Applied Analysis and Optimization
Volume 7, Number 3, 2023, 309–328

Yokohama Publishers
ISSN 2189-1664 Online Journal

A NONMONOTONE QUASI-NEWTON METHOD FOR MULTIOBJECTIVE OPTIMIZATION PROBLEMS ON RIEMANNIAN MANIFOLDS

XIAOBO LI* AND MANISH KRISHAN LAL†

Abstract. In this paper, we propose a nonmonotone quasi-Newton method for multiobjective optimization problems defined on Riemannian manifolds. Under some reasonable conditions, the global convergence result is established, and the superlinear local convergence rate of the proposed method to a Pareto critical point is also proved.

1. Introduction

Many practical problems in information and management sciences, engineering design, architecture, and machine learning can be modeled as multiobjective problems, see [8, 10, 19–21] and references therein. Due to competing multiple objectives in these problems, no single point can be a minimizer for all objective functions. Thus, we seek a Pareto optimal point [8, 19] and the necessary optimality conditions to find an optimal solution. Recently, some authors focused on extending the optimization methods from Euclidean spaces to Riemannian manifolds; see, e.g., [3, 7, 11, 12, 24]. As we know, the generalization of optimization methods from Euclidean spaces to Riemannian manifolds has some important advantages, for example, nonconvex probelms can be translated into convex problems with appropriate Riemannian metrics, and constrained problems may become unconstrained problems from the Riemannian geometry viewpoint.

The development of algorithms for multiobjective optimization problems on Riemannian manifolds is rather very new. Some of the methods include the steepest descent method [3, 15], Newton's method [1, 11], projected gradient method [16] and trust-region method [17]. The monotone algorithms do not allow an increase in the values of the objective function in successive iterations, by generating a monotonically non-increasing sequence of objective-function values. This property which is important for proving convergence, however, also could lead to relatively short step sizes and thus motivates to consider nonmonotone algorithms, where this monotonicity requirement is relaxed and imposing nonmonotonicity on function values

2010 *Mathematics Subject Classification.* 65K05 · 65K10 · 90C48 · 49J40.

Key words and phrases. Multiobjective optimization problems; nonmonotone technique; Quasi-Newton method; global convergence; Riemannian manifolds.

*Xiaobo Li was supported by the Young Scientists Fund of the National Natural Science Foundation of China (No. 11901485), Natural Science Foundation of Sichuan(No. 2023NSFSC1354, No. 2022NSFSC0532), the Fundamental Research Funds for the Central Universities (No. PHD2023-057).

†M. Krishan Lal was supported by SERB-UBC fellowship.

have turned out to be a key property for enhancing the performance of the algorithms, see [16, 17, 20, 27]. Nonmonotone line search methods for multiobjective programs were first proposed by Mita et al. [20], where the authors established global convergence of the steepest descent method and Newton methods. Other related works [18] and [9, 22] discuss the global convergence of the Newton method and the local superlinear rate of convergence for the quasi-Newton method, respectively. 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.

Motivated by the results described above, in this paper, building on the approach of [9] and [22], we present a quasi-Newton method for multiobjective optimization problems equipped with a nonmonotone line search strategy on Riemannian manifolds. Under some reasonable conditions, we establish global convergence and the superlinear local convergence rate results. Since a Riemannian manifold, in general, does not have a linear structure, the 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, by utilizing a nonmonotone inexact line search technique, we extend the Newton method presented in [9] and [22] from $\mathbb{R}^n$ to Riemannian manifolds. Second, our results can be viewed as a generalization of results of [1] from single objective functions to multiobjective functions on Riemannian manifolds. Third, in this paper, we replace exponential mappings with retractions and utilize parallel transports for isometric vector transport, respectively, enhancing the efficiency of the method.

The remainder of our work is organized as follows: In Section 2, we provide the necessary notations, definitions, and concepts on Riemannian manifolds. In Section 4, we present the nonmonotone quasi-Newton method for multiobjective optimization problems on Riemannian manifolds. In Section 5 and Section 6, under certain reasonable conditions, we present global convergence results and demonstrate the attainment of a superlinear local convergence rate for the proposed nonmonotone quasi-Newton method applied to multiobjective optimization problems on Riemannian manifolds.

2. Preliminaries from Riemannian geometry

Throughout this paper, we assume that M is a finite-dimensional differentiable manifold. We denote by T_xM the tangent space of M at x with inner product $\langle \cdot, \cdot \rangle_x$ and induced norm $\| \cdot \|_x$.

The tangent bundle of M is denoted by $TM = \bigcup_{x \in M} T_xM$. We assume the reader is familiar with the basic knowledge of Riemannian manifolds, see, e.g., [5, 13]. 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, i.e., $\gamma(t_0) = x$ and $\gamma(t_1) = y$, we can define the length of γ by $l(\gamma) = \int_a^b \|\gamma'(t)\| dt$. Minimizing this length functional over the set of all curves we obtain a Riemannian distance $d(x, y)$ which induces the original topology on M. Now, we lay out the notations and definitions that are used throughout the paper. Let ∇ be the Levi-Civita connection associated with M. A vector field $V : M \to TM$ along γ is said to be parallel if

$\nabla_{\gamma'}^{V} = 0$. We say that γ is a geodesic when $\nabla_{\gamma'}^{\gamma'} = 0$; in the case $\|\gamma'\| = 1$, γ is said to be normalized.

Definition 2.1. Let $\gamma : \mathbb{R} \to M$ be a geodesic and let P denote the parallel transport along γ, which is defined by

$$P_{\gamma(a),\gamma(b)}(v) = V(\gamma(b)) \quad \forall\, a, b \in \mathbb{R}, \quad \text{and } v \in T_{\gamma(a)}M,$$

where V is the unique C^{∞} vector field satisfying $\nabla_{\gamma'(t)}V = 0$ and $V(\gamma(a)) = v$.

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 [25], any pair of points $x, y \in M$ can be joined by a minimal geodesic.

Definition 2.2. 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 exponential mapping $\exp_x$ provides a local parametrization of M via T_xM.

Definition 2.3. 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 identity mapping.

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

Parallel transport is often too expensive to compute in a practical method, so we can consider a more general vector transport (see, for example, [12, 24]), which is built upon the retraction R_x.

Definition 2.4. A vector transport $\mathcal{T} : TM \bigoplus TM \to TM$, $(\eta_x, \xi_x) \mapsto \mathcal{T}_{\eta_x}\xi_x$ with the associated retraction R_x is a smooth mapping such that, for all η_x in the domain of R_x and $\forall\, \xi_x, \zeta_x \in T_xM$,

(i) $\mathcal{T}_{\eta_x}\xi_x \in T_{R_x(\eta_x)}M$,
(ii) $\mathcal{T}_{0_x}\xi_x = \xi_x$,
(iii) $\mathcal{T}_{\eta_x}$ is a linear mapping.

Definition 2.5. A vector transport $\mathcal{T}$ is called an isometric vector transport denoted as $\mathcal{T}_S$, with an associated retraction R_x, if in addition to properties (i), (ii), (iii) from Definition 2.4, it satisfies

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

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

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

Let $L(TM, TM)$ denote fiber bundle with base space $M \times M$ such that the fiber over $(x, y) \in M \times M$ is $L(T_xM, T_yM)$, the set of all linear mappings from T_xM to T_yM. From [5, Section 4], we recall that a transporter L on M is a smooth section of the bundle $L(TM, TM)$. Furthermore, $L^{-1}(x, y) = L(y, x)$ and $L(x, z) = L(y, z)L(x, y)$.

Remark 2.6. Given a retraction R_x, for any $\eta_x, \xi_x \in T_xM$, the isometric vector transport $\mathcal{T}_S$ can be represented by

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

In this paper, from the locking condition proposed by Huang [11], we require

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

For the Stiefel manifold and the Grassman manifold, there always exist retractions such that the above equality holds, see e.g., [11]. Furthermore, from the above results, we have

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

3. Auxillary results

In this paper, let $I := \{1, 2, \dots, m\}$, $\mathbb{R}^m_+ = \{x \in \mathbb{R}^m \mid x_i \geq 0, i \in I\}$ and $\mathbb{R}^m_{++} = \{x \in \mathbb{R}^m \mid x_i > 0, i \in I\}$. 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 $y - x \in \mathbb{R}^m_{++}$.

