CrossMark

Convergence Rate of Descent Method with New Inexact Line-Search on Riemannian Manifolds

Xiao-bo Li[1] · Nan-jing Huang[2] · Qamrul Hasan Ansari[3,4] · Jen-Chih Yao[5]

Abstract

In this paper, we propose the descent method with new inexact line-search for unconstrained optimization problems on Riemannian manifolds. The global convergence of the proposed method is established under some appropriate assumptions. We further analyze some convergence rates, namely R-linear convergence rate, superlinear convergence rate and quadratic convergence rate, of the proposed descent method.

✉ Qamrul Hasan Ansari
  qhansari@gmail.com

  Xiao-bo Li
  lixiaobo.double@163.com

  Nan-jing Huang
  nanjinghuang@hotmail.com

  Jen-Chih Yao
  yaojc@mail.cmu.edu.tw

[1] School of Sciences, Southwest Petroleum University, Chengdu 610500, People's Republic of China

[2] Department of Mathematics, Sichuan University, Chengdu 610064, People's Republic of China

[3] Department of Mathematics, Aligarh Muslim University, Aligarh 202 002, India

[4] Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

[5] Research Center for Interneural Computing, China Medical University Hospital, China Medical University, Taichung, Taiwan

⌀ Springer

1 Introduction

Since the non-convex problems can be translated into convex problems with appropriate Riemannian metrics, and constrained optimization problems can be written as unconstrained optimization problems from the Riemannian geometry viewpoint, in recent years, many concepts, techniques, as well as methods of optimization theory have been extended from Euclidean space to Riemannian manifolds; see, for instance, [1–15]. It is an emerging area of research in convex analysis, optimization theory, game theory, etc; see, for example, [12,14] and the references therein.

The descent method is one of the oldest and widely known methods for solving minimization problems. Many convergence results for descent method have been established in the setting of Euclidean space. For instance, some line-search techniques for the general descent method in $\mathbb{R}^n$ are proposed by Nocedal and Wright [16]. They proved the global convergence and studied convergence rate of steepest descent method, Newton's method and coordinate descent method. Shi and Shen [17] also proved the global convergence and local convergence rate of descent method with new inexact line-search in $\mathbb{R}^n$. The new inexact line-search in [17] has many advantages comparing with some other similar line-searches, such as Armijo line-search and Wolfe line-search; see [18,19]. Based on the results in [17], a new inexact line-search for quasi-Newton method is proposed and some global convergence results are established by Shi [20]. Recently, some authors focused on extending the convergence results for descent methods from Euclidean space to Riemannian manifolds. In particular, the steepest descent method and the Newton method to solve convex optimization problems on Riemannian manifolds have been studied by Udrişte [14], in which the linear convergence result for the steepest descent method is obtained under the assumption of exact line-search. Smith [21] considered the steepest descent, Newton method and conjugate gradient method on Riemannian manifolds and analyzed their convergence properties by employing the Riemannian structure of the manifold. The full convergence result for the steepest descent method with the Armijo line-search on Riemannian manifolds has been studied by da Cruz Neto et al. [4,5]. A combination of generalized Armijo method and generalized Newton method in the setting of Riemannian manifolds is investigated by Yang [15]. The global convergence results for these methods with less restrictions are obtained. Estimations of linear/quadratic convergence rate for the generalized Armijo–Newton algorithms are also derived in [15]. Moreover, the global and local convergence analysis of the exact and approximate line-search such as the Newton's method, the gradient descent method and the trust-region method on Riemannian manifolds is studied in [22–27].

Motivated by the results described above, in this paper, we propose a new inexact line-search for the descent method on Riemannian manifolds and establish its convergence results. Since the computation of the exponential mapping and parallel transport can be quite expensive in the setting of manifolds, and many convergence results show that the nice properties of some algorithms hold for all suitably defined retractions and isometric vector transports on Riemannian manifolds (see, for example, [6,13,28]), we replace the geodesic by retraction and parallel transport by isometric vector transport, respectively. As a result, the global convergence result of the descent method, with new inexact line-search for unconstrained optimization problems on Riemannian

manifolds, is proved under some appropriate assumptions. Moreover, we analyze the convergence rate, such as R-linear convergence rate, superlinear convergence rate and quadratic convergence rate, of the descent method on Riemannian manifolds. To the best of our knowledge, these results have not been studied before, and it is very interesting to explore the convergence results for a new inexact line-search algorithm on manifolds by utilizing Riemannian techniques.

The present paper is organized as follows: In Sect. 2, we recall some notions and known results from Riemannian geometry, which will be used throughout the paper. We propose the descent method with new inexact line-search for unconstrained optimization problems on Riemannian manifolds. In Sect. 3, the global convergence result for the descent method on Riemannian manifolds is proved. Section 4 is devoted to analyze the convergence rate of the descent method with new inexact line-search.

2 Preliminaries

In this section, we recall some standard notations, definitions and results from Riemannian manifolds, which can be found in any introductory book on Riemannian geometry; see, for example, [2,3,7].

Let M be a finite-dimensional differentiable manifold and $x \in M$. The tangent space of M at x is denoted by $T_x M$ and the tangent bundle of M by $TM = \bigcup_{x \in M} T_x M$, which is naturally a manifold. We denote by $g_x(.,.)$ the inner product on $T_x M$ 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. Throughout the paper, unless otherwise specified, we assume that 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 $l(\gamma) = \int_a^b \|\gamma'(t)\| \mathrm{d}t$. 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.

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. Let $\gamma : \mathbb{R} \to M$ be a geodesic and $P_{\gamma[.,.]}$ denote the parallel transport along γ, which is defined by $P_{\gamma[\gamma(a), \gamma(b)]}(v) = V(\gamma(b))$ for all $a, b \in \mathbb{R}$ and $v \in T_{\gamma(a)} M$, where V is the unique C^∞ vector field such that $\nabla_{\gamma'(t)} V = 0$ 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 [29], any pair of points $x, y \in M$ can be joined by a minimal geodesic. The exponential mapping $\exp_x : T_x M \to M$ is defined by $\exp_x v = \gamma_v(1, x)$ for each $v \in T_x M$, where $\gamma(\cdot) = \gamma_v(\cdot, x)$ is the geodesic starting from x with velocity v, that is, $\gamma(0) = x$ 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 $T_x M$. However, the systematic use of the exponential mapping may not be desirable in all cases. Some local mappings to $T_x M$ may reduce the computational cost while preserving the useful convergence properties of the considered method. In this paper,

we relax the exponential mapping by a class of mappings called retractions, a concept that can be seen in [1,6,13,30,31].

Definition 2.1 Given $x \in M$, a retraction is a smooth mapping $R_x : T_x M \to M$ such that

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

As we know, the exponential mapping is a special retraction, and some retractions can be seen as an approximation of the exponential mapping.

The parallel transport is often too expensive to use in a practical method, so we can consider a more general vector transport (see, for example, [6,13,22]), which is built upon the retraction R_x. 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 $\xi_x, \zeta_x \in T_x M$, (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 $\mathscr{T}_S$ denote the isometric vector transport (see, for example, [13,22,31]) with R_x as the associated retraction. Then it satisfies (i), (ii), (iii) and the following condition (iv):

$$g(\mathscr{T}_{S(\eta_x)} \xi_x, \mathscr{T}_{S(\eta_x)} \zeta_x) = g(\xi_x, \zeta_x). \tag{1}$$

In most of the practical cases, $\mathscr{T}_{S(\eta_x)}$ exists for all $\eta_x \in T_x M$, and thus, we make this assumption throughout the paper. Furthermore, let $\mathscr{T}_{R_x}$ denote the derivative of the retraction, that is,

$$\mathscr{T}_{R_x(\eta_x)} \xi_x = DR_x(\eta_x)[\xi_x] = \left. \frac{\mathrm{d}}{\mathrm{d}t} R_x(\eta_x + t\xi_x) \right|_{t=0}. \tag{2}$$

Let $L(TM, TM)$ denote the fiber bundle with the base space $M \times M$ such that the fiber bundle over $(x, y) \in M \times M$ is $L(T_x M, T_y M)$, the set of all linear mappings from $T_x M$ to $T_y M$. We recall, from Section 4 in [6], that a transporter $\mathscr{L}$ on M to be a smooth section of the bundle $L(TM, TM)$, that is, for all $x, y \in M$, $\mathscr{L}(x, y) \in L(T_x M, T_y M)$ and $\mathscr{L}(x, x) = \mathrm{id}$, where id means the identity mapping. Furthermore, $\mathscr{L}(x, y)$ is isometric from $T_x M$ to $T_y M$, $\mathscr{L}^{-1}(x, y) = \mathscr{L}(y, x)$ and $\mathscr{L}(x, z) = \mathscr{L}(y, z)\mathscr{L}(x, y)$. Given a retraction R_x, for any $\eta_x, \xi_x \in T_x M$, the isometric vector transport $\mathscr{T}_S$ can be defined by

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

In this paper, we require

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

In some manifolds, there exist retractions such that the above equality holds, e.g., the Stiefel manifold and the Grassmann manifold [31]. Furthermore, from the above equality and (1), we have

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

In this paper, we consider the following unconstrained minimization problem

$$\min_{x \in M} f(x), \tag{3}$$

where $f : M \to \mathbb{R}$ is a continuously differentiable function with a lower bound and M is a complete Riemannian manifold.

Now, we describe the descent method with new inexact line-search on Riemannian manifolds for finding the solution of the minimization problem (3).

