alpha}_k \leq \frac{L_H}{2}\|R_{x^k}^{-1}p(x^k)\|^2\frac{\alpha_k}{\mu_1}. \tag{4.25}$$

Therefore,

$$0 < -\gamma\theta(x^k) \leq -\frac{2\mu_1(1-\zeta)}{L_H\Omega^2}\theta(x^k) < \alpha_k, \tag{4.26}$$

which concludes the proof.. □

In the following results, we obtain the iteration-complexity bound of the generalized conditional gradient method for composite multiobjective optimization problems on Riemannian manifolds. We define

$$f_{j^*}(x^0) := \max\{f_j(x^0) \mid j \in J\}, \tag{4.27}$$

$$f_{j^*}^{\inf} := \min\{f_j^* \mid j \in J\}, \tag{4.28}$$

where $f_j^* := \inf\{f_j(x) | x \in \mathrm{dom}(G)\}$ for all $j \in J$.

Theorem 4.2. *Assume that F satisfies (A1)-(A2). Then,* $\lim_{k\to\infty} F(x^k) = F(x^*)$ *for some* $x^* \in \mathrm{dom}(G)$. *Moreover,*
(i) $\lim_{k\to\infty}\theta(x^k) = 0$*;*
(ii) $\min\{|\theta(x^k)| : k = 0, \cdots, N-1\} \leq \sqrt{\frac{f_{j^*}(x^0) - f_{j^*}^{\inf}}{\zeta\gamma N}}$.

Proof. (i) By Step 1 of Algorithm 2, we obtain

$$F(R_{x^k}\alpha_k(R_{x^k}^{-1}p(x^k))) \preceq F(x^k) + \zeta\alpha_k\theta(x^k)e, \tag{4.29}$$

or equivalently,

$$\zeta\alpha_k\theta(x^k)e \preceq F(x^k) - F(x^{k+1}). \tag{4.30}$$

From Lemma 4.1, we have

$$0 \prec \zeta\gamma|\theta(x^k)|^2 e \preceq F(x^k) - F(x^{k+1}), \tag{4.31}$$

which implies that the sequence $\{F(x^k)\}$ is monotone decreasing.

By the similar proof of Theorem 4.1, we conclude that

$$\lim_{k\to\infty} F(x^k) = F(x^*). \tag{4.32}$$

This, together with (4.31), gives that

$$\lim_{k\to\infty}|\theta(x^k)|^2 = 0. \tag{4.33}$$

Therefore, $\lim_{k\to\infty}\theta(x^k)=0$.

(ii) From (4.31), for all $j\in J$, we obtain

$$\sum_{k=0}^{N-1}|\theta(x^k)|^2\le\sum_{k=0}^{N-1}\frac{1}{\zeta\gamma}[f_j(x^k)-f_j(x^{k+1})]\le\frac{1}{\zeta\gamma}[f_{j^*}(x^0)-f_{j^*}^{\inf}]. \tag{4.34}$$

Thus,

$$\min\{|\theta(x^k)|^2:k=0,\cdots,N-1\}\le\frac{[f_{j^*}(x^0)-f_{j^*}^{\inf}]}{\zeta\gamma N}, \tag{4.35}$$

which concludes the proof. □

Corollary 4.1. *Assume that F satisfies (A1)-(A2). Let $\varepsilon>0$ and define $K(\varepsilon):=\{k\in N:|\theta(x^k)|>\varepsilon\}$. Then,*

$$|K(\varepsilon)|\le\frac{[f_{j^*}(x^0)-f_{j^*}^{\inf}]}{\zeta\gamma}\frac{1}{\varepsilon^2}, \tag{4.36}$$

where $|K(\varepsilon)|$ denotes the number of elements of $K(\varepsilon)$.