In this paper, we consider the following problem:

$$\min_{x \in M} F(x),$$

where $F : M \to \mathbb{R}^m$, $F(x) = (f_1(x), f_2(x), \dots, f_m(x))$ is a twice continuously differentiable vectorial function in M. Recall that given a smooth function f defined on M, the gradient of f at x, denoted by $\mathrm{grad} f(x)$, is defined as the unique element of T_xM, that satisfies

$$\langle \mathrm{grad} f(x), d\rangle = Df(x)[d], \ \forall d \in T_xM.$$

We denote the Riemannian jacobian of F by

$$\mathrm{grad}F(x) := (\mathrm{grad} f_1(x), \dots, \mathrm{grad} f_m(x)),$$

and the image of the Riemannian jacobian of F at a point $x \in M$ by

$$\mathrm{Im}(\mathrm{grad}F(x)) := \mathrm{grad}F(x)v = \{(\langle \mathrm{grad} f_1(x), v\rangle, \dots, \langle \mathrm{grad} f_m(x), v\rangle) \mid \forall v \in T_xM\}.$$

Definition 3.1. A point $x \in M$ is a Pareto critical point of F if and only if

$$\mathrm{Im}(\mathrm{grad}F(x)) \bigcap (-\mathbb{R}^m_{++}) = \emptyset, x \in M.$$

Let $x \in M$ be a point that is not Pareto critical. Then, there exists a direction $v \in T_xM$, such that

$$\mathrm{grad}F(x)v = (\langle \mathrm{grad} f_1(x), v\rangle, \dots, \langle \mathrm{grad} f_m(x), v\rangle) \in -\mathbb{R}^m_{++}.$$

Clearly, $\forall i = 1, \dots, m$, we have $\langle \mathrm{grad} f_i(x), v\rangle < 0$.

In this case, v is called a descent direction for F at $x \in M$. In this paper, we call $v \in T_xM$ the quasi-Newton's direction if it is the optimal solution to the problem:

$$\min_{v \in T_xM} \max_{i=1,\dots,m} \{\langle \mathrm{grad} f_i(x), v\rangle + \frac{1}{2}\langle B_i(x)v, v\rangle\}, \tag{3.1}$$

where $B_i(x)$ is some approximation of Hess$f_i(x)$ and Hess$f_i(x)$ is the Hessian matrix of $f_i(x)$ for $i = 1, \dots, m$. Note that Hess$f_i(x)$ is a matrix in T_xM.

The solution and optimal value will be denoted by

$$\tau(x) := \min_{v \in T_xM} \max_{i=1,\dots,m} \{\langle \mathrm{grad} f_i(x), v\rangle + \frac{1}{2}\langle B_i(x)v, v\rangle\}, \tag{3.2}$$

$$v(x) := \arg\min_{v \in T_xM} \max_{i=1,\dots,m} \{\langle \mathrm{grad} f_i(x), v\rangle + \frac{1}{2}\langle B_i(x)v, v\rangle\}. \tag{3.3}$$

If we assume that $B_i(x), i = 1, \dots, m$ are positive definite, then according to (3.2) and (3.3), we have $\tau(x) \geq \langle \mathrm{grad} f_i(x), v(x)\rangle$ for all $i = 1, \dots, m$ and $x \in M$.

Lemma 3.2. Let $B_i(x), i = 1, \dots, m$ be symmetric positive definite for all $x \in M$ and consider τ as defined by (3.2). Then,

(i) For all $x \in M$, $\tau(x) \leq 0$;

(ii) The following conditions are equivalent:

(a) The point $x \in M$ is not Pareto critical point;

(b) $\tau(x) < 0$;

(c) $v(x) \neq 0$.

In particular, $x \in M$ is a Pareto critical point if and only if $\tau(x) = 0$.

Proof. For (i), observe that

$$\begin{aligned}\tau(x) &= \min_{v \in T_xM} \max_{i=1,\dots,m} \{\langle \mathrm{grad} f_i(x), v\rangle + \frac{1}{2}\langle B_i(x)v, v\rangle\}\\ &\leq \max_{i=1,\dots,m} \{\langle \mathrm{grad} f_i(x), 0\rangle + \frac{1}{2}\langle B_i(x)0, 0\rangle\} = 0.\end{aligned}$$

For (ii), we prove $(a) \Rightarrow (b)$. If $x \in M$ is not a Pareto critical point, then there exists a direction $\bar{v} \in T_xM$ such that

$$\langle \mathrm{grad} f_i(x), \bar{v}\rangle < 0, \forall i = 1, \dots, m,$$

and subsequently, we also have

$$\max_{i=1,\dots,m} \{\langle \mathrm{grad} f_i(x), \bar{v}\rangle\} < 0.$$

Note that $l\bar{v} \in T_xM$ for all $l > 0$. Thus,

$$\begin{aligned}\tau(x) &\leq \max_{i=1,\dots,m} \{\langle \mathrm{grad} f_i(x), l\bar{v}\rangle + \frac{1}{2}\langle B_i(x)l\bar{v}, l\bar{v}\rangle\}\\ &= l \max_{i=1,\dots,m} \{\langle \mathrm{grad} f_i(x), \bar{v}\rangle + \frac{1}{2}l\langle B_i(x)\bar{v}, \bar{v}\rangle\}.\end{aligned}$$

Observing $B_i(x), i = 1, \dots, m$ are positive definite, we obtain $\langle B_i(x)\bar{v}, \bar{v}\rangle > 0$ for all $\bar{v} \neq 0$. When we take $0 < l < \Gamma$ where

$$\Gamma := \frac{-2\max_{i=1,\dots,m}\langle \mathrm{grad} f_i(x), \bar{v}\rangle}{\max_{i=1,\dots,m}\langle B_i(x)\bar{v}, \bar{v}\rangle},$$

we can conclude that $\tau(x) < 0$.

$(b) \Rightarrow (c)$. If $v(x) = 0$, we obtain that $\tau(x) = 0$. Then, it follows from $\tau(x) < 0$ that $v(x) \neq 0$.

$(c) \Rightarrow (a)$. If $v(x) \neq 0$, then

$$\langle \mathrm{grad} f_i(x), v(x)\rangle + \frac{1}{2}\langle B_i(x)v(x), v(x)\rangle < 0, \forall i = 1, \ldots, m.$$

Thus, $\langle \mathrm{grad} f_i(x), v(x)\rangle < 0, \forall i = 1, \ldots, m$. In this case, x is not a Pareto critical point of F. It is obvious to check that $x \in M$ is a Pareto critical point if and only if $\tau(x) = 0$. □

In this paper, from the similar assumption in [15, Theorem 4.3] and [17, Theorem 4.4] on Riemannian manifolds, we assume the following assumptions hold.

(1) The functions $\mathrm{Hess} f_i(x)$, $i = 1, \ldots, m$ are uniformly continuous, *i.e.,* for all $\epsilon > 0$, there exists $\delta > 0$ such that for all $x, y \in M$, with $d(x, y) < \delta$, we have

$$\|L(x,y)\mathrm{Hess} f_i(x)v - \mathrm{Hess} f_i(y)L(x,y)v\| < \frac{\epsilon}{2}\|v\|, \ \forall v \in T_xM. \tag{3.4}$$

(2) For all $\epsilon > 0$, there exists $\delta > 0$ such that for all $x, y \in M$, with $d(x, y) < \delta$, we have for all $i = 1, \ldots, m$,

$$\|\mathrm{Hess} f_i(y)L(x,y)v - L(x,y)B_i(x)v\| < \frac{\epsilon}{2}\|v\|, \ \forall v \in T_xM. \tag{3.5}$$

These assumptions also extend the assumption in [9, Lemma 4.1] and of [22, Section 4] from $\mathbb{R}^n$ to Riemannian manifolds.

Lemma 3.3. The function $\tau(x)$ is continuous.