New Inexact Line-Search on Riemannian Manifolds

Given $\beta \in]0, 1[$ and $\sigma \in]0, 1[$, the Hessian approximation B_k is symmetric positive definite with respect to the metric g, and $s_k = \frac{-g(\operatorname{grad} f(x_k), d_k)}{g(B_k[d_k], d_k)}$, where $d_k \in T_{x_k}M$ is a descent direction with $g(\operatorname{grad} f(x_k), d_k) < 0$. The step size α_k is the largest one in $\{s_k, s_k\beta, s_k\beta^2, \ldots\}$ such that

$$f(R_{x_k}(\alpha_k d_k)) - f(x_k) \le \sigma\alpha_k \left(g(\operatorname{grad} f(x_k), d_k) + \frac{1}{2}\alpha_k g(B_k[d_k], d_k) \right). \tag{4}$$

In comparison with the original Armijo line-search on Riemannian manifolds [4,5], there are some advantages of the new inexact line-search. For example, let α_k denote the step size defined by the original Armijo line-search and α'_k denote the step size defined by the new inexact line-search; then, it is easy to see that $\alpha_k < \alpha'_k$.

Now we present a related descent method with the new inexact line-search on Riemannian manifolds for the minimization problem (3) as follows.

Algorithm 2.1 (Descent Method with New Inexact Line-search)

Step 1. *Given $x_0 \in M$, initial Hessian approximation B_0, which is symmetric positive definite with respect to the metric g, $k := 0$;*
Step 2. *If $\|\operatorname{grad} f(x_k)\| = 0$, then stop. Else go to step 3;*
Step 3. *Set $x_{k+1} = R_{x_k}(\alpha_k d_k)$, where d_k is a descent direction with $g(\operatorname{grad} f(x_k), d_k) < 0$, α_k is defined by the new inexact line-search on Riemannian manifolds;*
Step 4. *Define $p_k = \operatorname{grad} f(x_{k+1}) - \mathscr{L}(x_k, x_{k+1})\operatorname{grad} f(x_k)$ and modify B_k as B_{k+1} by using BFGS quasi-Newton algorithm or other quasi-Newton algorithms (see, for example, [6,13]);*
Step 5. *Set $k := k + 1$ and go to step 2.*

Remark 2.1 If M is a Riemannian manifold with nonnegative curvature, then by Proposition 5.4.1 in [22], the exponential mapping on M induced by ∇ is a retraction R, and

the parallel transport is isometric vector transport. Furthermore, if $d_k = -\text{grad} f(x_k)$ and $B_k = I$ (I denotes the unit matrix), then from (4), we get

$$f(x_{k+1}) - f(x_k) \leq \sigma \alpha_k \left(\frac{1}{2} \alpha_k - 1 \right) \|\text{grad} f(x_k)\|^2.$$

In this case, the descent method with new inexact line-search on Riemannian manifolds is a generalization of the steepest descent method with Armijo line-search in [4].

From Lemma 5.2.1 and Lemma 5.2.2 in [31], and the definition of $\mathscr{L}(.,.)$, we obtain the following result.

Lemma 2.1 *Suppose that the function $f : M \to \mathbb{R}$ is twice continuously differentiable on a set $S \subset M$. Then, there exists a constant $L_0 > 0$ such that, for all $x, y \in S$,*

$$\|\mathscr{L}(x, y)\text{grad} f(x) - \text{grad} f(y)\| \leq L_0 \|R_x^{-1} y\|. \tag{5}$$

3 Global Convergence

In this section, we study the global convergence property of the descent method with new inexact line-search on Riemannian manifolds under some appropriate assumptions.

Definition 3.1 The matrix B_k is said to be uniformly positive definite, iff there exist constants m' and $\overline{m}$ such that $0 < m' \leq \overline{m}$ and for any $k \in \mathbb{N}$,

$$m' \|d\|^2 \leq g(B_k[d], d) \leq \overline{m} \|d\|^2, \quad \forall d \in T_{x_k} M. \tag{6}$$

Proposition 3.1 *Assume that f is twice continuously differentiable on M, B_k is uniformly positive definite, d_k is a descent direction, and $\{x_k\}$ is the sequence generated by Algorithm 2.1. Then, there exists $\tau > 0$ such that*

$$f(x_k) - f(x_{k+1}) \geq \tau \left(\frac{g(\text{grad} f(x_k), d_k)}{\|d_k\|} \right)^2, \quad \forall k \in \mathbb{N}. \tag{7}$$

Proof Let $K_1 = \{k \in \mathbb{N} : \alpha_k = s_k\}$ and $K_2 = \{k \in \mathbb{N} : \alpha_k < s_k\}$. The proof is divided into two parts.

PART 1. If $k \in K_1$, then by (4) and (6), we have

$$f(x_k) - f(x_{k+1}) \geq -\alpha_k \sigma \left(g(\text{grad} f(x_k), d_k) + \frac{1}{2} \alpha_k g(B_k[d_k], \alpha_k) \right)$$

$$= \frac{\sigma g(\text{grad} f(x_k), d_k)}{g(B_k[d_k], \alpha_k)} (g(\text{grad} f(x_k), d_k)$$

$$-\frac{1}{2}\frac{g(\operatorname{grad} f(x_k), d_k)}{g(B_k[d_k], \alpha_k)}g(B_k[d_k], \alpha_k))$$

$$\geq \frac{\sigma}{2m}\left(\frac{g(\operatorname{grad} f(x_k), d_k)}{\|d_k\|}\right)^2, \quad \forall k \in K_1.$$

Thus,

$$f(x_k) - f(x_{k+1}) \geq \frac{\sigma}{2m}\left(\frac{g(\operatorname{grad} f(x_k), d_k)}{\|d_k\|}\right)^2, \quad \forall k \in K_1. \qquad (8)$$

PART 2. If $k \in K_2$, then $\alpha_k \beta^{-1} \in \{s_k, s_k \beta, s_k \beta^2, \ldots\}$. This implies that (4) does not hold, and so,

$$f(x_k) - f(R_{x_k}(\alpha_k \beta^{-1} d_k))$$
$$< -\sigma \alpha_k \beta^{-1}(g(\operatorname{grad} f(x_k), d_k) + \frac{1}{2}\alpha_k \beta^{-1}g(B_k[d_k], d_k)), \quad \forall k \in K_2. \quad (9)$$

Define $m(t) = f(R_{x_k}(td_k))$. By using mean value theorem on the left-hand side of the above inequality, there exists $\theta_k \in]0, 1[$ such that

$$\begin{aligned}
&f(x_k) - f(R_{x_k}(\alpha_k \beta^{-1} d_k)) \\
&= m(0) - m(\alpha_k \beta^{-1}) \\
&= \frac{dm(\theta_k \alpha_k \beta^{-1})}{dt}(0 - \alpha_k \beta^{-1}) \\
&= -\alpha_k \beta^{-1}g\left(\operatorname{grad} f(R_{x_k}(\theta_k \alpha_k \beta^{-1} d_k)), DR_{x_k}(\theta_k \alpha_k \beta^{-1} d_k)[d_k]\right) \\
&= -\alpha_k \beta^{-1}g\left(\operatorname{grad} f(R_{x_k}(\theta_k \alpha_k \beta^{-1} d_k)), \mathscr{L}(x_k, R_{x_k}(\theta_k \alpha_k \beta^{-1} d_k))d_k\right) \\
&< -\sigma \alpha_k \beta^{-1}\left(g(\operatorname{grad} f(x_k), d_k) + \frac{1}{2}\alpha_k \beta^{-1}g(B_k[d_k], d_k)\right), \quad \forall k \in K_2.
\end{aligned}$$

This shows that

$$g\left(\operatorname{grad} f(R_{x_k}(\theta_k \alpha_k \beta^{-1} d_k)), \mathscr{L}(x_k, R_{x_k}(\theta_k \alpha_k \beta^{-1} d_k))d_k\right)$$
$$> \sigma\left(g(\operatorname{grad} f(x_k), d_k) + \frac{1}{2}\alpha_k \beta^{-1}g(B_k[d_k], d_k)\right). \qquad (10)$$

Noting that

$$\|\mathscr{L}(x_k, R_{x_k}(\theta_k \alpha_k \beta^{-1} d_k))d_k\| = \|d_k\|,$$

it follows from Lemma 2.1 that there exists $L_0 > 0$ such that

$$\begin{aligned}
&g\big(\operatorname{grad} f(R_{x_k}(\theta_k \alpha_k \beta^{-1} d_k)) \\
&\quad - \mathscr{L}(x_k, R_{x_k}(\theta_k \alpha_k \beta^{-1} d_k))\operatorname{grad} f(x_k), \mathscr{L}(x_k, R_{x_k}(\theta_k \alpha_k \beta^{-1} d_k))d_k\big) \\
&\quad \leq \|\operatorname{grad} f(R_{x_k}(\theta_k \alpha_k \beta^{-1} d_k)) - \mathscr{L}(x_k, R_{x_k}(\theta_k \alpha_k \beta^{-1} d_k))\operatorname{grad} f(x_k)\|\|d_k\|
\end{aligned}$$

$$\leq L_0 \| R_{x_k}^{-1} R_{x_k}(\theta_k \alpha_k \beta^{-1} d_k) \| \| d_k \|$$
$$\leq L_0 \alpha_k \beta^{-1} \| d_k \|^2.$$

This together with (10) implies that

$$L_0 \alpha_k \beta^{-1} \| d_k \|^2$$
$$\geq g\big(\mathrm{grad}\, f(R_{x_k}(\theta_k \alpha_k \beta^{-1} d_k))$$
$$\quad - \mathscr{L}(x_k, R_{x_k}(\theta_k \alpha_k \beta^{-1} d_k)) \mathrm{grad}\, f(x_k), \mathscr{L}(x_k, R_{x_k}(\theta_k \alpha_k \beta^{-1} d_k)) d_k\big)$$
$$> \sigma(g(\mathrm{grad}\, f(x_k), d_k) + \frac{1}{2}\alpha_k \beta^{-1} g(B_k[d_k], d_k)) - g(\mathrm{grad}\, f(x_k), d_k)$$
$$> (\sigma - 1) g(\mathrm{grad}\, f(x_k), d_k), \quad \forall k \in K_2.$$

Thus, we obtain

$$\alpha_k > \frac{\beta(\sigma - 1)}{L_0} \cdot \frac{g(\mathrm{grad}\, f(x_k), d_k)}{\| d_k \|^2}, \quad \forall k \in K_2.$$