Proof. The proof follows from (4.35) straightforwardly. □

Corollary 4.2. *Assume that F satisfies (A1)-(A2) and $\varepsilon>0$. If $|\theta(x^k)|>\varepsilon$, then Algorithm 1 with Armijo line search performs, at most, $1+\frac{\ln(\gamma\varepsilon)}{\ln(\mu_2)}$ evaluations of F to compute the step size α_k.*

Proof. Let t_k and ω_k be the numbers of inner iterations and evaluations of F with Armijo line search to compute α_k, respectively. By Algorithm 2, we get $\omega_k=t_k+1$ and $\mu_2^{t_k}\ge\alpha_k$. Hence, from Lemma 4.1, it follows that $\mu_2^{t_k}\ge\gamma|\theta(x^k)|$. Since $|\theta(x^k)|>\varepsilon$, we have $\mu_2^{t_k}\ge\gamma\varepsilon$. Therefore, $t_k\le\frac{\ln(\gamma\varepsilon)}{\ln(\mu_2)}$, which concludes the proof. □

The following result follows from Corollaries 4.1 and 4.2 straightforwardly.

Theorem 4.3. *Assume that F satisfies (A1)-(A2) and $\varepsilon>0$. Then, Algorithm 1 with Armijo step size generates a point x^k such that $|\theta(x^k)|\le\varepsilon$, performing, at most,*

$$m[(1+\frac{\ln(\gamma\varepsilon)}{\ln(\mu_2)})\frac{[f_{j^*}(x^0)-f_{j^*}^{\inf}]}{\zeta\gamma}\frac{1}{\varepsilon^2}+1]=\mathscr{O}(|\ln(\varepsilon)|\varepsilon^{-2}) \tag{4.37}$$

evaluations of functions $f_1,\cdots,f_m$.

Remark 4.1. If $M=\mathbb{R}^n$ and $R_x\eta=x+\eta$, then Theorems 4.1, 4.2 and 4.3 can reduce to Theorems 8, 10 and 13 of [24].

4.2. **Convergence analysis using the nonmonotone line search step size.**

Proposition 4.2. *Assume that F satisfies (A1)-(A2). Let $\{x^k\}$ be the sequence generated by Algorithm 1 with the nonmonotone step size. Then,*

$$F(x^k)\preceq C^k. \tag{4.38}$$

Moreover, if x^k is not a Pareto critical point of F, then there exists α_k satisfying the nonmonotone line search condition.

Proof. Since $\theta(x^k) < 0$, by the nonmonotone line search (3.24), it follows that

$$F(x^k) \preceq C^k.$$

Now, we prove that there exists α_k satisfying the nonmonotone line search condition. Since H is differentiable and G is convex, for all $j \in J$, we conclude that

$$\begin{aligned} & f_j(R_{x^k}\alpha_k(R_{x^k}^{-1}p(x^k))) = g_j(R_{x^k}\alpha_k(R_{x^k}^{-1}p(x^k))) + h_j(R_{x^k}\alpha_k(R_{x^k}^{-1}p(x^k))) \\ & \leq (1-\alpha_k)g_j(x^k) + \alpha_k g_j(p(x^k)) + h_j(x^k) + \alpha_k\langle \mathrm{grad} h_j(x^k),\, R_{x^k}^{-1}p(x^k)\rangle + \frac{o(\alpha_k)}{\alpha_k} \\ & = f_j(x^k) + \alpha_k[\langle \mathrm{grad} h_j(x^k),\, R_{x^k}^{-1}p(x^k)\rangle + g_j(p(x^k)) - g_j(x^k)] + \frac{o(\alpha_k)}{\alpha_k}. \end{aligned} \tag{4.39}$$

For all $j \in J$, since $\langle \mathrm{grad} h_j(x^k),\, R_{x^k}^{-1}p(x^k)\rangle \leq \max_{j\in J}\langle \mathrm{grad} h_j(x^k),\, R_{x^k}^{-1}p(x^k)\rangle$ and $f_j(x^k) \leq C_j^k$, we have

$$F(R_{x^k}\alpha_k(R_{x^k}^{-1}p(x^k))) \preceq C^k + \sigma\alpha_k\theta(x^k)e + \alpha_k((1-\sigma)\theta(x^k) + \frac{o(\alpha_k)}{\alpha_k})e. \tag{4.40}$$

Observing that x^k is not a Pareto critical point, $\theta(x^k) < 0, \sigma \in (0,1)$, and that for $\alpha_k > 0$, small enough, the condition (3.24) holds. ☐