Proof. We show that τ is continuous on a fixed arbitrarily chosen compact set $W \subseteq M$. Due to Lemma 3.2 (i), for any $x \in M, i = 1, \ldots, m$,

$$\frac{1}{2}\langle B_i(x)v(x), v(x)\rangle \leq -\langle \mathrm{grad} f_i(x), v(x)\rangle. \tag{3.6}$$

Since F is twice continuously differentiable and $B_i(x)$ is positive definite for all $x \in W$ and $i = 1, \ldots, m$, there exist $K, L > 0$ such that

$$K = \max_{i=1,\ldots,m}\{\|\mathrm{grad} f_i(x)\|, L\} = \min_{i=1,\ldots,m}\langle B_i(x)e, e\rangle, \ \forall x \in M,$$

where $e \in T_xM$ and $\|e\| = 1$. Combining (3.6) and using Cauchy–Schwarz inequality, we get

$$\frac{1}{2}L\|v(x)\|^2 \leq \|\mathrm{grad} f_i(x)\|\|v(x)\| \leq K\|v(x)\|, \quad \forall\ x \in W \text{ and } i = 1, \ldots, m.$$

Hence,

$$\|v(x)\| \leq 2(K/L), \quad \forall x \in W. \tag{3.7}$$

Now, define a family of functions $\{\phi_{x,i}\}_{x\in W}, i = 1, \ldots, m$, where

$$\phi_{x,i}(z) := \langle \mathrm{grad} f_i(z), L(x,z)v(x)\rangle + \frac{1}{2}\langle B_i(z)L(x,z)v(x), L(x,z)v(x)\rangle,$$

by adding and subtracting the term $\frac{1}{2}\langle \mathrm{Hess} f_i(z)L(x,z)v(x), L(x,z)v(x)\rangle$, we get

$$\phi_{x,i}(z) = \langle \mathrm{grad} f_i(z), L(x,z)v(x)\rangle + \frac{1}{2}\langle \mathrm{Hess} f_i(z)L(x,z)v(x), L(x,z)v(x)\rangle$$

$$+\frac{1}{2}\langle B_i(z)-\mathrm{Hess}f_i(z)L(x,z)v(x),L(x,z)v(x)\rangle. \tag{3.8}$$

Since the first two terms of (3.8) are continuous, only the last term must be considered.

Define an open cover $\{B(z,\delta_z)\}_{z\in W}$ with radius $\delta_z>0$. From (3.4), (3.5) and $L(w,z)L(z,w)=\mathrm{Id}$, we have for all $z\in W$ and small $\epsilon_z>0$, there exists some $\delta_z>0$ such that for all $w\in B(z,\delta_z)$,

$$\|L(w,z)[B_i(w)L(z,w)v]-\mathrm{Hess}f_i(z)v\|\le\frac{\epsilon_z}{2}\|v\|,\ \forall v\in T_zM, \tag{3.9}$$

and

$$\|L(w,z)[\mathrm{Hess}f_i(w)L(z,w)v]-\mathrm{Hess}f_i(z)v\|\le\frac{\epsilon_z}{2}\|v\|,\ \forall v\in T_zM. \tag{3.10}$$

Regarding the last term of (3.8), for $w_1,w_2\in B(z,\delta_z)$ with $d(w_1,w_2)<\delta$, where $\delta>0$ is small enough, by (3.7), (3.8) and (3.9),

$$\begin{aligned}
&|\frac{1}{2}\langle(B_i(w_1)-\mathrm{Hess}f_i(w_1))L(x,w_1)v(x),L(x,w_1)v(x)\rangle\\
&-\frac{1}{2}\langle(B_i(w_2)-\mathrm{Hess}f_i(w_2))L(x,w_2)v(x),L(x,w_2)v(x)\rangle|\\
\le\ &\frac{1}{2}\|v(x)\|\|(B_i(w_1)-\mathrm{Hess}f_i(w_1))L(x,w_1)v(x)\|\\
&+\frac{1}{2}\|v(x)\|\|(B_i(w_2)-\mathrm{Hess}f_i(w_2))L(x,w_2)v(x)\|\\
\le\ &\frac{1}{2}\|v(x)\|[\|L(w_1,z)[B_i(w_1)L(x,w_1)v(x)]-\mathrm{Hess}f_i(z)L(x,z)v(x)\|\\
&+\|\mathrm{Hess}f_i(z)L(x,z)v(x)-L(w_1,z)[\mathrm{Hess}f_i(w_1)L(x,w_1)v(x)]\|\\
&+\|L(w_2,z)[B_i(w_2)L(x,w_1)v(x)]-\mathrm{Hess}f_i(z)L(x,z)v(x)\|\\
&+\|\mathrm{Hess}f_i(z)L(x,z)v(x)-L(w_2,z)[\mathrm{Hess}f_i(w_2)L(x,w_2)v(x)]\|]\\
\le\ &\frac{1}{2}\|v(x)\|[\frac{\epsilon_z}{2}\|v(x)\|+\frac{\epsilon_z}{2}\|v(x)\|+\frac{\epsilon_z}{2}\|v(x)\|+\frac{\epsilon_z}{2}\|v(x)\|]=\epsilon_z\|v(x)\|^2\\
\le\ &4\epsilon_z\frac{K^2}{L^2}.
\end{aligned}$$

Thus, $\phi_{x,i}$ is uniformly continuous for all $x\in W$ and $i=1,\dots,m$, and so $\phi_x=\max_{i=1,\dots,m}\phi_{x,i}$ is also uniformly continuous. Take $\epsilon>0$, there exists $\delta>0$ such that for all $y_1,y_2\in W, d(y_1,y_2)<\delta$ implies $|\phi_x(y_1)-\phi_x(y_2)|<\epsilon$ for all $x\in W$. Thus, for $d(y_1,y_2)<\delta$, we have

$$\begin{aligned}
\tau(y_2)\le&\max_{i=1,\dots,m}\{\langle\mathrm{grad}f_i(y_2),L(y_1,y_2)v(y_1)\rangle\\
&\qquad+\frac{1}{2}\langle B_i(y_2)L(y_1,y_2)v(y_1),L(y_1,y_2)v(y_1)\rangle\}\\
=&\ \phi_{y_1}(y_2)\le\phi_{y_1}(y_1)+|\phi_{y_1}(y_2)-\phi_{y_1}(y_1)|<\tau(y_1)+\epsilon.
\end{aligned}$$

Hence, $\tau(y_2)-\tau(y_1)<\epsilon$. Similarly, we can obtain $|\tau(y_2)-\tau(y_1)|<\epsilon$, which implies the continuity of τ. □

4. Nonmonotone quasi-Newton's algorithm

Now, based on the above results, we introduce the nonmonotone quasi-Newton's algorithm for multiobjective optimization on Riemannian manifolds, which extends the quasi-Newton method in [9, 18, 22] from $\mathbb{R}^n$ to Riemannian manifolds.

Algorithm 1. (Nonmonotone quasi-Newton's algorithm for multiobjective optimization on Riemannian manifolds)

- choose $\sigma \in (0,1), \beta \in (0,1), Q_0 = 1, 0 \leq \eta_{\min} \leq \eta_{\max} \leq 1, \eta_0 \in [\eta_{\min}, \eta_{\max}]$ and $\epsilon > 0$. Let $x_0 \in M, C_i^0 = f_i(x_0)$ and symmetric positive definite matrix $B_i(x_0) = I$ in $T_{x_0}M$ for $i = 1, \dots, m$;
- set $k = 0$;
- **while** $\tau(x_k) > -\epsilon$ **do**, where

$$v(x_k) = \operatorname{argmin}_{v \in T_{x_k}M} \max_{i=1,\dots,m} \{\langle \operatorname{grad} f_i(x_k), v\rangle + \frac{1}{2}\langle B_i(x_k)v, v\rangle\},$$

$$\tau(x_k) = \min_{v \in T_{x_k}M} \max_{i=1,\dots,m} \{\langle \operatorname{grad} f_i(x_k), v\rangle + \frac{1}{2}\langle B_i(x_k)v, v\rangle\};$$

- set $x_{k+1} = R_{x_k}\alpha_k v(x_k)$, where α_k is determined by the nonmonotone inexact line search rule: for all $i = 1, \dots, m$, choose α_k as the largest one in $\{1, \beta, \beta^2, \dots\}$ such that

$$f_i(x_{k+1}) \leq C_i^k + \sigma\alpha_k\tau(x_k),$$

where the cost update is defined as

$$C_i^{k+1} := \frac{\eta_k Q_k C_i^k + f_i(x_{k+1})}{Q_{k+1}},$$

and $Q_{k+1} := \eta_k Q_k + 1$. Generate η_{k+1} by an adaptive formula (see [27]) and $B_i(x_{k+1})$ by using the BFGS quasi-Newton algorithm or different variants of quasi-Newton algorithms (see [12, 14, 23]);

- $k = k + 1$;

- **end while**

5. 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.