Let

$$\alpha_k' = \frac{\beta(\sigma - 1)}{L_0} \cdot \frac{g(\mathrm{grad}\, f(x_k), d_k)}{\| d_k \|^2}.$$

Then,

$$\alpha_k' < \alpha_k < s_k. \tag{11}$$

Now, from (4) and (11), one has

$$f(x_k) - f(x_{k+1}) \geq -\sigma \alpha_k \left(g(\mathrm{grad}\, f(x_k), d_k) + \frac{1}{2}\alpha_k g(B_k[d_k], d_k) \right)$$
$$\geq -\sigma \max_{\alpha_k' \leq \alpha \leq s_k} \left\{ \alpha \left(g(\mathrm{grad}\, f(x_k), d_k) + \frac{1}{2}\alpha g(B_k[d_k], d_k) \right) \right\}$$
$$= -\sigma \alpha_k' \left(g(\mathrm{grad}\, f(x_k), d_k) + \frac{1}{2}\alpha_k g(B_k[d_k], d_k) \right)$$
$$\geq -\sigma \alpha_k' g(\mathrm{grad}\, f(x_k), d_k) - \frac{\sigma}{2}\alpha_k' s_k g(B_k[d_k], d_k)$$
$$= \frac{\beta \sigma (1 - \sigma)}{2 L_0} \left(\frac{g(\mathrm{grad}\, f(x_k), d_k)}{\| d_k \|} \right)^2, \quad \forall k \in K_2.$$

Thus,

$$f(x_k) - f(x_{k+1}) \geq \frac{\beta \sigma (1 - \sigma)}{2 L_0} \left(\frac{g(\mathrm{grad}\, f(x_k), d_k)}{\| d_k \|} \right)^2, \quad \forall k \in K_2. \tag{12}$$

Let

$$\tau = \min \left\{ \frac{\sigma}{2\overline{m}}, \frac{\beta \sigma (1 - \sigma)}{2 L_0} \right\}.$$

By (8) and (12), we obtain

$$f(x_k) - f(x_{k+1}) \geq \tau \left(\frac{g(\operatorname{grad} f(x_k), d_k)}{\|d_k\|} \right)^2, \quad \forall k \in \mathbb{N}.$$

This completes the proof.   □

Corollary 3.1 *Suppose that all the assumptions of Proposition* 3.1 *are satisfied and* d_k *satisfies*

$$\cos \theta_k = \frac{-g(\operatorname{grad} f(x_k), d_k)}{\|\operatorname{grad} f(x_k)\| \|d_k\|} \geq \mu, \tag{13}$$

where $0 < \mu \leq 1$ *is a constant and* θ_k *denotes the angle between* $-\operatorname{grad} f(x_k)$ *and* d_k. *Then,*

$$f(x_k) - f(x_{k+1}) \geq \tau \mu^2 \|\operatorname{grad} f(x_k)\|^2, \quad \forall k \in \mathbb{N}.$$

Proof It follows from (13) that

$$g^2(\operatorname{grad} f(x_k), d_k) \geq \mu^2 \|\operatorname{grad} f(x_k)\|^2 \|d_k\|^2.$$

This together with (7) implies that

$$f(x_k) - f(x_{k+1}) \geq \tau \mu^2 \|\operatorname{grad} f(x_k)\|^2, \quad \forall k \in \mathbb{N}.$$

This completes the proof.   □

Theorem 3.1 *Suppose that all the assumptions of Corollary* 3.1 *are satisfied. Then,*

$$\sum_{k=1}^{+\infty} \|\operatorname{grad} f(x_k)\|^2 < +\infty,$$

and thus,

$$\lim_{k \to +\infty} \|\operatorname{grad} f(x_k)\| = 0. \tag{14}$$

Proof Since f has a lower bound on M, by Corollary 3.1, the result follows.   □

Remark 3.1 Theorem 3.1 shows that all accumulation points of the sequence $\{x_k\}$ generated by Algorithm 2.1 are stationary points. Moreover, if we consider $M = \mathbb{R}^n$ and $R_x(\eta) = x + \eta$, then Theorem 3.1 reduces to Corollary 3.4 in [17].

4 Convergence Rate

In this section, we analyze the convergence rate, such as R-linear convergence rate, superlinear convergence rate and quadratic convergence rate, of the descent method with new inexact line-search on Riemannian manifolds.

Definition 4.1 Let $\{x_k\}$ be a sequence converging to x^*. The convergence is said to be

(a) *R*-linear, iff there exist a constant $0 \leq \theta < 1$ and a positive integer N such that

$$[d(x_k, x^*)]^{\frac{1}{k}} \leq \theta, \quad \forall k > N;$$

(b) superlinear, iff there exist a sequence $\{\alpha_k\}$ converging to 0 and a positive integer N such that
$$d(x_{k+1}, x^*) \leq \alpha_k d(x_k, x^*), \quad \forall k > N;$$

(c) quadratic, iff there exist a constant $\theta \geq 0$ and a positive integer N such that

$$d(x_{k+1}, x^*) \leq \theta d^2(x_k, x^*), \quad \forall k > N.$$

The convergence rate depends on the property of uniformly retraction-convexity, which is defined as follows.

Definition 4.2 For a function $f : M \to \mathbb{R}$ on the Riemannian manifold M with the retraction R, define the function $m_{x,\eta}(t) = f(R_x(t\eta))$ for $x \in M$ and $\eta \in T_x M$. The function f is said to be uniformly retraction-convex on $S \subset M$, iff $m_{x,\eta}(t)$ is uniformly convex, i.e., there exists a constant $c > 0$ such that

$$\alpha m_{x,\eta}(t_1) + (1-\alpha)m_{x,\eta}(t_2) - m_{x,\eta}(t_2 + \alpha(t_1 - t_2)) \geq c\alpha(1-\alpha)(t_1 - t_2)^2,$$

for all $x \in S$, $\eta \in T_x M$ with $\|\eta\| = 1$, $\alpha \in {]0, 1[}$ and $t_1, t_2 \geq 0$ such that $R_x(\varsigma\eta) \in S$ for all $\varsigma \in [0, \max(t_1, t_2)]$.

Remark 4.1 From Section 3.4 in [32], if there exists a constant $c > 0$ such that $\frac{d^2 m_{x,\eta}(t)}{dt^2} \geq c$, then $m_{x,\eta}(t)$ is uniformly convex, and so f is uniformly retraction-convex. Next, we provide a sufficient condition to ensure the uniformly retraction-convexity of f. Assume that x^* is a stationary point of f and the Hessian matrix of f at x^*, denoted by $\mathrm{Hess} f(x^*)$, is positive definite. Then, from Lemma 3.1 in [6], there exist $c_1, c_2 > 0$, $\bar{t} > 0$ and a neighborhood N of x^* such that $c_1 \leq \frac{d^2 m_{x,\eta}(t)}{dt^2} \leq c_2$ for all $x \in N$ and $t \leq \bar{t}$. Thus, f is uniformly retraction-convex on a neighborhood of x^*.

Lemma 4.1 *Assume that f is twice continuously differentiable and uniformly retraction-convex on M, and x^* is a stationary point of f. Moreover, assume that*

$$\frac{\mathrm{D}}{\mathrm{d}t}\mathrm{D}R_x(t\eta)[\eta] = 0, \quad \forall x \in M, \eta \in T_x M, \tag{15}$$

where $\mathrm{D}/\mathrm{d}t$ denotes the covariant derivative along the curve $t \mapsto R_x(t\eta)$ (see Chapter 2 in [33]). Then, there exist constants m' and $\overline{m}$ such that $0 < m' \leq \overline{m}$ and

$$m'\|\eta\|^2 \leq g(\mathrm{Hess} f(x)[\eta], \eta) \leq \overline{m}\|\eta\|^2, \quad \forall x \in M, \eta \in T_x M; \tag{16}$$

and thus,

$$\frac{1}{2}m'\|R_{x^*}^{-1}x\|^2 \le f(x) - f(x^*) \le \frac{1}{2}\overline{m}\|R_{x^*}^{-1}x\|^2, \quad \forall x \in M; \tag{17}$$

$$m'\|R_x^{-1}y\|^2 \le g(\mathcal{L}(y,x)\mathrm{grad}\,f(y) - \mathrm{grad}\,f(x), R_x^{-1}y) \le \overline{m}\|R_x^{-1}y\|^2, \quad \forall x, y \in M; \tag{18}$$

$$\|\mathrm{grad}\,f(x)\| \ge m'\|R_{x^*}^{-1}x\|, \quad \forall x \in M. \tag{19}$$

Proof Define $m_{x,\eta}(t) = f(R_x(t\eta))$, for all $x \in M$, $\eta \in T_x M$ with $\|\eta\| = 1$ and $t \ge 0$. Let $y = R_x(t\eta)$. Then, it is easy to see that $t\eta = R_x^{-1}y$. Since f is uniformly retraction-convex on M, there exists a constant $c > 0$ such that

$$\alpha m_{x,\eta}(0) + (1-\alpha)m_{x,\eta}(t) - m_{x,\eta}(t + \alpha(0-t)) \ge c\alpha(1-\alpha)t^2, \quad \forall \alpha \in]0, 1[.$$

This implies that

$$
\begin{aligned}
m_{x,\eta}(0) - m_{x,\eta}(t) &\ge \lim_{\alpha \to 0}\left\{\frac{m_{x,\eta}(t + \alpha(0-t)) - m_{x,\eta}(t)}{\alpha} + c(1-\alpha)t^2\right\} \\
&= g(\mathrm{grad}\,f(R_x(t\eta)), (-t)\mathrm{D}R_x(t\eta)[\eta]) + ct^2 \\
&= g(\mathrm{grad}\,f(R_x(t\eta)), \mathcal{L}(x, R_x(t\eta))(-t\eta)) + ct^2. \tag{20}
\end{aligned}
$$

By the similar argument, we also have

$$m_{x,\eta}(t) - m_{x,\eta}(0) \ge g(\mathrm{grad}\,f(x), t\eta) + ct^2. \tag{21}$$