Theorem 4.4. *Assume that F satisfies (A1)-(A2). Let $\{x^k\}$ be the sequence generated by Algorithm 1 with the nonmonotone line search step size. Then,*

$$\alpha_k \geq \min\{\beta, \frac{2\beta(1-\sigma)}{L_H}\frac{|\theta(x^k)|}{\|R_{x^k}^{-1}p(x^k)\|^2}\}. \tag{4.41}$$

Proof. If $\alpha_k \geq \beta$, then (4.41) holds.
If $\alpha_k < \beta$, then by (3.24) and Proposition 4.2, we have

$$F(R_{x^k}\alpha_k\beta^{-1}(R_{x^k}^{-1}p(x^k))) \succ C^k + \sigma\alpha_k\beta^{-1}\theta(x^k)e \succeq F(x^k) + \sigma\alpha_k\beta^{-1}\theta(x^k)e. \tag{4.42}$$

For all $j \in J$, define $\phi_j^k : \mathbb{R} \to \mathbb{R}\bigcup\{+\infty\}$ by

$$\phi_j^k(t) := h_j(R_{x^k}t(R_{x^k}^{-1}p(x^k))).$$

Note that

$$\phi_j^k(\alpha_k\beta^{-1}) - \phi_j^k(0) = \alpha_k\beta^{-1}\frac{d\phi_j^k(0)}{dt} + \int_0^{\alpha_k\beta^{-1}}[\frac{d\phi_j^k(t)}{dt} - \frac{d\phi_j^k(0)}{dt}]dt.$$

Define $v(x^k) = R_{x^k}^{-1}p(x^k)$. From [13, 19], we have

$$\frac{d\phi_j^k(t)}{dt} = \langle \mathrm{grad} h_j(R_{x^k}tv(x^k)),\, DR_{x^k}tv(x^k)[v(x^k)]\rangle.$$

Thus, for all $j \in J$, it holds

$$
\begin{aligned}
& h_j(R_{x^k}\alpha_k\beta^{-1}v(x^k)) - h_j(x^k) \\
=& \alpha_k\beta^{-1}\langle \mathrm{grad} h_j(x^k),\ DR_{x^k}0v(x^k)[v(x^k)]\rangle \\
& + \int_0^{\alpha_k\beta^{-1}} [\langle \mathrm{grad} h_j(R_{x^k}tv(x^k)),\ DR_{x^k}tv(x^k)[v(x^k)]\rangle - \langle \mathrm{grad} h_j(x^k),\ DR_{x^k}0v(x^k)[v(x^k)]\rangle]dt \\
=& \alpha_k\beta^{-1}\langle \mathrm{grad} h_j(x^k),\ v(x^k)\rangle \\
& + \int_0^{\alpha_k\beta^{-1}} \langle \mathrm{grad} h_j(R_{x^k}tv(x^k)) - \mathfrak{L}(x^k,\ R_{x^k}tv(x^k))\mathrm{grad} h_j(x^k), \mathfrak{L}(x^k, R_{x^k}tv(x^k))v(x^k)\rangle dt \\
\le& \alpha_k\beta^{-1}\langle \mathrm{grad} h_j(x^k),\ v(x^k)\rangle \\
& + \int_0^{\alpha_k\beta^{-1}} \|\mathrm{grad} h_j(R_{x^k}tv(x^k)) - \mathfrak{L}(x^k, R_{x^k}tv(x^k))\mathrm{grad} h_j(x^k)\|\|\mathfrak{L}(x^k, R_{x^k}tv(x^k))v(x^k)\| dt \\
\le& \alpha_k\beta^{-1}\langle \mathrm{grad} h_j(x^k),\ v(x^k)\rangle + \int_0^{\alpha_k\beta^{-1}} tL_H\|v(x^k)\|^2 dt \\
\le& \alpha_k\beta^{-1}\max_{j\in J}\langle \mathrm{grad} h_j(x^k),\ v(x^k)\rangle + \frac{L_H}{2}[\alpha_k\beta^{-1}]^2\|v(x^k)\|^2.
\end{aligned}
$$