Similar to the proof of in [20, Lemma 3.2], we obtain the the following result.

Lemma 5.1. Let $\{x_k\} \subseteq M$ be a sequence generated by Algorithm 1. Then, for any $i = 1, \dots, m$,

$$f_i(x_k) \leq C_i^k.$$

Moreover, if x_k is not a Pareto critical point of F, then there exists α_k satisfying the nonmonotone inexact conditions of the line search procedure such that

$$f_i(x_{k+1}) \leq C_i^k + \sigma\alpha_k\tau(x_k). \tag{5.1}$$

Proof. From Lemma 3.2, we obtain $f_i(x_k) \le C_i^k,\ \forall i = 1, \dots, m$. Now, we show that there exists α_k satisfying the nonmonotone conditions. Define $g_i^k(t) := f_i(R_{x_k} tv(x_k))$. By Taylor's theorem, we have for all $i = 1, \dots, m$,

$$g_i^k(t) = g_i^k(0) + t\frac{dg_i^k(0)}{dt} + o(t).$$

That is, for all $i = 1, \dots, m$,

$$\begin{aligned} f_i(R_{x_k}\alpha_k v(x_k)) &= f_i(x_k) + \alpha_k\langle \mathrm{grad} f_i(x_k), DR_{x_k}0v(x_k)[v(x_k)]\rangle + o(\alpha_k) \\ &= f_i(x_k) + \alpha_k\langle \mathrm{grad} f_i(x_k), v(x_k)\rangle + o(\alpha_k). \end{aligned}$$

Since $\langle \mathrm{grad} f_i(x_k), v(x_k)\rangle \le \tau(x_k)$ and $f_i(x_k) \le C_i^k,\ i = 1, \dots, m$, we have

$$\begin{aligned} f_i(R_{x_k}\alpha_k v(x_k)) &\le C_i^k + \alpha_k\tau(x_k) + o(\alpha_k) \\ &= C_i^k + \alpha_k\sigma\tau(x_k) + \alpha_k[(1-\sigma)\tau(x_k) + \frac{o(\alpha_k)}{\alpha_k}]. \end{aligned}$$

Observing that x_k is not a Pareto critical point, $\tau(x_k) < 0$, and thus for $\alpha_k > 0$ small enough, we obtain (5.1) holds. ☐

Lemma 5.2. Suppose that Algorithm 1 is employed and $\mathrm{grad} f_i(x), i = 1, \dots, m$ satisfying the following Lipschitz condition with Lipschitz constant $\tilde{L}$, that is,

$$\|L(x, x_k)\mathrm{grad} f_i(x) - \mathrm{grad} f_i(x_k)\| \le \tilde{L}d(x, x_k),$$

for all x on the segment connecting x_k and $R_{x_k}\alpha_k\beta^{-1}v(x_k)$ if $\alpha_k \le \beta$. Then

$$\alpha_k \ge \min\{\beta, \frac{2\beta(1-\sigma)}{\tilde{L}}\frac{|\tau(x_k)|}{\|v(x_k)\|^2}\}. \tag{5.2}$$

Proof. If $\alpha_k \ge \beta$, then (5.2) trivially holds. If $\alpha_k < \beta$, then from (5.1) and Lemma 5.1, we have for all $i = 1, \dots, m$,

$$\begin{aligned} f_i(R_{x_k}\alpha_k\beta^{-1}v(x_k)) &> C_i^k + \sigma\alpha_k\beta^{-1}\tau(x_k) \\ &\ge f_i(x_k) + \sigma\alpha_k\beta^{-1}\tau(x_k). \end{aligned} \tag{5.3}$$

Defining $\phi_i^k(t) := f_i(R_{x_k}tv(x_k))$, since $\mathrm{grad} f_i(x),\ i = 1, \dots, m$ are Lipschitz continuous, we have

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

Clearly,

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

This implies that for all $i = 1, \dots, m$,

$$\begin{aligned} & f_i(R_{x_k}\alpha_k\beta^{-1}v(x_k)) - f_i(x_k) \\ = \ & \alpha_k\beta^{-1}\langle \mathrm{grad} f_i(x_k), DR_{x_k}0v(x_k)[v(x_k)]\rangle \\ + \ & \int_0^{\frac{\alpha_k}{\beta}}[\langle \mathrm{grad} f_i(R_{x_k}tv(x_k)), DR_{x_k}tv(x_k)[v(x_k)]\rangle \\ & -\langle \mathrm{grad} f_i(x_k), DR_{x_k}tv(x_k)[v(x_k)]\rangle]dt \end{aligned}$$

$$
\begin{aligned}
&= \alpha_k\beta^{-1}\langle \mathrm{grad} f_i(x_k), v(x_k)\rangle \\
&+ \int_0^{\frac{\alpha_k}{\beta}} \langle \mathrm{grad} f_i(R_{x_k} tv(x_k)) \\
&\quad -L(x_k, R_{x_k} tv(x_k))\mathrm{grad} f_i(x_k), L(x_k, R_{x_k} tv(x_k))v(x_k)\rangle dt \\
&\leq \alpha_k\beta^{-1}\langle \mathrm{grad} f_i(x_k), v(x_k)\rangle \\
&+ \int_0^{\frac{\alpha_k}{\beta}} \|\mathrm{grad} f_i(R_{x_k} tv(x_k)) \\
&\quad -L(x_k, R_{x_k} tv(x_k))\mathrm{grad} f_i(x_k)\| \|L(x_k, R_{x_k} tv(x_k))v(x_k)\| dt \\
&\leq \alpha_k\beta^{-1}\langle \mathrm{grad} f_i(x_k), v(x_k)\rangle + \int_0^{\frac{\alpha_k}{\beta}} t\tilde{L}\|v(x_k)\|^2 dt \\
&\leq \alpha_k\beta^{-1}\tau(x_k) + \frac{\tilde{L}}{2}[\alpha_k\beta^{-1}]^2\|v(x_k)\|^2.
\end{aligned}
$$

This together with (5.3) imply (5.2). □

Theorem 5.3. Suppose that $f_i(x)$ $\forall i = 1, \ldots, m$ are bounded from below and $\eta_{\max} < 1$, then there exists $c > 0$ such that

$$|\tau(x_k)| \geq c\|v(x_k)\|^2, \ k \in \mathbb{N}.$$

Moreover, assume that the assumptions of Lemma 5.2 hold. Then, every limit point of the sequence $\{x_k\}$ generalized by Algorithm 1 is a Pareto critical point of F.

Proof. We show that for all $i = 1, \ldots, m$,

$$f_i(x_{k+1}) \leq C_i^k - \rho|\tau(x_k)|, \tag{5.4}$$

where $\rho = \min\{\sigma\beta, (2\sigma\beta(1-\sigma)c)/\tilde{L}\}$. The following two cases are possible.

Case 1. If $\alpha_k \geq \beta$, then by (5.1), it follows that for all $i = 1, \ldots, m$,

$$
\begin{aligned}
f_i(x_{k+1}) &\leq C_i^k + \sigma\alpha_k\tau(x_k) \\
&= C_i^k - \sigma\alpha_k|\tau(x_k)| \\
&\leq C_i^k - \sigma\beta|\tau(x_k)|,
\end{aligned}
$$

which implies (5.4).