Combining (20) and (21), we get

$$g(\mathrm{grad}\,f(R_x(t\eta)), -\mathcal{L}(x, R_x(t\eta))t\eta) + g(\mathrm{grad}\,f(x), t\eta) + 2ct^2 \le 0.$$

Hence,

$$g(\mathcal{L}(y,x)\mathrm{grad}\,f(y) - \mathrm{grad}\,f(x), \eta) \ge 2ct\|\eta\|^2. \tag{22}$$

Clearly,

$$
\begin{aligned}
\frac{\mathrm{d}m_{x,\eta}(t)}{\mathrm{d}t} &= g(\mathrm{grad}\,f(R_x(t\eta)), \mathrm{D}R_x(t\eta)[\eta]) \\
&= g(\mathrm{grad}\,f(y), \mathcal{L}(x, y)\eta) \\
&= g(\mathcal{L}(y,x)\mathrm{grad}\,f(y), \eta), \tag{23}
\end{aligned}
$$

and

$$\frac{\mathrm{d}m_{x,\eta}(0)}{\mathrm{d}t} = g(\mathrm{grad}\,f(x), \eta). \tag{24}$$

Then, it follows from (23) that

$$
\frac{\mathrm{d}^2 m_{x,\eta}(t)}{\mathrm{d}t^2} = g\left(\mathrm{Hess}\, f(R_x(t\eta))[\mathrm{D}R_x(t\eta)[\eta]], \mathrm{D}R_x(t\eta)[\eta]\right)
$$
$$
+ g\left(\mathrm{grad}\, f(R_x(t\eta)), \frac{\mathrm{D}}{\mathrm{d}t}\mathrm{D}R_x(t\eta)[\eta]\right),
$$

Since $\frac{\mathrm{D}}{\mathrm{d}t}\mathrm{D}R_x(t\eta)[\eta] = 0$, we have

$$
\frac{\mathrm{d}^2 m_{x,\eta}(t)}{\mathrm{d}t^2} = g(\mathrm{Hess}\, f(R_x(t\eta))[\mathrm{D}R_x(t\eta)[\eta]], \mathrm{D}R_x(t\eta)[\eta]). \tag{25}
$$

Thus, by (22), (23), (24) and (25), we have

$$
\begin{aligned}
g(\mathrm{Hess}\, f(x)[\eta], \eta) &= \frac{\mathrm{d}^2 m_{x,\eta}(0)}{\mathrm{d}t^2} \\
&= \lim_{t\to 0} \frac{g(\mathscr{L}(y,x)\mathrm{grad}\, f(y), \eta) - g(\mathrm{grad}\, f(x), \eta)}{t} \\
&= \lim_{t\to 0} \frac{g(\mathscr{L}(y,x)\mathrm{grad}\, f(y) - \mathrm{grad}\, f(x), \eta)}{t} \ge 2c\|\eta\|^2. \quad (26)
\end{aligned}
$$

Since f is twice continuously differentiable on M, from Lemma 2.1, for $t > 0$, we obtain

$$
\begin{aligned}
\frac{g(\mathscr{L}(y,x)\mathrm{grad}\, f(y) - \mathrm{grad}\, f(x), \eta)}{t} &\le \frac{\|\mathscr{L}(y,x)\mathrm{grad}\, f(y) - \mathrm{grad}\, f(x)\|\|\eta\|}{t} \\
&\le \frac{L_0\|t\eta\|\|\eta\|}{t} = L_0\|\eta\|^2. \quad (27)
\end{aligned}
$$

Taking $2c = m'$ and $L_0 = \overline{m}$, then, it follows from (26) and (27) that

$$
m'\|\eta\|^2 \le g(\mathrm{Hess}\, f(x)[\eta], \eta) \le \overline{m}\|\eta\|^2, \quad \forall x \in M, \eta \in T_x M. \tag{28}
$$

Thus, (16) holds.

For any $x \in M$ and $x \ne x^*$, let $z = \|R_{x^*}^{-1} x\|$ and $p = \frac{R_{x^*}^{-1} x}{z}$. Define

$$
m_{x^*,p}(t) = f(R_{x^*}(tp)).
$$

From Taylor theorem, we have

$$
m_{x^*,p}(z) - m_{x^*,p}(0) = \frac{\mathrm{d}m_{x^*,p}(0)}{\mathrm{d}t}(z - 0) + \int_0^1 (1-t)\frac{\mathrm{d}^2 m_{x^*,p}(tz)}{\mathrm{d}t^2} z^2 \mathrm{d}t. \tag{29}
$$

Noting that

$$\frac{\mathrm{d}^2 m_{x^*,p}(tz)}{\mathrm{d}t^2} = g(\mathrm{Hess}\, f(R_{x^*}(tzp))[DR_{x^*}(tzp)[p]], DR_{x^*}(tzp)[p]),$$

then, it follows from (29) that

$$
\begin{aligned}
m_{x^*,p}(z) - m_{x^*,p}(0) &= g(\mathrm{grad}\, f(x^*), R_{x^*}^{-1}x) + \int_0^1 (1-t)g(\mathrm{Hess}\, f(R_{x^*}(tR_{x^*}^{-1}x)) \\
&\quad \times [\mathscr{L}(x^*, R_{x^*}(tR_{x^*}^{-1}x))R_{x^*}^{-1}x], \mathscr{L}(x^*, R_{x^*}(tR_{x^*}^{-1}x))R_{x^*}^{-1}x)\mathrm{d}t.
\end{aligned}
\tag{30}
$$

By combining (28) and (30), we obtain

$$
\begin{aligned}
\frac{1}{2}m'\|R_{x^*}^{-1}x\|^2 &= \int_0^1 (1-t)m'\|\mathscr{L}(x^*, R_{x^*}(tR_{x^*}^{-1}x))R_{x^*}^{-1}x\|^2 \\
&\leq f(x) - f(x^*) \leq \int_0^1 (1-t)\overline{m}\|\mathscr{L}(x^*, R_{x^*}(tR_{x^*}^{-1}x))R_{x^*}^{-1}x\|^2 \\
&= \frac{1}{2}\overline{m}\|R_{x^*}^{-1}x\|^2.
\end{aligned}
\tag{31}
$$

Hence, (17) holds.

For any $x, y \in M$ and $x \neq y$, let $\widetilde{z} = \|R_x^{-1}y\|$ and $\widetilde{p} = \frac{R_x^{-1}y}{\widetilde{z}}$. Define $m_{x,p}(t) = f(R_x(t\widetilde{p}))$. Then, we have

$$\frac{\mathrm{d}m_{x,p}(\widetilde{z})}{\mathrm{d}t} - \frac{\mathrm{d}m_{x,p}(0)}{\mathrm{d}t} = \int_0^{\widetilde{z}} \frac{\mathrm{d}^2 m_{x,p}(t)}{\mathrm{d}t^2}\mathrm{d}t.$$

Therefore,

$$
\begin{aligned}
&g(\mathrm{grad}\, f(y), \mathscr{L}(x, y)\widetilde{p}) - g(\mathrm{grad}\, f(x), \widetilde{p}) \\
&\quad = \int_0^{\widetilde{z}} g(\mathrm{Hess}\, f(R_x(t\widetilde{p}))[DR_x(t\widetilde{p})[\widetilde{p}]], DR_x(t\widetilde{p})[\widetilde{p}])\mathrm{d}t,
\end{aligned}
\tag{32}
$$

since

$$
\begin{aligned}
m'\|R_x^{-1}y\| &= \int_0^{\widetilde{z}} m'\|\widetilde{p}\|^2\mathrm{d}t \leq \int_0^{\widetilde{z}} g(\mathrm{Hess}\, f(R_x(t\widetilde{p}))[DR_x(t\widetilde{p})[\widetilde{p}]], DR_x(t\widetilde{p})[\widetilde{p}])\mathrm{d}t \\
&\leq \int_0^{\widetilde{z}} \overline{m}\|\widetilde{p}\|^2\mathrm{d}t = \overline{m}\|R_x^{-1}y\|.
\end{aligned}
$$

From (32) and the above inequality, we have

$$m'\|R_x^{-1}y\|^2 \leq g(\mathscr{L}(y, x)\mathrm{grad}\, f(y) - \mathrm{grad}\, f(x), R_x^{-1}y) \leq \overline{m}\|R_x^{-1}y\|^2.$$

Then, inequality (18) holds. Furthermore, we obtain

$$m'\|R_x^{-1}y\| \le \|\operatorname{grad} f(y) - \mathscr{L}(x, y)\operatorname{grad} f(x)\|.$$

In particular, when $y = x$ and $x = x^*$, one has

$$m'\|R_{x^*}^{-1}x\| \le \|\operatorname{grad} f(x)\|.$$

This completes the proof.   □

Remark 4.2 If the retraction R_x is an exponential mapping, then from [22], assumption (15) holds. We further illustrate assumption (15) by the following example.

We consider the sphere $\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n : x^\top x = 1\}$ with its structure of Riemannian submanifold of the Euclidean space $\mathbb{R}^n$. The projection retraction R_x is defined by

$$R_x(\eta) = (x + \eta)/\|x + \eta\|, \quad \forall x \in \mathbb{S}^{n-1}, \ \eta \in T_x\mathbb{S}^{n-1}.$$

The chosen isomeric vector transport $\mathscr{T}$ on $\mathbb{S}^{n-1}$ is

$$\mathscr{T}_{R_x(\eta)}\xi = \xi - \frac{2y^\top \xi}{\|x + y\|^2}(x + y), \quad \forall x \in M, \ \eta, \xi \in T_x\mathbb{S}^{n-1},$$

where $y = R_x(\eta)$. For further details, see, for example [22,31]. From Section 8.1 in [22] and Section 5 in [34], we obtain

$$\frac{\mathrm{D}}{\mathrm{d}t}\mathrm{D}R_x(t\eta)[\eta] = \frac{\mathrm{D}}{\mathrm{d}t}\mathscr{T}_{R_x(t\eta)}\eta = 0, \quad \forall x \in M, \ \eta \in T_x\mathbb{S}^{n-1}.$$

Thus, assumption (15) holds on $\mathbb{S}^{n-1}$ for the projection retraction.