By (4.42), for all $j \in J$, we have

$$
g_j(R_{x^k}\alpha_k\beta^{-1}v(x^k)) - g_j(x^k) + h_j(R_{x^k}\alpha_k\beta^{-1}v(x^k)) - h_j(x^k) \ge \sigma\alpha_k\beta^{-1}\theta(x^k), \tag{4.43}
$$

which, together with the above inequality, implies that

$$
\begin{aligned}
& g_j(R_{x^k}\alpha_k\beta^{-1}v(x^k)) - g_j(x^k) + \alpha_k\beta^{-1}\max_{j\in J}\langle \mathrm{grad} h_j(x^k),\ v(x^k)\rangle + \frac{L_H}{2}[\alpha_k\beta^{-1}]^2\|v(x^k)\|^2 \\
\ge& \sigma\alpha_k\beta^{-1}\theta(x^k), \quad \forall j \in J.
\end{aligned} \tag{4.44}
$$

Since G is convex, from Lemma 3.1, we have

$$
g_j(R_{x^k}\alpha_k\beta^{-1}v(x^k)) - g_j(x^k) \le \alpha_k\beta^{-1}[g_j(p(x^k)) - g_j(x^k)], \quad \forall j \in J.
$$

This, together with (4.44), gives that

$$
\alpha_k\beta^{-1}\theta(x^k) + \frac{L_H}{2}[\alpha_k\beta^{-1}]^2\|v(x^k)\|^2 \ge \sigma\alpha_k\beta^{-1}\theta(x^k).
$$

Thus, $\alpha_k \ge \frac{2\beta(1-\sigma)}{L_H}\frac{|\theta(x^k)|}{\|R_{x^k}^{-1}p(x^k)\|^2}$. □

Theorem 4.5. *Assume that F satisfies (A1)-(A2). F is bounded from below, $\eta_{\max} < 1$, and there exists $c > 0$ such that*

$$
|\theta(x^k)| \ge c\|R_{x^k}^{-1}p(x^k)\|^2, \quad \forall k \in \mathbb{N}.
$$

Then, every limit point of the sequence $\{x^k\}$ generated by Algorithm 1 with the nonmonotone line search is a Pareto critical point of F.

Proof. We show that

$$
F(x^{k+1}) \preceq C^k - \bar{\rho}|\theta(x^k)|e, \tag{4.45}
$$

where $\bar{\rho} = \min\{\sigma\beta, \frac{2\sigma\beta(1-\sigma)c}{L_H}\}$. The following two cases are possible.
Case 1. If $\alpha_k \geq \beta$, then by (3.24),

$$F(x^{k+1}) \preceq C^k + \sigma\alpha_k\theta(x^k)e = C^k - \sigma\alpha_k|\theta(x^k)|e \preceq C^k - \sigma\beta|\theta(x^k)|e,$$

which implies that (4.45).
Case 2. If $\alpha_k < \beta$, then by (4.41),

$$\alpha_k \geq \frac{2\beta(1-\sigma)}{L_H}\frac{|\theta(x^k)|}{\|R_{x^k}^{-1}p(x^k)\|^2}.$$

From (3.24), we have

$$F(x^{k+1}) \preceq C^k - \frac{2\sigma\beta(1-\sigma)}{L_H}\frac{|\theta(x_k)|^2}{\|R_{x^k}^{-1}p(x^k)\|^2}e.$$

Since $|\theta(x^k)| \geq c\|R_{x^k}^{-1}p(x^k)\|^2$, we know

$$F(x^{k+1}) \preceq C^k - \frac{2\sigma\beta(1-\sigma)c}{L_H}|\theta(x^k)|e,$$

which implies (4.45). Consequently,

$$C^{k+1} = \frac{\eta_k Q_k C^k + F(x^{k+1})}{Q_{k+1}} \preceq \frac{\eta_k Q_k C^k + C^k - \bar{\rho}|\theta(x^k)|e}{Q_{k+1}} = C^k - \frac{\bar{\rho}|\theta(x^k)|}{Q_{k+1}}e. \tag{4.46}$$