Case 2. If $\alpha_k < \beta$, then by (5.2),

$$\alpha_k \geq \frac{2\beta(1-\sigma)}{\tilde{L}} \frac{|\tau(x_k)|}{\|v(x_k)\|^2},$$

and by (5.1), we have for all $i = 1, \ldots, m$,

$$f_i(x_{k+1}) \leq C_i^k - \frac{2\sigma\beta(1-\sigma)}{\tilde{L}} \frac{|\tau(x_k)|^2}{\|v(x_k)\|^2}.$$

Since $|\tau(x_k)| \geq c\|v(x_k)\|^2$, this implies that for all $i = 1, \ldots, m$,

$$f_i(x_{k+1}) \leq C_i^k - \frac{2\sigma\beta(1-\sigma)c}{\tilde{L}}|\tau(x_k)|,$$

which implies (5.4). Combining the cost update relation defined in Algorithm 1 and (5.4), we have for all $i = 1, \dots, m$,

$$\begin{aligned} C_i^{k+1} &= \frac{\eta_k Q_k C_i^k + f_i(x_{k+1})}{Q_{k+1}} \\ &\le \frac{\eta_k Q_k C_i^k + C_i^k - \rho|\tau(x_k)|}{Q_{k+1}} \\ &= C_i^k - \frac{\rho|\tau(x_k)|}{Q_{k+1}}. \end{aligned} \tag{5.5}$$

Since $f_i(x)$ is bounded from below and $f_i(x_k) \le C_i^k$ for all $k \in N$ and $i = 1, \dots, m$, we conclude that C_i^k is bounded from below. It follows from (5.5) that

$$\sum_{k=0}^{\infty} \frac{|\tau(x_k)|}{Q_{k+1}} = \sum_{k=0}^{\infty} \frac{1}{\rho}(C_i^k - C_i^{k+1}) < +\infty. \tag{5.6}$$

Suppose that x^* is a limit point of $\{x_k\}$. Assume that the subsequence $\{x_k\}_{k\in K}$ converges to x^*, we establish $\tau(x^*) = 0$. By contradiction, assume that $\tau(x^*) < 0$, which implies that there exists $\epsilon > 0, \delta > 0$ such that for all $k \in K$, $d(x_k, x^*) \le \delta$, we have $|\tau(x_k)| \ge \epsilon > 0$. This implies that

$$\sum_{k=0}^{\infty} \frac{|\tau(x_k)|}{Q_{k+1}} \ge \sum_{k\in\{k\in K|d(x_k,x^*)\le\delta\}} \frac{\epsilon}{Q_{k+1}}. \tag{5.7}$$

Since $\eta_{\max} < 1$, from [27, Theorem 2.2] , we have

$$Q_{k+1} \le \frac{1}{1-\eta_{\max}}. \tag{5.8}$$

Consequently, following (5.8), we have

$$\sum_{k=0}^{\infty} \frac{|\tau(x_k)|}{Q_{k+1}} \ge \sum_{k\in\{k\in K|d(x_k,x^*)\le\delta\}} (1-\eta_{\max})\epsilon = +\infty.$$

This contradicts to (5.6). Thus, we have $\tau(x^*) = 0$. From Lemma 3.2, we obtain x^* is a Pareto critical point of F. □

Remark 5.4. If $M = \mathbb{R}^n$ and $R_x\eta = x + \eta$, then Theorem 5.3 can reduce to Theorem 4.2 of [18]. Moreover, Theorem 5.3 can be regarded as a generalization of Section 5 of [1] from single objective functions to multiobjective functions on Riemannian manifolds.

6. Local superlinear convergence

In this section, under some assumptions, we prove that the sequence $\{x_k\}$ generated by Algorithm 1 converges to a Pareto critical point of F superlinearly, which generates the results in [9, 18, 22] from $\mathbb{R}^n$ to Riemannian manifolds.

Observe that (3.1) is equivalent to

$$\begin{aligned} &\min g(t, v) = t \\ \text{s.t. } &\langle \mathrm{grad} f_i(x), v\rangle + \frac{1}{2}\langle B_i(x)[v], v\rangle - t \le 0, \ \forall i = 1, \dots, m. \end{aligned}$$

The Lagrangian of this problem is

$$\mathcal{L}((t,s),\lambda) = t + \sum_{i=1}^{m} \lambda_i(\langle \mathrm{grad} f_i(x), v\rangle + \frac{1}{2}\langle B_i(x)[v], v\rangle - t).$$

Direct calculation of the KKT conditions yields

$$\sum_{i=1}^{m} \lambda_i = 1,$$

and

$$\sum_{i=1}^{m} \lambda_i(\mathrm{grad} f_i(x) + B_i(x)[v]) = 0.$$

This implies that

$$\sum_{i=1}^{m} \lambda_i B_i(x)[v] = -\sum_{i=1}^{m} \lambda_i \mathrm{grad} f_i(x). \tag{6.1}$$

Similar to the proof of lemmas 4.2 and 4.3 in [22], we have the following result.

Lemma 6.1. Assume that the sequence $\{x_k\}$ is generated by Algorithm 1 and $a > 0$ such that for all $i = 1, \ldots, m$ and $k \in N$, we have

$$\langle B_i(x_k)z, z\rangle \geq a\|z\|^2, \ \forall z \in T_{x_k}M.$$

Then, we have
 (a) $|\tau(x_k)| \geq \frac{a}{2}\|v(x_k)\|^2$;
 (b) $|\tau(x_k)| < \frac{1}{2a}\|\sum_{i=1}^{m} \lambda_i \mathrm{grad} f_i(x_k)\|^2$, for all $\lambda_i \geq 0, \ i = 1, \ldots, m$ with $\sum_{i=1}^{m} \lambda_i = 1$.

In this section, similar to the assumptions in theorem 4.4 of [17] on Riemannian manifolds, we assume the following assumptions hold,

$$\lim_{k\to\infty} \frac{\|\mathrm{Hess} f_i(x^*)L(x_k, x^*)v(x_k) - L(x_k, x^*)[B_i(x_k)v(x_k)\|]}{\|v(x_k)\|} = 0, \tag{6.2}$$

where x^* is a Pareto critical point of F, $\{x_k\}$ is the sequence generated by Algorithm 1.

From the similar proof of proposition 4.4 in [18], we have the the following result.

Lemma 6.2. Assume that the sequence $\{x_k\}$ is generated by Algorithm 1 and $a > 0$ such that for all $i = 1, \ldots, m$ and $k \in N$, we have

$$\langle B_i(x_k)z, z\rangle \geq a\|z\|^2, \ \forall z \in T_{x_k}M.$$

Then, we have

$$\lim_{k\to\infty} \tau(x_k) = 0, \tag{6.3}$$

$$\lim_{k\to\infty} \|v(x_k)\| = 0. \tag{6.4}$$

In order to prove the superlinear convergence, we need the following results.

Lemma 6.3. [24] Let $S \subset M$ be an open set and the retraction $R_x : T_xM \to M, x \in S$ has equicontinuous derivatives at x in the sense that

$$\forall \epsilon > 0,\ \exists \delta > 0, \forall x \in S : \|v\| < \delta \Rightarrow \|P_{\gamma[x,R_x(v)]}\mathrm{D}R_x(0) - \mathrm{D}R_x(v)\| < \epsilon.$$

Then, for any $\epsilon > 0$, there exists an $\epsilon' > 0$ such that, for all $x \in S$ and $v, w \in T_xM$ with $\|v\|, \|w\| < \epsilon'$,

$$(1-\epsilon)\|w - v\| \leq d(R_x(v), R_x(w)) \leq (1+\epsilon)\|w - v\|.$$

Theorem 6.4. Let $\{x_k\}$ be a sequence generated by Algorithm 1. Assume that $a > 0$ such that for all $k \in N$ and $i = 1, \ldots, m$,

$$\langle B_i(x_k)z, z\rangle \geq a\|z\|^2,\ \forall z \in T_{x_k}M. \tag{6.5}$$

Also, assume that

$$\frac{D}{dt}DR_{x_k}(tz)[z] = 0,\ \forall z \in T_{x_k}M,$$

where D/dt denotes the covariant derivative along the curve $t \to R_x(tz)$, and all the assumptions of Theorem 5.3 and Lemma 6.3 hold. Then, for sufficiently large k, $\alpha_k = 1$, and the sequence $\{x_k\}$ converges to a Pareto critical point x^*, superlinearly.