Lemma 4.2 [13] *Let $S \subset M$ be an open set, $x \in S$ and the retraction $R_x : T_xM \to M$ 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 $\epsilon' > 0$ such that, for all $x \in S$ and $v, w \in T_xM$ with $\|v\|, \|w\| < \epsilon'$,

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

Theorem 4.1 *Suppose that all the assumptions of Lemma 4.1 are satisfied, B_k is uniformly positive definite, and d_k satisfies (13). Furthermore, assume that $\mathrm{D}R_{x_k}$ is equicontinuous on a neighborhood U of the stationary point x^*. Then, the sequence $\{x_k\}$ generated by Algorithm 2.1 converges R-linearly to x^*.*

Proof From (14) and (19), we have

$$\lim_{k \to +\infty} x_k = x^*.$$

It follows from Corollary 3.1, (17) and (19) that

$$f(x_k) - f(x_{k+1}) \geq \tau \mu^2 \|\text{grad} f(x_k)\|^2 \geq \tau \mu^2 m'^2 \|R_{x^*}^{-1} x_k\|^2$$
$$\geq \frac{2\tau \mu^2 m'^2}{\overline{m}} (f(x_k) - f(x^*)).$$

Set $\theta = \sqrt{2 \frac{\tau \mu^2 m'^2}{\overline{m}}} = m' \sqrt{2 \frac{\tau \mu^2}{\overline{m}}}$. Then,

$$f(x_k) - f(x_{k+1}) \geq \theta^2 (f(x_k) - f(x^*)). \tag{33}$$

We can prove that $\theta < 1$. Indeed, since f is twice continuously differentiable on M, it follows from Lemma 2.1 that there exists a constant $L_0 > 0$ such that

$$\|\mathscr{L}(y, x)\text{grad} f(y) - \text{grad} f(x)\| \leq L_0 \|R_x^{-1} y\|, \quad \forall x, y \in M.$$

Then,

$$g(\mathscr{L}(y, x)\text{grad} f(y) - \text{grad} f(x), R_x^{-1} y) \leq \|\mathscr{L}(y, x)\text{grad} f(y) - \text{grad} f(x)\| \|R_x^{-1} y\|$$
$$\leq L_0 \|R_x^{-1} y\|^2, \quad \forall x, y \in M.$$

Taking $L_0 = \overline{m}$, then by the definition of τ in the proof of Proposition 3.1, we obtain

$$\theta^2 = \frac{2\tau \mu^2 m'^2}{\overline{m}} \leq \frac{2\tau m'^2}{\overline{m}} \leq \frac{2m'^2}{\overline{m}} \cdot \frac{\beta \sigma (1 - \sigma)}{2L_0} \leq \beta \sigma (1 - \sigma) < 1.$$

Set $\omega = \sqrt{1 - \theta^2}$. Then, $\omega < 1$, and from (33), we obtain

$$f(x_{k+1}) - f(x^*) \leq (1 - \theta^2)(f(x_k) - f(x^*))$$
$$= \omega^2 (f(x_k) - f(x^*)) \leq \cdots \leq \omega^{2(k+1)} (f(x_0) - f(x^*)).$$

From (17), we have

$$\|R_{x^*}^{-1} x_{k+1}\|^2 \leq \frac{2}{m'} (f(x_{k+1}) - f(x^*)) \leq \frac{2\omega^{2(k+1)}}{m'} (f(x_0) - f(x^*)). \tag{34}$$

It follows from Lemma 4.2 that, for k sufficiently large,

$$
\begin{aligned}
d(x_{k+1}, x^*) &= d(R_{x^*}(R_{x^*}^{-1} x_{k+1}), R_{x^*}(0_{x^*})) \\
&\le \frac{3}{2} \| R_{x^*}^{-1} x_{k+1} - 0_{x^*} \| \\
&= \frac{3}{2} \| R_{x^*}^{-1} x_{k+1} \|.
\end{aligned}
$$

By (34), we get

$$
\begin{aligned}
\lim_{k \to +\infty} [d(x_{k+1}, x^*)]^{\frac{1}{k+1}} &\le \lim_{k \to +\infty} \left[\frac{3}{2} \| R_{x^*}^{-1} x_{k+1} \| \right]^{\frac{1}{k+1}} \\
&\le \lim_{k \to +\infty} \left(\omega^{k+1} \sqrt{\frac{2 \times \frac{9}{4}(f(x_0) - f(x^*))}{m'}} \right)^{\frac{1}{k+1}} = \omega < 1.
\end{aligned}
$$

Thus, x_k converges R-linearly to x^*. This completes the proof. $\square$

In the following, we assume that

(H) B_k and d_k generated by Algorithm 2.1 satisfy the following condition

$$
\lim_{k \to +\infty} \frac{\| B_k[d_k] - \mathscr{L}(x^*, x_k)(\mathrm{Hess}\, f(x^*)[\mathscr{L}(x_k, x^*)d_k]) \|}{\| d_k \|} = 0.
$$

Lemma 4.3 *Suppose that all the assumptions of Lemma 4.1 are satisfied, B_k is uniformly positive definite, $d_k = -B_k^{-1} \mathrm{grad}\, f(x_k)$ and the assumption (H) holds. Then, there exists $k' \in \mathbb{N}$ such that*

$$
\alpha_k = 1, \quad \forall k \ge k'.
$$

Proof From (14) and (19), we have

$$
\lim_{k \to +\infty} x_k = x^*. \tag{35}
$$

Since B_k is uniformly positive definite, there exists a constant $q > 0$ such that

$$
\| d_k \| = \| - B_k^{-1} \mathrm{grad}\, f(x_k) \| \le \| B_k^{-1} \| \| \mathrm{grad}\, f(x_k) \| \le q \| \mathrm{grad}\, f(x_k) \|.
$$

Then, from (14) and the above inequality, we obtain

$$
\lim_{k \to +\infty} \| d_k \| = 0, \tag{36}
$$

and thus,

$$
\lim_{k \to +\infty} R_{x_k}(t d_k) = x^*. \tag{37}
$$

Assumption (H) implies that

$$
\begin{aligned}
&\|g(B_k[d_k], d_k) - g(\mathscr{L}(x^*, x_k)(\mathrm{Hess}\, f(x^*)[\mathscr{L}(x_k, x^*)d_k]), d_k)\| \\
&= \|g(B_k[d_k] - \mathscr{L}(x^*, x_k)(\mathrm{Hess}\, f(x^*)[\mathscr{L}(x_k, x^*)d_k]), d_k)\| \\
&\le \|B_k[d_k] - \mathscr{L}(x^*, x_k)(\mathrm{Hess}\, f(x^*)[\mathscr{L}(x_k, x^*)d_k])\|\|d_k\| \\
&= \|d_k\|o(\|d_k\|),
\end{aligned}
$$

and thus,

$$
g(B_k[d_k] - \mathscr{L}(x^*, x_k)(\mathrm{Hess}\, f(x^*)[\mathscr{L}(x_k, x^*)d_k]), d_k) = o(\|d_k\|^2). \tag{38}
$$

Clearly,

$$
\begin{aligned}
&g(\mathrm{Hess}\, f(R_{x_k}(td_k))[\mathscr{L}(x_k, R_{x_k}(td_k))d_k] \\
&\quad - \mathscr{L}(x^*, R_{x_k}(td_k))(\mathrm{Hess}\, f(x^*)[\mathscr{L}(x_k, x^*)d_k]), \mathscr{L}(x_k, R_{x_k}(td_k))d_k) \\
&\le \|\mathrm{Hess}\, f(R_{x_k}(td_k))[\mathscr{L}(x_k, R_{x_k}(td_k))d_k] \\
&\quad - \mathscr{L}(x^*, R_{x_k}(td_k))(\mathrm{Hess}\, f(x^*)[\mathscr{L}(x_k, x^*)d_k])\|\|d_k\| \\
&= \|\mathrm{Hess}\, f(R_{x_k}(td_k))[\mathscr{L}(x^*, R_{x_k}(td_k))(\mathscr{L}(x_k, x^*)d_k)] \\
&\quad - \mathscr{L}(x^*, R_{x_k}(td_k))(\mathrm{Hess}\, f(x^*)[\mathscr{L}(x_k, x^*)d_k])\|\|d_k\| \\
&= \|(\mathrm{Hess}\, f(R_{x_k}(td_k))\mathscr{L}(x^*, R_{x_k}(td_k)) \\
&\quad - \mathscr{L}(x^*, R_{x_k}(td_k))\mathrm{Hess}\, f(x^*))[\mathscr{L}(x_k, x^*)d_k]\|\|d_k\| \\
&\le \|\mathrm{Hess}\, f(R_{x_k}(td_k))\mathscr{L}(x^*, R_{x_k}(td_k)) \\
&\quad - \mathscr{L}(x^*, R_{x_k}(td_k))\mathrm{Hess}\, f(x^*)\|\|\mathscr{L}(x_k, x^*)d_k\|\|d_k\| \\
&= \|\mathrm{Hess}\, f(R_{x_k}(td_k))\mathscr{L}(x^*, R_{x_k}(td_k)) - \mathscr{L}(x^*, R_{x_k}(td_k))\mathrm{Hess}\, f(x^*)\|\|d_k\|^2.
\end{aligned}
$$

This together with (37) implies that, for k sufficiently large,

$$
\begin{aligned}
&g(\mathrm{Hess}\, f(R_{x_k}(td_k))[\mathscr{L}(x_k, R_{x_k}(td_k))d_k] - \mathscr{L}(x^*, R_{x_k}(td_k)) \\
&\quad (\mathrm{Hess}\, f(x^*)[\mathscr{L}(x_k, x^*)d_k]), \mathscr{L}(x_k, R_{x_k}(td_k))d_k) = o(\|d_k\|^2).
\end{aligned}
$$