Since $F(x)$ is bounded from below and $F(x^k) \preceq C^k$, we conclude that C^k is bounded from below. Together with (4.46), it holds

$$\sum_{k=0}^{\infty}\frac{|\theta(x^k)|}{Q_{k+1}} \leq \sum_{k=0}^{\infty}\frac{1}{\bar{\rho}}(C_j^k - C_j^{k+1}) < +\infty, \quad \forall j \in J. \tag{4.47}$$

Suppose that x^* is a limit point of $\{x^k\}$. Thus, we have $\theta(x^*) = 0$. By contradiction, assume that $\theta(x^*) < 0$, which implies that there exists $\varepsilon > 0, \delta > 0$ such that for all $k \in \mathbb{K}$, $d(x^k, x^*) \leq \delta$, we have $|\theta(x^k)| \geq \varepsilon > 0$. Therefore, we can get

$$\sum_{k=0}^{\infty}\frac{|\theta(x^k)|}{Q_{k+1}} \geq \sum_{k\in\{k\in\mathbb{K}|d(x^k,x^*)\leq\delta\}}\frac{\varepsilon}{Q_{k+1}}. \tag{4.48}$$

Since $\eta_{\max} < 1$, by the similar proof of Theorem 2.2 in [18] , we have

$$Q_{k+1} \leq \frac{1}{1-\eta_{\max}}. \tag{4.49}$$

Together with (4.48), it holds

$$\sum_{k=0}^{\infty}\frac{|\theta(x^k)|}{Q_{k+1}} \geq \sum_{k\in\{k\in\mathbb{K}|d(x^k,x^*)\leq\delta\}}(1-\eta_{\max})\varepsilon = +\infty,$$

which contradicts to (4.47). Thus, we have that $\theta(x^*) = 0$. It follows from Lemma 3.2 that x^* is a Pareto critical point of F, which concludes the proof. □

Theorem 4.6. *Suppose that all assumptions of Theorem 4.5 are satisfied. The sequence $\{x^k\}$ is generated by Algorithm 1 with the nonmonotone line search. Then,* $\lim_{k\to\infty} F(x^k) = F(x^*)$ *for some* $x^* \in \mathrm{dom}(G)$*. Moreover,*

(i) $\lim_{k\to\infty}\theta(x^k)=0$*;*
(ii) $\min\{|\theta(x^k)|:k=0,\cdots,N-1\}\leq\frac{1}{(1-\eta_{\max})\bar{\rho}N}(f_{j^*}(x^0)-f_{j^*}^{\inf})$.

Proof. (i) By the similar proof of Theorems 4.1 and 4.5, we obtain $\lim_{k\to\infty}F(x^k)=F(x^*)$ and $\lim_{k\to\infty}\theta(x^k)=0$.

(ii) From the proof of Theorem 4.5, we obtain the sequence $\{C_j^k\}$ is bounded from below, and

$$\sum_{k=0}^{N-1}\frac{|\theta(x_k)|}{Q_{k+1}}\leq\sum_{k=0}^{N-1}\frac{1}{\bar{\rho}}(C_j^k-C_j^{k+1}),\quad\forall j\in J.$$

Since $C^0=F(x^0)$, for all $j\in J$, we have $C_j^0=f_j(x^0)$. Moreover, from $F(x^k)\preceq C^k$, we know

$$\sum_{k=0}^{N-1}\frac{|\theta(x_k)|}{Q_{k+1}}\leq\sum_{k=0}^{N-1}\frac{1}{\bar{\rho}}(C_j^k-C_j^{k+1})\leq\frac{1}{\bar{\rho}}(f_{j^*}(x^0)-f_{j^*}^{\inf}).$$

Therefore,

$$\min\{|\theta(x^k)|:k=0,1,\cdots,N-1\}\leq\frac{Q_{k+1}}{\bar{\rho}N}(f_{j^*}(x^0)-f_{j^*}^{\inf})\leq\frac{1}{(1-\eta_{\max})\bar{\rho}N}(f_{j^*}(x^0)-f_{j^*}^{\inf}),$$

which concludes the proof. □

Corollary 4.3. *Suppose that all assumptions of Theorem 4.5 are satisfied and* $\varepsilon>0$*. Define the set* $K(\varepsilon):=\{k\in\mathbb{N}:|\theta(x^k)|>\varepsilon\}$*. Then,*