Proof. First, by Lemma 6.2, we have

$$\lim_{k\to\infty} \tau(x_k) = 0, \tag{6.6}$$

and

$$\lim_{k\to\infty} v(x_k) = 0. \tag{6.7}$$

By Theorem 5.3, we have $\lim_{k\to\infty} x_k = x^*$, which together with (6.7) and the continuity of $R_{x_k}(\cdot)$ imply that $\lim_{k\to\infty} d(R_{x_k}tv(x_k), x^*) = 0$, where $t \in [0, 1]$. Then, (6.2) implies that

$$\begin{aligned}
&\langle L(x_k, x^*)[B_i(x_k)v(x_k)] - \mathrm{Hess}f_i(x^*)L(x_k, x^*)v(x_k), L(x_k, x^*)v(x_k)\rangle \\
&\leq \|L(x_k, x^*)[B_i(x_k)v(x_k)] - \mathrm{Hess}f_i(x^*)L(x_k, x^*)v(x_k)\|\|v(x_k)\| \\
&= o(\|v(x_k)\|^2).
\end{aligned} \tag{6.8}$$

Second, we show that for k sufficiently large, we have $\alpha_k = 1$.

Defining $\phi_i^k(t) := f_i(R_{x_k}tv(x_k)), t \in [0, 1]$. Clearly,

$$\begin{aligned}
\frac{d^2\phi_i^k(t)}{dt^2} &= \langle \mathrm{Hess}f_i(R_{x_k}tv(x_k))DR_{x_k}tv(x_k)[v(x_k)], DR_{x_k}tv(x_k)[v(x_k)]\rangle \\
&+ \langle \mathrm{grad}f_i(R_{x_k}tv(x_k)), \frac{D}{dt}DR_{x_k}(tv(x_k))[v(x_k)]\rangle.
\end{aligned}$$

Since $\frac{D}{dt}DR_{x_k}(tv(x_k))[v(x_k)] = 0$, we have for all $i = 1, \ldots, m$,

$$\begin{aligned}
\frac{d^2\phi_i^k(t)}{dt^2} &= \langle \mathrm{Hess}f_i(R_{x_k}tv(x_k))DR_{x_k}tv(x_k)[v(x_k)], DR_{x_k}tv(x_k)[v(x_k)]\rangle \\
&= \langle \mathrm{Hess}f_i(R_{x_k}tv(x_k))L(x_k, R_{x_k}tv(x_k))v(x_k), L(x_k, R_{x_k}tv(x_k))v(x_k)\rangle.
\end{aligned}$$

By mean value theorem, (3.4), (6.2) and (6.8), we have for k sufficiently large, $\sigma \in (0,1)$ and $\epsilon > 0$,

$$\begin{aligned}
& \phi_i^k(1) - \phi_i^k(0) \\
= \;& \frac{d\phi_i^k(0)}{dt}(1-0) + \int_0^1 (1-t)\frac{d^2\phi_i^k(t)}{dt^2}dt \\
= \;& \langle \mathrm{grad} f_i(x_k), v(x_k)\rangle \\
+ \;& \int_0^1 (1-t)\langle \mathrm{Hess} f_i(R_{x_k} tv(x_k))L(x_k, R_{x_k} tv(x_k))v(x_k), L(x_k, R_{x_k} tv(x_k))v(x_k)\rangle dt \\
= \;& \langle \mathrm{grad} f_i(x_k), v(x_k)\rangle + \frac{1}{2}\langle B_i(x_k)v(x_k), v(x_k)\rangle \\
+ \;& \int_0^1 (1-t)[\langle \mathrm{Hess} f_i(R_{x_k} tv(x_k))L(x_k, R_{x_k} tv(x_k))v(x_k), L(x_k, R_{x_k} tv(x_k))v(x_k)\rangle \\
- \;& \langle \mathrm{Hess} f_i(x^*)L(x_k, x^*)v(x_k), L(x_k, x^*)v(x_k)\rangle]dt \\
+ \;& \frac{1}{2}\langle \mathrm{Hess} f_i(x^*)L(x_k, x^*)v(x_k) - L(x_k, x^*)[B_i(x_k)v(x_k)], L(x_k, x^*)v(x_k)\rangle \\
\le \;& [\langle \mathrm{grad} f_i(x_k), v(x_k)\rangle + \frac{1}{2}\langle B_i(x_k)v(x_k), v(x_k)\rangle] + \frac{1}{2}\epsilon\|v(x_k)\|^2 + o(\|v(x_k)\|^2) \\
\le \;& \sigma\tau(x_k).
\end{aligned}$$

That is, for all $i = 1, \ldots, m$,

$$f_i(R_{x_k} v(x_k)) \le f_i(x_k) + \sigma\tau(x_k) \le C_i^k + \sigma\tau(x_k).$$

Therefore, for k sufficiently large, $\alpha_k = 1$.

Based on the above results, there exists $k_0 \in N$ such that for all $k \ge k_0$,

$$x_{k+1} = R_{x_k} v(x_k).$$

Define $q(x_{k+1}) := \sum_{i=1}^m \lambda_i(x_k)\mathrm{grad} f_i(x_{k+1})$, where $\lambda_i(x_k), i = 1, \ldots, m$ are defined by the KKT condition in Section 3 of [22]. Note that $\lambda_i(x_k) \ge 0,\ i = 1, \ldots, m$ and $\sum_{i=1}^m \lambda_i(x_k) = 1$. By Lemma 6.1(b), we have

$$\tau(x_{k+1}) \le \frac{1}{2a}\|q(x_{k+1})\|^2. \tag{6.9}$$

Next, define

$$G(x_k) := \sum_{i=1}^m \lambda_i(x_k) f_i(x_k), \tag{6.10}$$

and

$$H(x_k) := \sum_{i=1}^m \lambda_i(x_k) B_i(x_k).$$

We have $q(x_{k+1}) = \mathrm{grad} G(x_{k+1})$ and from (6.1), we have

$$H(x_k)[v(x_k)] = -\mathrm{grad} G(x_k). \tag{6.11}$$

By (3.4), for all $\epsilon > 0$, there exists $\delta > 0$ such that for all $x, y \in M$ with $d(x, y) < \delta$, it is easily shown that

$$\|L(x,y)[\mathrm{Hess}G(x)L(y,x)v] - \mathrm{Hess}G(y)v\| \leq \frac{\epsilon}{2}\|v\|, \ \forall v \in T_yM, \tag{6.12}$$

and by (3.5), it is easily shown that for all $\epsilon > 0$, and for all $k \geq k_0$, we have for all $i = 1, \ldots, m$,

$$\|H(x_k)v(x_k) - \mathrm{Hess}G(x_k)v(x_k)\| < \frac{\epsilon}{2}\|v(x_k)\|. \tag{6.13}$$

Thus, using (6.12), (6.13) and Taylor's theorem, for k sufficiently large, we have

$$\begin{aligned}
&\|L(R_{x_k}v(x_k), x_k)\mathrm{grad}G(R_{x_k}v(x_k)) - (\mathrm{grad}G(x_k) + H(x_k)v(x_k))\| \\
&\leq \|L(R_{x_k}v(x_k), x_k)\mathrm{grad}G(R_{x_k}v(x_k)) - \mathrm{grad}G(x_k) - \mathrm{Hess}G(x_k)v(x_k)\| \\
&\quad + \|\mathrm{Hess}G(x_k)v(x_k) - H(x_k)v(x_k)\| \\
&= \int_0^1 [L(R_{x_k}tv(x_k), x_k)[\mathrm{Hess}G(R_{x_k}tv(x_k))DR_{x_k}tv(x_k)[v(x_k)]] \\
&\quad - \mathrm{Hess}G(x_k)v(x_k)]dt + \|\mathrm{Hess}G(x_k)v(x_k) - H(x_k)v(x_k)\| \\
&= \int_0^1 [L(R_{x_k}tv(x_k), x_k)[\mathrm{Hess}G(R_{x_k}tv(x_k))L(x_k, R_{x_k}tv(x_k))v(x_k)] \\
&\quad - \mathrm{Hess}G(x_k)v(x_k)]dt + \|\mathrm{Hess}G(x_k)v(x_k) - H(x_k)v(x_k)\| \\
&\leq \frac{\epsilon}{2}\|v(x_k)\| + \frac{\epsilon}{2}\|v(x_k)\| \\
&= \epsilon\|v(x_k)\|.
\end{aligned}$$