Define $m(t) = f(R_{x_k}(td_k))$. By using Taylor theorem, (38) and the above equality, we obtain for k sufficiently large,

$$
\begin{aligned}
m(1) - m(0) &= \left.\frac{\mathrm{d}m(t)}{\mathrm{d}t}\right|_{t=0} + \int_0^1 (1-t)\frac{\mathrm{d}^2 m}{\mathrm{d}t^2}(t)\mathrm{d}t \\
&= g(\mathrm{grad}\, f(x_k), d_k) + \int_0^1 (1-t)g(\mathrm{Hess}\, f(R_{x_k}(td_k))[DR_{x_k}(td_k)[d_k]], DR_{x_k}(td_k)[d_k])\mathrm{d}t \\
&= g(\mathrm{grad}\, f(x_k), d_k) + \frac{1}{2}g(B_k[d_k], d_k) \\
&\quad + \int_0^1 (1-t)\{g(\mathrm{Hess}\, f(R_{x_k}(td_k))[\mathscr{L}(x_k, R_{x_k}(td_k))d_k], \mathscr{L}(x_k, R_{x_k}(td_k))d_k)
\end{aligned}
$$

$$- g(\operatorname{Hess} f(x^*)[\mathscr{L}(x_k, x^*)d_k], \mathscr{L}(x_k, x^*)d_k)\}\mathrm{d}t - \frac{1}{2}g(B_k[d_k], d_k)$$

$$+ \frac{1}{2}g(\operatorname{Hess} f(x^*)[\mathscr{L}(x_k, x^*)d_k], \mathscr{L}(x_k, x^*)d_k)$$

$$= g(\operatorname{grad} f(x_k), d_k) + \frac{1}{2}g(B_k[d_k], d_k)$$

$$+ \int_0^1 (1 - t)\{g(\operatorname{Hess} f(R_{x_k}(td_k))[\mathscr{L}(x_k, R_{x_k}(td_k))d_k]$$

$$- \mathscr{L}(x^*, R_{x_k}(td_k))(\operatorname{Hess} f(x^*)[L(x_k, x^*)d_k]), \mathscr{L}(x_k, R_{x_k}(td_k))d_k)\}\mathrm{d}t$$

$$+ \frac{1}{2}g(\mathscr{L}(x^*, x_k)(\operatorname{Hess} f(x^*)[\mathscr{L}(x_k, x^*)d_k]) - B_k[d_k], d_k)$$

$$= g(\operatorname{grad} f(x_k), d_k) + \frac{1}{2}g(B_k[d_k], d_k) + o(\|d_k\|^2)$$

$$= -\frac{1}{2}g(\operatorname{grad} f(x_k), B_k^{-1}\operatorname{grad} f(x_k)) + o(\|d_k\|^2).$$

Thus, for k sufficiently large, we obtain

$$f(R_{x_k}(d_k)) - f(x_k) \le \sigma(g(\operatorname{grad} f(x_k), d_k) + \frac{1}{2}g(B_k[d_k], d_k)).$$

This implies that there exists $k' > 0$ such that

$$\alpha_k = 1, \quad \forall k \ge k'.$$

This completes the proof. $\qquad\square$

Theorem 4.2 *Suppose that all the assumptions of Lemma* 4.3 *are satisfied, and* $\mathrm{D}R_{x_k}$ *is equicontinuous on a neighborhood U of the stationary point x^*. Then, the sequence* $\{x_k\}$ *generated by Algorithm* 2.1 *converges superlinearly to x^*.*

Proof From Lemma 4.3, we obtain that there exists $k' > 0$ such that

$$x_{k+1} = R_{x_k}(d_k), \quad \forall k \ge k',$$

where $d_k = -B_k^{-1}\operatorname{grad} f(x_k)$. Let γ be the curve defined by $\gamma(t) = R_{x_k}(td_k)$. Then, we have

$$\|P_{\gamma[R_{x_k}(td_k), x_k]}\operatorname{Hess} f(R_{x_k}(td_k))[\mathscr{L}(x_k, R_{x_k}(td_k))d_k]$$

$$- \mathscr{L}(x^*, x_k)(\operatorname{Hess} f(x^*)[\mathscr{L}(x_k, x^*)d_k])\|$$

$$= \|(P_{\gamma[R_{x_k}(td_k), x_k]}\operatorname{Hess} f(R_{x_k}(td_k))\mathscr{L}(x^*, R_{x_k}(td_k))$$

$$- \mathscr{L}(x^*, x_k)\operatorname{Hess} f(x^*))[\mathscr{L}(x_k, x^*)d_k]\|$$

$$\le \|P_{\gamma[R_{x_k}(td_k), x_k]}\operatorname{Hess} f(R_{x_k}(td_k))\mathscr{L}(x^*, R_{x_k}(td_k))$$

$$- \mathscr{L}(x^*, x_k)\operatorname{Hess} f(x^*)\|\|d_k\|.$$

It follows from (35) and (37) that, for k sufficiently large,

$$\|P_{\gamma[R_{x_k}(td_k),x_k]}\text{Hess} f(R_{x_k}(td_k))[\mathscr{L}(x_k, R_{x_k}(td_k))d_k]$$
$$- \mathscr{L}(x^*, x_k)(\text{Hess} f(x^*)[\mathscr{L}(x_k, x^*)d_k])\| = o(\|d_k\|). \tag{39}$$

Noting that

$$P_{\gamma[x_{k+1},x_k]}\text{grad} f(x_{k+1}) - \text{grad} f(x_k)$$
$$= \int_0^1 P_{\gamma[R_{x_k}(td_k),x_k]}\text{Hess} f(R_{x_k}(td_k))[DR_{x_k}(td_k)[d_k]]dt$$
$$= \int_0^1 P_{\gamma[R_{x_k}(td_k),x_k]}\text{Hess} f(R_{x_k}(td_k))[\mathscr{L}(x_k, R_{x_k}(td_k))d_k]dt$$
$$= \mathscr{L}(x^*, x_k)(\text{Hess} f(x^*)[\mathscr{L}(x_k, x^*)d_k])$$
$$+ \int_0^1 \{P_{\gamma[R_{x_k}(td_k),x_k]}\text{Hess} f(R_{x_k}(td_k))[\mathscr{L}(x_k, R_{x_k}(td_k))d_k]$$
$$- \mathscr{L}(x^*, x_k)(\text{Hess} f(x^*)[\mathscr{L}(x_k, x^*)d_k])\}dt.$$

Since $d_k = -B_k^{-1}\text{grad} f(x_k)$, we have $\text{grad} f(x_k) = -B_k d_k$ and

$$P_{\gamma[x_{k+1},x_k]}\text{grad} f(x_{k+1})$$
$$= \mathscr{L}(x^*, x_k)(\text{Hess} f(x^*)[\mathscr{L}(x_k, x^*)d_k]) - B_k d_k$$
$$+ \int_0^1 \{P_{\gamma[R_{x_k}(td_k),x_k]}\text{Hess} f(R_{x_k}(td_k))[\mathscr{L}(x_k, R_{x_k}(td_k))d_k]$$
$$- \mathscr{L}(x^*, x_k)(\text{Hess} f(x^*)[\mathscr{L}(x_k, x^*)d_k])\}dt.$$

Consequently, for k sufficiently large, it follows from (39) that

$$\|P_{\gamma[x_{k+1},x_k]}\text{grad} f(x_{k+1})\|$$
$$\leq \|\mathscr{L}(x^*, x_k)(\text{Hess} f(x^*)[\mathscr{L}(x_k, x^*)d_k]) - B_k d_k\| + o(\|d_k\|).$$

By assumption (H), we have

$$\lim_{k\to+\infty} \frac{\|\text{grad} f(x_{k+1})\|}{\|d_k\|}$$
$$\leq \lim_{k\to+\infty} \frac{\|\mathscr{L}(x^*, x_k)(\text{Hess} f(x^*)[\mathscr{L}(x_k, x^*)d_k]) - B_k d_k\| + o(\|d_k\|)}{\|d_k\|} = 0. \tag{40}$$

From Lemma 4.2, we obtain for k sufficiently large,

$$\frac{1}{2}\|R_{x^*}^{-1}x_{k+1}\| \leq d(x_{k+1}, x^*) = d(R_{x^*}(R_{x^*}^{-1}x_{k+1}), R_{x^*}(0_{x^*})) \leq \frac{3}{2}\|R_{x^*}^{-1}x_{k+1}\|,$$

and

$$\frac{1}{2}\|R_{x_k}^{-1}x_{k+1}\| \le d(x_{k+1}, x_k) = d(R_{x_k}(R_{x_k}^{-1}x_{k+1}), R_{x_k}(0_{x_k})) \le \frac{3}{2}\|R_{x_k}^{-1}x_{k+1}\|.$$

This together with (19) implies that

$$\frac{\|\operatorname{grad} f(x_{k+1})\|}{\|d_k\|} \ge \frac{m'\|R_{x^*}^{-1}x_{k+1}\|}{\|R_{x_k}^{-1}x_{k+1}\|} \ge \frac{m' \times \frac{2}{3}d(x_{k+1}, x^*)}{2d(x_{k+1}, x_k)}$$

$$\ge \frac{m' \times \frac{1}{3}d(x_{k+1}, x^*)}{d(x_{k+1}, x^*) + d(x_k, x^*)} = \frac{\frac{m'}{3}\frac{d(x_{k+1}, x^*)}{d(x_k, x^*)}}{1 + \frac{d(x_{k+1}, x^*)}{d(x_k, x^*)}}.$$

Consequently, from (40), we obtain

$$\lim_{k \to +\infty} \frac{d(x_{k+1}, x^*)}{d(x_k, x^*)} = 0,$$

which implies that $\{x_k\}$ converges superlinearly to x^*. This completes the proof. $\square$

Theorem 4.3 *Suppose that all the assumptions of Lemma 4.3 are satisfied, $B_k =$ Hess $f(x_k)$ for k sufficiently large, and DR_{x_k} is equicontinuous on a neighborhood U of the stationary point x^*. Moreover, assume there exists a constant $\widetilde{L} > 0$ such that*

$$\|\operatorname{Hess} f(y) - P_{\gamma[x,y]}\operatorname{Hess} f(x)\mathscr{L}(y, x)\| \le \widetilde{L}d(x, y), \quad \forall x, y \in U.$$

Then, the sequence $\{x_k\}$ generated by Algorithm 2.1 converges quadratically to x^*.