$$|K(\varepsilon)|\leq\frac{f_{j^*}(x^0)-f_{j^*}^{\inf}}{(1-\eta_{\max})\bar{\rho}\varepsilon},\tag{4.50}$$

where $|K(\varepsilon)|$ *denotes the number of elements of* $K(\varepsilon)$*.*

Proof. The proof follows from Theorem 4.6 straightforwardly. □

Theorem 4.7. *Suppose that all assumptions of Theorem 4.5 are satisfied and* $\varepsilon>0$*. Then, Algorithm 1 with the nonmonotone line search generates a point* x^k *such that* $|\theta(x^k)|\leq\varepsilon$*, performing, at most,*

$$m[(\max(1,\ln(\frac{2\beta(1-\sigma)\varepsilon}{L_H\Omega^2})/\ln\beta)+1)\frac{[f_{j^*}(x^0)-f_{j^*}^{\inf}]}{(1-\eta_{\max})\bar{\rho}\varepsilon}+1]=\mathscr{O}(|\ln(\varepsilon)|\varepsilon^{-1})$$

evaluations of functions $f_1,\cdots,f_m$*.*

Proof. Let t_k be the number of inner iterations of F with the nonmonotone line search to compute α_k. By Algorithm 3, we have $\alpha_k=\beta^{t_k}$. Together with Theorem 4.4, it holds $\beta^{t_k}=\alpha_k\geq\min\{\beta,\frac{2\beta(1-\sigma)}{L_H}\frac{|\theta(x^k)|}{\|R_{x^k}^{-1}p(x^k)\|^2}\}$. Since $|\theta(x^k)|>\varepsilon$ and $\Omega\geq\|R_{x^k}^{-1}p(x^k)\|$, it follows that $t_k\leq\max\{1,\ln(\frac{2\beta(1-\sigma)\varepsilon}{L_H\Omega^2})/\ln\beta\}$. This, together with Corollary 4.3, gives the result. □

Remark 4.2. Proposition 4.2 and Theorem 4.5 can be regarded as a generalization of Proposition 4.5 and Theorem 4.6 of [20] from smooth objective functions to non-smooth functions, respectively. Moreover, exponential mappings and parallel transports in [20] are extended to retractions and isometric vector transports in this paper, respectively.

Next, we apply the generalized conditional gradient method for composite multiobjective problems to sphere S^{n-1}.

Example 4.1. On the unit sphere S^{n-1} considered as a Riemannian manifold of $\mathbb{R}^n$, the inner product inherited from the standard inner product on $\mathbb{R}^n$ is given by

$$g(\xi_x, \eta_x) := \xi_x^T \eta_x, \quad \forall \xi_x, \eta_x \in T_x S^{n-1},$$

and the projection is given by $P_x \xi_x = (I - xx^T)\xi_x$. From Section 4 in [19], we have $R_x(\eta_x) = \frac{x+\eta_x}{\|x+\eta_x\|}$. The tangent space to S^{n-1} is

$$T_x S^{n-1} = \{\xi \in \mathbb{R}^n : x^T \xi = 0\}.$$

Let $S^1 = \{x = (x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$. We will apply Algorithm 1 with Armijo step size to find a Pareto critical point of $F = (\bar{f}_1, \bar{f}_2)$. For each $j = 1, 2$, we define $\bar{f}_j : \mathbb{R}^2 \to \mathbb{R}$ by

$$\bar{f}_j(x) = \bar{g}_j(x) + \bar{h}_j(x), \tag{4.51}$$

where $\bar{g}_j : \mathbb{R}^2 \to \mathbb{R}$ is defined by

$$\bar{g}_j(x) = \delta_{S^1}(x) = \begin{cases} 0, & x \in S^1, \\ +\infty, & \text{otherwise}, \end{cases}$$

and $\bar{h}_j : \mathbb{R}^2 \to \mathbb{R}$ is defined by

$$\bar{h}_j(x) = \max_{z \in Z_j} \|(x_1^2 + 3x_2^2) z\|, \tag{4.52}$$

where $Z_j = \{z \in \mathbb{R}^2 | 0 \prec z \preceq 2e\}$, $e = (1,1)^T \in \mathbb{R}^2$. By the definition of $\bar{g}_j$, the optimization problem (4.51) can be rewritten as $f_j(x) : S^1 \to \mathbb{R}$, where