Since $q(x_{k+1}) = \mathrm{grad}G(x_{k+1})$, from (6.11) and (6.12), we have

$$\begin{aligned}
&\|q(x_{k+1})\| = \|\mathrm{grad}G(x_{k+1})\| \\
&= \|\mathrm{grad}G(R_{x_k}v(x_k))\| \\
&= \|L(R_{x_k}v(x_k), x_k)\mathrm{grad}G(R_{x_k}v(x_k))\| \\
&\leq \epsilon\|v(x_k)\|.
\end{aligned} \tag{6.14}$$

This, together with (6.9) imply that

$$|\tau(x_{k+1})| < \frac{\epsilon^2}{2a}\|v(x_k)\|^2.$$

By Lemma 6.1(a), we have

$$\frac{a}{2}\|v(x_{k+1})\|^2 < \frac{\epsilon^2}{2a}\|v(x_k)\|^2.$$

Thus, we obtain

$$\|v(x_{k+1})\| < \frac{\epsilon}{a}\|v(x_k)\|, \tag{6.15}$$

and then

$$\|v(x_{k+1})\| < (\epsilon/a)^{k+1-k_0}\|v(x_{k_0})\|. \tag{6.16}$$

For all $k > l \geq k_0$ and k_0 sufficiently large, from Lemma 6.3, let $\epsilon = 1$, we obtain

$$\begin{aligned} d(x_k, x_l) &\leq \sum_{i=l}^{k-1} d(x_{i+1}, x_i) \\ &= \sum_{i=l}^{k-1} d(R_{x_i}(v(x_i)), R_{x_i}0) \leq \sum_{i=l}^{k-1} 2\|v(x_i)\| \\ &\leq \sum_{i=l}^{k-1} 2(\epsilon/a)^{i-k_0}\|v(x_{k_0})\| \leq 2\|v(x_{k_0})\| \sum_{i=0}^{\infty} (\epsilon/a)^i = 2\frac{\|v(x_{k_0})\|}{1-\epsilon/a}. \end{aligned} \tag{6.17}$$

For k_0 sufficiently large, we obtain $\|v(x_{k_0})\| < \frac{\epsilon}{2(1-\epsilon/a)}$, and so

$$\frac{\|v(x_{k_0})\|}{1-\epsilon/a} < \frac{\epsilon(1-\epsilon/a)}{2(1-\epsilon/a)} = \frac{\epsilon}{2}.$$

Therefore, for all $\epsilon > 0$ and k_0 sufficiently large, $d(x_k, x_l) < \epsilon$. Hence, $\{x_k\}$ is a Cauchy sequence and there exists $x^* \in M$ such that $\lim_{k\to\infty} x_k = x^*$. Since $\tau(x)$ is continuous by Lemma 3.3, we have $\lim_{k\to\infty} \tau(x_k) = \tau(x^*)$. Hence, x^* is a Pareto critical point of F.

Now, we establish the superlinear convergence rate of $\{x_k\}$. Choose $t > 0$ and define $\tilde{\epsilon} = (at/1 + 2t)$, note that $\tilde{\epsilon}/a < 1$. By (6.17), for k_0 sufficiently large and $k > k_0$, $d(x_k, x_{k_0}) \leq \frac{2\|v(x_{k_0})\|}{1-\tilde{\epsilon}/a}$. Letting $k \to \infty$, we get $d(x^*, x_{k_0}) \leq \frac{2\|v(x_{k_0})\|}{1-\tilde{\epsilon}/a}$. Note that for all $k \geq k_0$, we have

$$d(x^*, x_{l+1}) \leq \frac{2\|v(x_{l+1})\|}{1-\tilde{\epsilon}/a} \leq \frac{2\tilde{\epsilon}/a}{1-\tilde{\epsilon}/a}\|v(x_l)\|. \tag{6.18}$$

By the triangular inequality and Lemma 6.3, let $\epsilon = \frac{1}{2}$, we have

$$\begin{aligned} d(x^*, x_l) &\geq d(x_{l+1}, x_l) - d(x^*, x_{l+1}) \\ &= d(R_{x_{l+1}}(v(x_l)), R_{x_l}(0)) - d(x^*, x_{l+1}) \\ &\geq \frac{1}{2}\|v(x_l)\| - \frac{2\tilde{\epsilon}/a}{1-\tilde{\epsilon}/a}\|v(x_l)\| \\ &= \frac{1-3\tilde{\epsilon}/a}{2(1-\tilde{\epsilon}/a)}\|v(x_l)\|. \end{aligned} \tag{6.19}$$

By (6.18) and (6.19), we have

$$d(x^*, x_{l+1}) \leq \frac{2\tilde{\epsilon}/a}{1-3\tilde{\epsilon}/a} d(x^*, x_l). \tag{6.20}$$

Then, the sequence $\{x_k\}$ converges to x^* superlinearly. □

Remark 6.5. If $M = \mathbb{R}^n$ and $R_x\eta = x + \eta$, then Theorem 6.4 can reduce to [9, Theorem 5.1] and to [22, Theorem 7] by using nonmonotone inexact search technique.

Next, we apply the nonmonotone quasi-Newton method to Stiefel manifolds and sphere S^{n-1}, we also obtain the local convergence rate of the proposed method.

Example 6.6. Let $\mathrm{St}(p,n)(p \le n)$ denote the set of all $n \times p$ orthonormal matrices, that is

$$\mathrm{St}(p,n)(p \le n) := \{X \in \mathbb{R}^{n\times p} \mid X^T X = I_p\},$$

where I_p denotes the $p \times p$ identity matrix. The set $\mathrm{St}(p,n)$ is called Stiefel manifold. The tangent space at x and the associated orthogonal projection are given by

$$T_X\mathrm{St}(p,n) = \{Z \in \mathbb{R}^{n\times p} \mid X^T Z + Z^T X = 0\},$$

$$P_X \xi_X = (I - XX^T)\xi_X + X\mathrm{skew}(X^T \xi_X),$$

where $\mathrm{skew}(A) = \frac{A-A^T}{2}$. The canonical product is given by

$$g(Z_1, Z_2) = \langle Z_1, Z_2\rangle = \mathrm{tr}(Z_1^T Z_2),\ \forall Z_1, Z_2 \in T_X\mathrm{St}(p,n).$$

From [1], we use the retraction given by

$$R_X(\eta_X) = \mathrm{qf}(X + \eta_X),$$

where $\mathrm{qf}(A)$ denotes the Q factor of decomposition of $A \in R_*^{n\times p}$ as $A = QR$, where $R_*^{n\times p}$ denotes the set of all nonsingular $n \times p$ matrixes, $Q \in \mathrm{St}(p,n)$, and R is an upper triangular $n \times p$ matrix with strictly positive diagonal elements. Moreover,

$$\mathcal{T}_{\eta_X}\xi_X = P_{R_X(\eta_X)}\xi_X = (I - Y^T Y)\xi_X + Y\mathrm{skew}(Y^T\xi_X),$$

where $Y := R_X(\eta_X)$.

Consider the multiobjective optimization $F : M \to \mathbb{R}^m$, $F(X) = (f_1(X), f_2(X), \ldots, f_m(X))$, where

$$f_i(X) = -\|\mathrm{diag}(X^T A_i X)\|^2,\ i = 1, 2, \ldots, m,$$

where $A_i = iI_{n\times n},\ i = 1, 2, \ldots, m$, and $\|\mathrm{diag}(X)\|^2$ returns the sum of the squared diagonal elements of X. This problem has applications in independent component analysis for blind source separation. From [26, Section 2], the gradient of f_i is

$$\mathrm{grad} f_i(X) = P_X \mathrm{grad}\bar{f}_i(X),\ i = 1, 2, \ldots, m,$$

where

$$\bar{f}_i : \mathbb{R}^{n\times p} \to \mathbb{R} : X \mapsto \bar{f}_i(X) = -\|\mathrm{diag}(X^T A_i X)\|^2,\ i = 1, 2, \ldots, m.$$

Moreover,

$$\mathrm{Hess} f_i(X)[\xi_X] = P_X(D\mathrm{grad}\bar{f}_i(X)[\xi_X]) - \xi_X \mathrm{sym}(X^T \mathrm{grad}\bar{f}_i(X)),\ i = 1, 2, \ldots, m,$$

where $\mathrm{sym}(U) = \frac{U+U^T}{2}$. By [1], if the columns of X are eigenvectors of A, we get X as a Pareto critical point of f. We choose $\mathrm{Hess} f_i(X) = B_i(X)$ for $i = 1, 2, \ldots, m$, $\eta_k = \frac{1}{2}$, we can check all the assumptions of Theorem 6.4 hold. Assume that x^* is a Pareto critical point F, using Theorem 6.4, the sequence $\{X_k\}$ generated by Algorithm 1 converges to x^*, superlinearly.