Proof From Lemma 4.3, there exists $k' > 0$ such that

$$x_{k+1} = R_{x_k}(d_k), \quad \forall k \ge k',$$

where $d_k = -(\operatorname{Hess} f(x_k))^{-1}\operatorname{grad} f(x_k)$. Let γ be the curve defined by $\gamma(t) = R_{x_k}(tR_{x_k}^{-1}x^*)$. Then, it follows from Lemma 4.2 that, for k sufficiently large,

$$\|\operatorname{Hess} f(x_k)[d_k - R_{x_k}^{-1}x^*]\|$$

$$= \left\| P_{\gamma[x^*,x_k]}\operatorname{grad} f(x^*) - \operatorname{grad} f(x_k) - \operatorname{Hess} f(x_k)[R_{x_k}^{-1}x^*] \right\|$$

$$= \left\| \int_0^1 P_{\gamma[R_{x_k}(tR_{x_k}^{-1}x^*),x_k]}\operatorname{Hess} f(R_{x_k}(tR_{x_k}^{-1}x^*))[DR_{x_k}(tR_{x_k}^{-1}x^*)[R_{x_k}^{-1}x^*]]\mathrm{d}t \right.$$

$$\left. - \operatorname{Hess} f(x_k)[R_{x_k}^{-1}x^*] \right\|$$

$$\le \int_0^1 \|(\operatorname{Hess} f(R_{x_k}(tR_{x_k}^{-1}x^*)) - P_{\gamma[x_k, R_{x_k}(tR_{x_k}^{-1}x^*)]}\operatorname{Hess} f(x_k)\mathscr{L}(R_{x_k}(tR_{x_k}^{-1}x^*), x_k))$$

$$[\mathscr{L}(x_k, R_{x_k}(tR_{x_k}^{-1}x^*))R_{x_k}^{-1}x^*]\|\mathrm{d}t$$

$$\le \int_0^1 \|\operatorname{Hess} f(R_{x_k}(tR_{x_k}^{-1}x^*)) - P_{\gamma[x_k, R_{x_k}(tR_{x_k}^{-1}x^*)]}\operatorname{Hess} f(x_k)\mathscr{L}(R_{x_k}(tR_{x_k}^{-1}x^*), x_k)\|\|R_{x_k}^{-1}x^*\|\mathrm{d}t$$

$$\leq \int_0^1 \widetilde{L} d(x_k, R_{x_k}(t R_{x_k}^{-1} x^*)) \| R_{x_k}^{-1} x^* \| dt$$

$$= \widetilde{L} \| R_{x_k}^{-1} x^* \| \int_0^1 d(x_k, R_{x_k}(t R_{x_k}^{-1} x^*)) dt$$

$$= \widetilde{L} \| R_{x_k}^{-1} x^* \| \int_0^1 d(R_{x_k}(0_{x_k}), R_{x_k}(t R_{x_k}^{-1} x^*)) dt$$

$$\leq \widetilde{L} \| R_{x_k}^{-1} x^* \| \int_0^1 \frac{3}{2} \| t R_{x_k}^{-1} x^* \| dt = \frac{3}{4} \widetilde{L} \| R_{x_k}^{-1} x^* \|^2. \tag{41}$$

Since f is uniformly retraction-convex on M, from (16), there exists $m' > 0$ such that

$$\| \operatorname{Hess} f(x_k)[d_k - R_{x_k}^{-1} x^*] \| \geq m' \| d_k - R_{x_k}^{-1} x^* \|. \tag{42}$$

Furthermore, it follows from Lemma 4.2 that, for k sufficiently large,

$$d(x_{k+1}, x^*) = d(R_{x_k}(d_k), R_{x_k}(R_{x_k}^{-1} x^*)) \leq \frac{3}{2} \| R_{x_k}^{-1} x^* - d_k \|, \tag{43}$$

and

$$d(x_k, x^*) = d(R_{x_k}(0_{x_k}), R_{x_k}(R_{x_k}^{-1} x^*)) \geq \frac{1}{2} \| R_{x_k}^{-1} x^* \|. \tag{44}$$

By (41), (42), (43) and (44), we obtain

$$\frac{2}{3} d(x_{k+1}, x^*) \leq \| d_k - R_{x_k}^{-1} x^* \| \leq \frac{1}{m'} \| \operatorname{Hess} f(x_k)[d_k - R_{x_k}^{-1} x^*] \|$$

$$\leq \frac{3\widetilde{L}}{4m'} \| R_{x_k}^{-1} x^* \|^2 \leq \frac{3\widetilde{L}}{m'} d^2(x_k, x^*).$$

Thus,

$$d(x_{k+1}, x^*) \leq \frac{9\widetilde{L}}{2m'} d^2(x_k, x^*),$$

which implies that $\{x_k\}$ converges quadratically to x^*. This completes the proof. □

Remark 4.3 Smith [21, Theorem 4.4] presented Newton's method on Riemannian manifolds and proved that its convergence is quadratic. If the retraction R_x is the exponential mapping, and the isometric vector transport is the parallel transport, then Theorem 4.3 can be seen as a generalization of Theorem 4.4 in [21]. If $M = \mathbb{R}^n$ and $R_x(\eta) = x + \eta$, then Theorems 4.1, 4.2 and 4.3 reduce to Theorems 4.1, 5.1 and 5.3 in [17].

Example 4.1 Let $\operatorname{St}(p, n)$ $(p \leq n)$ denote the set of all $n \times p$ orthonormal matrices, that is,

$$\operatorname{St}(p, n) := \left\{ X \in \mathbb{R}^{n \times p} : X^\top X = I_p \right\},$$

where I_p denotes the $p \times p$ identity matrix. The set $\operatorname{St}(p, n)$ is called Stiefel manifold. The tangent space of $\operatorname{St}(p, n)$ is given by

$$T_X \operatorname{St}(p, n) = \left\{ X\Omega + X_\perp K : \Omega^\top = -\Omega, \ K \in \mathbb{R}^{(n-p) \times p} \right\},$$

where $X_\perp$ is any $n \times (n - p)$ orthonormal matrix such that $(X\ X_\perp)$ is an orthogonal matrix. The canonical inner product is given by

$$g(Z_1, Z_2) = \text{trace}(Z_1^\top Z_2), \quad \forall Z_1, Z_2 \in T_X\text{St}(p, n).$$

As in [6], the isometric vector transport from $T_{X_1}\text{St}(p, n)$ to $T_{X_2}\text{St}(p, n)$ is given by

$$\mathscr{L}(X_1, X_2) = \mathscr{B}_{X_2}\mathscr{B}_{X_1}^\flat,$$

where $a^\flat$ denotes the flat of $a \in T_X\text{St}(p, n)$, i.e., $a^\flat : T_X\text{St}(p, n) \to \mathbb{R} : v \to g(a, v)$. Moreover, a smooth function $\mathscr{B} : U \to \mathbb{R}^{np \times (\frac{np-p(p+1)}{2})} : X \mapsto \mathscr{B}_X$ is defined on an open set U of $\text{St}(p, n)$, and the columns of $\mathscr{B}_X$ form an orthonormal basis of $T_X\text{St}(p, n)$. Let $X \in \text{St}(p, n)$. Then, from [6,31], we obtain

$$(R_X(\eta)\ R_X(\eta)_\perp) = (X\ X_\perp) \exp\begin{pmatrix} \Omega & -K^\top \\ K & 0_{(n-p)\times(n-p)} \end{pmatrix},$$

where $\Omega = X^\top\eta$ and $K = X_\perp^\top\eta$. The function $Y = R_X(\eta)$ is the desired retraction by the isometric vector transport $\mathscr{L}(\cdot, \cdot)$. Consider the cost function

$$f(X) = \text{trace}(X^\top A X),$$

where $A = \text{diag}(\mu_1, \ldots, \mu_p)$ with $0 < \mu_1 < \cdots < \mu_p$. From Chapter 11 in [31], the gradient of f is

$$\text{grad} f(X) = P_X(2AX),$$

where $\text{sym}(Q) = \frac{Q+Q^\top}{2}$ and $P_X(V) = X\text{sym}(X^\top V)$. Moreover, the Hessian of f on $\eta \in T_X\text{St}(p, n)$ is given by

$$\text{Hess} f(X)\eta = P_X\left(2A\eta - \eta\text{sym}(X^\top(2AX))\right).$$

Let $\beta, \sigma \in]0, 1[$, $B_k = \text{Hess} f(x_k)$ and $d_k = -B_k^{-1}\text{grad} f(x_k)$ for all $k \in \mathbb{N}$. Define $m_{X,\eta}(t) = f(R_X(t\eta))$ for all $X \in \text{St}(p, n)$ and $\eta \in T_X\text{St}(p, n)$. From the above equality, it is easy to check that $\frac{d^2 m_{X,\eta}(t)}{dt^2} > 0$. Thus, by Remark 4.1, f is uniformly retraction-convex. Furthermore, the vector transport $\mathscr{L}(\cdot, \cdot)$ is parallel translation. From Section 8.1 in [22] and Section 5 in [34], it follows that assumption (15) holds. Thus, all the assumptions of Theorem 4.1 are satisfied. Assume that X^* is a stationary point of f. Then, the sequence $\{X_k\}$ generated by Algorithm 2.1 converges R-linearly to the stationary point X^*.