$$x \mapsto f_j(x) = \max_{z \in Z_j} \|(x_1^2 + 3x_2^2) z\|.$$

From Section 4 in [19], we obtain

$$\operatorname{grad} f_j(x) = P_x \bar{f}_j(x) = [\operatorname{grad} \bar{f}_j(x)]^T - xx^T [\operatorname{grad} \bar{f}_j(x)]^T.$$

By setting $A = \operatorname{diag}(1,3)$ and $\zeta = 0.1$, results in Table 1 show that the generalized conditional gradient method is effective under different Armijo condition numbers. In Table 1, ”iter” and ”grad” denote the number of iterations, and the norm of the gradient for the composite optimization problem (4.51), respectively.

TABLE 1. Numerical results for the composite optimization problem (4.51).

$\mu_1 = 0.5, \mu_2 = 0.6$			$\mu_1 = 0.2, \mu_2 = 0.8$			$\mu_1 = 0.3, \mu_2 = 0.7$		
x	iter	grad	x	iter	grad	x	iter	grad
(0.242536, 0.970143)	1	9.4120e-1	(0.242536, 0.970143)	1	9.4120e-1	(0.540232, 0.540232)	2	1.8185e0
(0.841516, 0.540232)	2	1.8185e0	(0.980767, -0.178040)	3	7.0080e-1	(0.998657, 0.051803)	4	2.0690e-1
(0.980760, 0.195170)	3	7.6570e-1	(0.999562, -0.029607)	5	1.1840e-1	(0.999819, -0.019032)	5	7.6100e-2
(0.999471, 0.032526)	4	1.3000e-1	(0.999996, -0.002845)	7	1.1400e-2	(0.999997, -0.002629)	7	1.0500e-2
(0.999990, 0.044516)	5	1.7800e-2	(1.000000, 0.000884)	8	3.5000e-3	(1.000000, -0.000364)	9	1.5000e-3
(1.000000, 0.000605)	6	2.4000e-3	(1.000000, 0.000085)	10	3.4132e-4	(1.000000, 0.000135)	10	5.4212e-4
(1.000000, 0.000082)	7	3.2939e-4	(1.000000, -0.000027)	11	1.0605e-4	(1.000000, 0.000019)	12	7.4904e-5
(1.000000, 0.000011)	8	4.4797e-5	(1.000000, -0.000003)	12	1.0239e-5	(1.000000, 0.000003)	14	1.0360e-5
(1.000000, 0.000002)	9	6.0924e-6	(1.000000, 0.000001)	14	3.1815e-6	(1.000000, -0.000001)	15	3.8560e-6

5. CONCLUSIONS

In this paper, the generalized conditional gradient method for the composite multiobjective optimization problem was proposed on Riemannian manifolds. Under some conditions, the global convergence and the iteration-complexity bound was investigated. Future research directions involve establishing explicit asymptotic and non-asymptotic convergence rates, numerically competitive Riemannian proximal subgradient algorithm, and proximal augmented Lagrangian methods for composite multiobjective optimization problems.

Acknowledgments
The authors are grateful to the reviewers for useful suggestions which improved the contents of this paper. Xiaobo Li was supported by the National Natural Science Foundation of China under Grant No. 11901485, Natural Science Foundation of Sichuan under Grant No. 2023NSFSC1354, No. 2022NSFSC1821, and No. 2022NSFSC0532, and the Fundamental Research Funds for the Central Universities under Grant No. PHD2023-057. Kai Tu was supported by the National Natural Science Foundation of China under Grant No. 12101436.