Example 6.7. On the unit sphere S^{n-1} considered as a Riemannian manifold of $\mathbb{R}^n$, the inner product inherited from $\mathbb{R}^n$ is given by

$$\langle \xi, \eta\rangle = \xi^T \eta,$$

and the projections are given by

$$P_x \xi_x = (I - xx^T)\xi_x.$$

From [1, Section 4], we obtain the retraction on S^{n-1} is

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

and

$$T_x(\eta_x) = \frac{x+\eta_x}{\|x+\eta_x\|}, \ \forall \eta_x \in T_xS^{n-1},$$

Moreover,

$$\mathcal{T}_{\eta_x}\xi_x = (I - \frac{(x+\eta_x)(x+\eta_x)^T}{\|x+\eta_x\|^2})\xi_x,$$

and

$$\mathcal{T}_{\eta_x}^{-1}y = (I - \frac{(x_k+\eta_x)x_k^T}{x_k^T(x_k+\eta_x)})y, \ \forall y \in T_{x_{k+1}}S^{n-1}.$$

Consider the multiobjective optimization $F : M \to \mathbb{R}^m, \ F(x) = (f_1(x), f_2(x), \ldots, f_m(x))$, where

$$f_i(x) = x^T A_i x, \ i = 1, 2, \ldots, m,$$

where $A_i = iI_{n\times n}, \ i = 1, 2, \ldots, m.$ By Section 4 in [1], we get

$$\mathrm{grad} f_i(x) = 2P_x(A_ix) = 2(A_ix - xx^TA_ix), \ i = 1, 2, \ldots, m,$$

where P_x is the orthogonal projector onto T_xS^{n-1}, *i.e.*,

$$P_xz = z - xx^Tz.$$

Moreover, the affine connection is given by

$$\nabla_\eta\xi = P_x(D\xi(x)\eta), \ \forall \xi \in S^{n-1}, \eta \in T_xS^{n-1},$$

and so

$$\mathrm{Hess} f_i(x)[\eta] = \nabla_\eta \mathrm{grad} f_i(x) = 2P_x(D\mathrm{grad} f_i(x)[\eta]) = 2P_x(A_i\eta - \eta x^TA_ix).$$

If we choose $\mathrm{Hess} f_i(x_k) = B_i(x_k)$ for $i = 1, 2, \ldots, m$, $\eta_k = \frac{1}{2}$, we can check all the assumptions of Theorem 6.4 hold. Assume that x^* is a Pareto critical point of F, using Theorem 6.4, the sequence $\{x_k\}$ generated by Algorithm 1 converges to x^*, superlinearly.

References

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

[2] G. C. Bento and J. X. da Cruz Neto, *A subgradient method for multiobjective optimization on Riemannian manifolds*, J. Optim. Theory Appl. **159** (2013), 125–137.

[3] G. C. Bento, O. P. Ferreira and P. R. Oliveira, *Unconstrained steepest descent method for multicriteria optimization on Riemmanian manifolds*, J. Optim. Theory Appl. **154** (2012), 88–107.

[4] D. Burago, Y. Burago and S. Ivanov, *A Course in Metric Geometry*, Graduate Studies in Math., Vol. 33, Amer. Math. Soc., Providence, RI, 2001.

[5] I. Chavel, *Riemannian Geometry - A Modern Introduction*, Cambridge University Press, London, 1993.

[6] J. X. da Cruz Neto and P. R. Oliveira, *Geodesic Methods in Riemannian Manifolds*, Universidade Federal do Rio de Janeiro, Research Report, PESC/COPPE, 1995.

[7] J. X. da Cruz Neto, L. L. Lima and P. R. Oliveira, *Geodesic Algorithm in Riemannian Manifolds*, Balk. J. Geom. Appl. **3** (1998), 89–100.

[8] G. Eichfrlder, *An adaptive scalarization method in multiobjective optimization*, SIAM J. Optim. **19** (2009), 1694–1718.

[9] J. Fliege, L. M. Graña Drummond and B. F. Svaiter, *Newton's method for multiobjective optimization*, SIAM J. Optim. **20** (2009), 602–626.

[10] E. H. Fukuda and L. M. Graña Drummond, *A survey on multiobjective descent methods*, Pesq. Oper. **34** (2014), 585–620.

[11] W. Huang, *Optimization Algorithms on Riemannian Manifolds with Applications*, Ph.D. Thesis, Department of Mathematics, Florida State University, Tallahassee, FL, 2013.

[12] W. Huang, K. A. Gallivan and P. A. Absil, *A Broyden class of quasi-newton methods for Riemannian optimization*, SIAM J. Optim. **25** (2015), 1660–1685.

[13] W. Klingenberg, *A Course in Differential Geometry*, Springer, Berlin, 1978.

[14] 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. `https://doi.org/10.23952/asvao.5.2023.2.12`

[15] X. B. Li, N. J. Huang, Q. H. Ansari and J. C. Yao, *Convergence rate of descent method with new inexact line-search on Riemannian manifolds*, J. Optim. Theory Appl. **180** (2019), 830–854.

[16] X. B. Li and M. Krishan Lal, *Convergence Analysis of a Nonmonotone Projected Gradient Method for Multiobjective Optimization Problems on Riemannian Manifolds*, J. Nonlinear Convex Anal. **23(8)** (2022), 1673-1691.

[17] X. B. Li, X. Wang and M. Krishan Lal, *A Nonmonotone Trust Region Method for Unconstrained Optimization Problems on Riemannian Manifolds*, J. Optim. Theory Appl. **188** (2021), 547–570.

[18] N. Mahdavi-Amiri and F. S. Sadaghiani, *A Superlinearly convergent nonmonotone quasi-Newton method for unconstrained multiobjective optimization*, Optim. Method Softw. **35** (2020), 1223–1247.

[19] K. Miettinen, *Nonlinear multiobjective optimization*, Vol. 12, Springer Science Business Media, 2012.

[20] K. Mita, E. H. Fukuda and N. Yamashita, *Nonmonotone line searches for unconstrained multiobjective optimization problems*, J. Global Optim. **75** (2019), 63–90.

[21] P. M. Pardalos and A. Žilinskas, *Non-convex multi-objective optimization*, Springer International Publishing, 2017.

[22] Ž. Povalej, *Quasi-Newton's method for multiobjective optimization*, J. Comput. Appl. Math. **255**(2014), 765–777.

[23] L. F. Prudente and D. R. Souza, *A Quasi-Newton method with Wolfe line searches for multiobjective optimization*, J. Optim. Theory Appl., **194(3)**, 1107–1140.

[24] W. Ring and B. Wirth, *Optimization methods on Riemannian manifolds and their application to shape space*, SIAM J. Optim. **22** (2012), 596–627.

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

[26] F. J. Theis, T. P. Cason and P.-A. Absil, *Soft Dimension Reduction for ICA by Joint Diagonalization on the Stiefel Manifold*, In: Proceedings of the 8th International Conference on Independent Analysis and Signal Separation, **5441** (2009), 351–361.

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

Manuscript received April 7 2023
revised September 1 2023

X. B. Li
School of Sciences, Civil Aviation Flight University of China, Guanghan, 618300, People Republic of China.
E-mail address: lixiaobo.double@163.com

M. Krishan Lal
Mathematics, Irving K. Barber Faculty of Science, UBC Okanagan, Kelowna, British Columbia V1V 1V7, Canada.
E-mail address: manish.krishanlal@ubc.ca