Example 4.2 Let $M = \mathbb{H} := \{(x_1, x_2) \in \mathbb{R}^2 : x_2 > 0\}$ be the Poincaré plane endowed with the Riemannian metric given by

$$g_{ij} = \frac{1}{x_2^2}\delta_{ij}, \quad i, j = 1, 2.$$

Then, $\mathbb{H}$ is a Hadamard manifold with the constant sectional curvature -1 (see [14]). The geodesics in $\mathbb{H}$ are vertical semilines

$$C_a : x = a, \ y = ce^t, \quad t \in]-\infty, +\infty[,$$

and semicircles

$$C_{b,r} : x = b - r \tanh t, \ y = \frac{r}{\cosh t}, \quad t \in]-\infty, +\infty[.$$

For fixed $(x_0, y_0) \in \mathbb{H}$ and for the vector $p = (p_1, p_2) \in T_{(x_0, y_0)}\mathbb{H}$, we have

$$
\begin{aligned}
R_{(x_0, y_0)}(tp) &= \exp_{(x_0, y_0)}(tp) \\
&= \begin{cases} (x_0, y_0 e^t), & \text{if } p_1 = 0, \ p_2 = y_0, \\ \left(x_0 + \frac{p_2}{p_1}\|p\| + \frac{\|p\|^2}{p_1}\tanh t, -y_0 \frac{\|p\|}{p_1}\frac{1}{\cosh t}\right), & \text{if } p_1 < 0, \ t \in [0, +\infty[, \end{cases}
\end{aligned}
$$

where $\|p\|^2 = \|p_1\|^2 + \|p_2\|^2$. For further details, see Chapter 1 in [14]. Let $f : \mathbb{H} \to \mathbb{R}$ be a twice continuously differentiable function. Then, as in [14], the gradient of f and the Hessian of f are given by

$$\operatorname{grad} f((x_1, x_2)) = \left(x_2^2 \frac{\partial f}{\partial x_1}, \ x_2^2 \frac{\partial f}{\partial x_2}\right),$$

and

$$\operatorname{Hess} f((x_1, x_2)) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} - \frac{1}{x_2}\frac{\partial f}{\partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} + \frac{1}{x_2}\frac{\partial f}{\partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} + \frac{1}{x_2}\frac{\partial f}{\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} + \frac{1}{x_2}\frac{\partial f}{\partial x_2}, \end{pmatrix},$$

respectively. Let $C \subseteq H$ be defined by

$$C := \{(x_1, x_2) \in \mathbb{H} : (x_1 - 1)^2 + x_2^2 \le 5, \ x_1 \ge x_2 \ge 1\}.$$

Then, C is geodesic convex in $\mathbb{H}$ (see, [14]). Consider the cost function

$$f((x_1, x_2)) = \ln^2 \frac{x_1}{x_2}, \quad \forall (x_1, x_2) \in C.$$

Then, we obtain

$$\operatorname{Hess} f((x_1, x_2)) = \begin{pmatrix} \frac{2}{x_1^2}(1 - \ln \frac{x_1}{x_2}) + \frac{2}{x_2^2}\ln \frac{x_1}{x_2} & \frac{2}{x_1 x_2}(\ln \frac{x_1}{x_2} - 1) \\ \frac{2}{x_1 x_2}(\ln \frac{x_1}{x_2} - 1) & \frac{2}{x_2^2} \end{pmatrix}, \quad \forall (x_1, x_2) \in C.$$

It is easy to check f is uniformly retraction-convex on C. However, f is not uniformly convex on C in Euclidean sense. Suppose that the isometric vector transport is parallel transport. Let $\beta, \sigma \in]0, 1[$, $B_k = \operatorname{Hess} f(x_k)$ and $d_k = -B_k^{-1}\operatorname{grad} f(x_k)$

for all $k \in \mathbb{N}$. Then, all the assumptions of Theorems 4.1, 4.2 and 4.3 are satisfied. Hence, the sequence $\{x_k\}$ generated by Algorithm 2.1 converges R-linearly / super-linearly/quadratically to the stationary point x^*.

5 Conclusions

We proposed a descent method with new inexact line-search for unconstrained optimization problems on Riemannian manifolds. By using the retraction and isometric vector transport, the global convergence of the descent method with new inexact line-search on Riemannian manifolds is obtained, under some appropriate assumptions. Moreover, the R-linear / superlinear / quadratic convergence rate of the descent method, with new inexact line-search, is extended from linear spaces to Riemannian manifolds. In the future, we shall explore some other efficiently computable retractions and vector transport, which can be adjusted in the given problems. Moreover, it is interesting to do some numerical experiments and comparisons with other line-search algorithms for practical problems on Riemannian manifolds.

References

1. Absil, P.A., Baker, C.G., Gallivan, K.A.: Trust-region methods on Riemannian manifolds. Found. Comput. Math. **7**, 303–330 (2007)
2. Burago, D., Burago, Y., Ivanov, S.: A Course in Metric Geometry. Graduate Studies in Math., vol. 33. Amer. Math. Soc, Providence (2001)
3. Chavel, I.: Riemannian Geometry—A Modern Introduction. Cambridge University Press, London (1993)
4. da Cruz Neto, J.X., de Lima, L.L., Oliveira, P.R.: Geodesic algorithms in Riemannian geometry. Balkan J. Geom. Appl. **3**(2), 89–100 (1998)
5. da Cruz Neto, J.X., Ferreira, O.P., Lucambio Perez, L.R.: A proximal regularization of the steepest descent method in Riemannian manifold. Balkan J. Geom. Appl. **4**(2), 1–8 (1999)
6. Huang, W., Gallivan, K.A., Absil, P.A.: A Broyden class of quasi-Newton methods for Riemannian optimization. SIAM J. Optim. **25**, 1660–1685 (2015)
7. Klingenberg, W.: A Course in Differential Geometry. Springer, Berlin (1978)
8. Li, S.L., Li, C., Liou, Y.C., Yao, J.C.: Existence of solutions for variational inequalities on Riemannian manifolds. Nonlinear Anal. **71**, 5695–5706 (2009)
9. Li, C., Mordukhovich, B.S., Wang, J.H., Yao, J.C.: Weak sharp minima on Riemannian manifolds. SIAM J. Optim. **21**, 1523–1560 (2011)
10. Li, C., Yao, J.C.: Variational inequalities for set-valued vector fields on Riemannian manifolds: convexity of the solution set and the proximal point algorithm. SIAM J. Control Optim. **50**, 2486–2514 (2012)
11. Li, X.B., Zhou, L.W., Huang, N.J.: Gap functions and global error bounds for generalized mixed variational inequalities on Hadamard manifolds. J. Optim. Theory Appl. **168**(3), 830–849 (2016)
12. Rapcsák, T.: Smooth Nonlinear Optimization in $\mathbb{R}^n$. Kluwer Academic Publishers, Dordrecht (1997)
13. Ring, W., Wirth, B.: Optimization methods on Riemannian manifolds and their application to shape space. SIAM J. Optim. **22**(2), 596–627 (2012)

14. Udrişte, C.: Convex Functions and Optimization Methods on Riemannian Manifolds. Kluwer Academic Publishers, Dordrecht (1994)
15. Yang, Y.: Globally convergent optimization algorithms on Riemannian manifolds: uniform framework for unconstrained and constrained optimization. J. Optim. Theory Appl. **132**, 245–265 (2007)
16. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer Series in Operations Research. Springer, New York (1999)
17. Shi, Z.J., Shen, J.: Convergence of descent method with new line search. J. Appl. Math. Comput. **20**, 239–254 (2006)
18. Armijo, L.: Minimization of functions having Lipschitz continuous first partial derivatives. Pac. J. Math. **16**, 1–3 (1966)
19. Wolfe, P.: Convergence conditions for ascent methods. SIAM Rev. **11**, 226–235 (1969)
20. Shi, Z.J.: Convergence of quasi-Newton method with new inexact line search. J. Math. Anal. Appl. **315**, 120–131 (2006)
21. Smith, S.T.: Optimization techniques on Riemannian manifolds: Hamiltonian and gradient flows. Algorithm Control **3**, 113–136 (1994)
22. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton (2008)
23. Ferreira, O.P., Svaiter, B.F.: Kantorovich's theorem on Newton's method in Riemannian manifolds. J. Complex. **18**, 304–329 (2002)
24. Adler, R.L., Dedieu, J.P., Malajovich, J.Y., Martens, M., Shub, M.: Newton's method on Riemannian manifolds and a geodesic model for the human spine. IMA J. Numer. Anal. **22**, 359–390 (2002)
25. Dedieu, J.P., Priouret, P., Malajovich, G.: Newton's method on Riemannian manifolds: covariant alpha theory. IMA J. Numer. Anal. **23**, 395–419 (2003)
26. Li, C., Wang, J.H.: Newton's method for sections on Riemannian manifolds: generalized covariant α-theory. J. Complex. **24**, 423–451 (2008)
27. Li, C., Wang, J.H.: Convergence of the Newton method and uniqueness of zeros of vector fields on Riemannian manifolds. Sci. China Math. **48**, 1465–1479 (2005)
28. Brace, I., Manton, J.H.: An improved BFGS-on-manifold algorithm for computing weighted low rank approximations. In: Proceeding of the 17th International Symposium on Mathematical Theory of Networks and Systems, pp. 1735–1738 (2006)
29. Sakai, T.: Riemannian Geometry, Translations of Mathematical Monographs. American Mathematical Society, Providence (1996)
30. Shub, M.: Some remarks on dynamical systems and numerical analysis. In: Lara-Carrero, L., Lwowicz, J. (eds.) Dynamical Systems and Partial Differential Equations: Proceedings of VII ELAM Caracas, pp. 69–92. Equinoccio Universidad Simon Bolivar (1986)
31. Huang, W.: Optimization algorithms on Riemannian manifolds with applications. Ph.D. Thesis, Department of Mathematics, Florida State University, Tallahassee, FL (2013)
32. Ortega, J.W., Rheinboldt, W.C.: Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York (2003)
33. do Carmo, M.P.: Riemannian Geometry, Mathematics: Theory Applications. Birkhäuser, Boston (1992)
34. Huang, W., Absil, P.A., Gallivan, K.A.: A Riemannian symmetric rank-one trust-region method. Math. Program. **150**, 179–216 (2015)