REFERENCES

[1] G. Eichfrlder, An adaptive scalarization method in multiobjective optimization, SIAM J. Optim. 19 (2009), 1694-1718.
[2] K. Kumar, D. Ghosh, A. Upadhayay, J.C. Yao, X. Zhao, Quasi-Newton methods for multiobjective optimization problems: A systematic review, Appl. Set-Valued Anal. Optim. 5 (2023), 291-321.
[3] K. Mita, E. H. Fukuda and N. Yamashita, Nonmonotone line searches for unconstrained multiobjective optimization problems, J. Global Optim. 75 (2019), 63-90.
[4] K. Miettinen, Nonlinear multiobjective optimization, Vol. 12, Springer Science Business Media, 2012.
[5] P. M. Pardalos, A. Žilinskas, Non-convex multi-objective optimization, Springer International Publishing, 2017.
[6] P. B. Assuncão, O. P. Ferreira, L. F. Prudente, Conditional gradient method for multiobjective optimization, Comput. Optim. Appl. 78 (2021), 741-768.
[7] R. I. Bot, S. M. Grad, Inertial forward-backward methods for solving vetor optimization problems, Optimization 67 (2018), 959-974.
[8] H. Tanabe, E. H. Fukuda, N. Yamashita, Proximal gradient methods for multiobjective optimization and their applications, Comput. Optim. Appl. 72 (2019), 339-361.
[9] B.B. Upadhyay, Lijie Li, Priyanka Mishra, Nonsmooth interval-valued multiobjective optimization problems and generalized variational inequalities on Hadamard manifolds, Appl. Set-Valued Anal. Optim. 5 (2023), 69-84.
[10] D. Burago, Y. Burago, S. Ivanov, A Course in Metric Geometry, Graduate Studies in Math., Vol. 33, Amer. Math. Soc., Providence, RI, 2001.
[11] I. Chavel, Riemannian Geometry-A Modern Introduction, Cambridge University Press, London, 1993.
[12] J. X. da Cruz Neto, P. R. Oliveira, Geodesic Methods in Riemannian Manifolds, Universidade Federal do Rio de Janeiro, Research Report, PESC/COPPE, 1995.
[13] W. Huang, Optimization Algorithms on Riemannian Manifolds with Applications, Ph.D. Thesis, Department of Mathematics, Florida State University, Tallahassee, FL, 2013.
[14] W. Huang, K.A. Gallivan, P.A. Absil, A Broyden class of quasi-newton methods for Riemannian optimization, SIAM J. Optim. 25 (2015), 1660-1685.
[15] W. Klingenberg, A Course in Differential Geometry, Springer, Berlin, 1978.
[16] W. Ring, B. Wirth, Optimization methods on Riemannian manifolds and their application to shape space, SIAM J. Optim. 22 (2012), 596–627.

[17] T. Sakai, Riemannian Geometry, Translations of Mathematical Monographs, American Mathematical Society, Providence, RI, 1996.

[18] H. C. Zhang, W. W. Hager, A Nonmonotone line search technique and its application to unconstrained optimization, SIAM. J. Optim. 14 (2004), 1043-1056.

[19] P. A. Absil, R. Mahony, R. Sepulchre, Optimization Algorithms on Matrix Manifolds, Princeton University Press, 2008.

[20] X. B. Li, M. Krishan Lal, Convergence analysis of a nonmonotone projected gradient method for multiobjective optimization problems on Riemannian manifolds, J. Nonlinear Convex Anal. 23 (2022), 1673-1691.

[21] X. B. Li, X. F. Wang, M. Krishan Lal, A nonmonotone trust region method for unconstrained optimization problems on Riemannian manifolds, J. Optim. Theory Appl. 188 (2021), 547-570.

[22] A. Beck, M. Teboulle, A conditional gradient method with linear rate of convergence for solving convex linear systems, Math. Methods Oper. Res. 59 (2004), 235-247.

[23] K. Bredies, D. A. Lorenz, L. P. Maass, A generalized conditional gradient method and its connection to an iterative shrinkage method, Comput. Optim. Appl. 42 (2009), 173-193.

[24] P. B. Assunc*ão*, O. P. Ferreira, L. F. Prudente, A generalized conditional gradient method for multiobjective composite optimization problems, arXiv preprint: 2302. 12912v1, 2023.

[25] D. Bertsekas, A. Nedic, A. Ozdaglar, Convex Analysis and Optimization, Athena Scientific Optimization and Computation Series, Athena Scientific, 2003.