A Unifying Local Convergence Result for Newton's Method in Riemannian Manifolds

Felipe Alvarez[*], Jérôme Bolte[†] , Julien Munier[‡]

Abstract: We consider the problem of finding a singularity of a vector field X on a complete Riemannian manifold. In this regard we prove a unified result for local convergence of Newton's method. Inspired by previous work of Zabrejko and Nguen on Kantorovich's majorant method, our approach relies on the introduction of an abstract one-dimensional Newton's method obtained using an adequate Lipschitz-type radial function of the covariant derivative of X. The main theorem gives in particular a synthetic view of several famous results, namely the Kantorovich, Smale and Nesterov-Nemirovskii theorems. Concerning real-analytic vector fields an application of the central result leads to improvements of some recent developments in this area.

Key-words: vector field, Newton's method, quadratic convergence, existence of a singularity

[*] Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático (CNRS UMI 2807), Universidad de Chile, Blanco Encalada 2120, Santiago, Chile. Email: falvarez@dim.uchile.cl.

[†] Equipe Combinatoire et Optimisation, Case 189, UFR 921, Université Pierre et Marie Curie, 4 Place Jussieu, 75252 Paris Cedex 5, France. Fax: (33) 1 44 27 27 24, Email: bolte@math.jussieu.fr

[‡] Institut de Mathématiques et Modélisation de Montpellier (CNRS UMR 5149), Université Montpellier II, case 051, Place E. Bataillon, 34095 Montpellier Cedex 5, France. Email: munier@math.univ-montp2.fr

Un résultat unificateur sur la convergence locale de la méthode de Newton dans les variétés riemanniennes

Résumé : Etant donné un champ de vecteurs X sur une variété riemannienne, le problème central de ce papier est de mettre en évidence un procédé algorithmique permettant d'en détecter une singularité. Dans cette optique, un résultat unificateur concernant la convergence locale de l'algorithme de Newton est proposé. Inspiré par les travaux de Zabrejko et Nguen sur la méthode des majorants de Kantorovich, l'approche développée repose sur l'introduction d'une méthode de Newton en dimension un - cette dernière étant obtenue via une paramétrisation radiale d'un module de Lipschitz adéquat de la dérivée covariante de X. Le résultat principal qui en découle offre une vue synthétique de quelques résultats célèbres, à savoir les théorèmes de Kantorovich, Smale et Nesterov-Nemirovskii. Des résultats récents concernant les champs de vecteurs analytiques sont améliorés.

Mots-clés : champ de vecteurs, méthode de Newton, convergence quadratique, existence d'une singularité

1 Introduction

In this paper we investigate the existence, local uniqueness and iterative approximation of solutions to the problem of finding a singularity of a continuously differentiable vector field X defined on a connected and finite-dimensional manifold M. In fact, we assume that M is endowed with a Riemannian metric g with (M, g) being complete, and we consider the iterates generated by a Riemannian version of Newton's method applied to X. The choice of this method is based on the following prominent feature: under mild nondegeneracy conditions, there exists a non-trivial set of initial points sufficiently close to the solution so that Newton's method converges with quadratic rate.

A rather unsatisfying aspect of this qualitative property is that "sufficiently close" may depend explicitly on the solution which one does not know *a priori*. However, in many interesting cases, some implicit quantitative criteria have been provided to verify the proximity of the starting point to a solution and thus ensure quadratic convergence of Newton's method.

The celebrated Kantorovich theorem [12] on Newton's method in Banach spaces gives the first set of quantitative assumptions ensuring existence and uniqueness of a solution in a prescribed ball around the starting point together with quadratic convergence of the method. Kantorovich's result requires the knowledge of a local Lipschitz constant for the first derivative of X. In a similar spirit, a local analysis of Newton's method applied to analytic mappings led Smale to introduce in [18] the so called α-test, a fundamental proximity criterion in point estimation theory which uses information about all the derivatives of the data only at the initial point. In the specific case of minimization problems, Nesterov and Nemirovskii developed in [14] a proximity test for Newton's method under the so called self-concordancy condition on the objective function. This local convergence result is a key piece of their breakthrough in the study of the computational complexity of central path algorithms in mathematical programming. Together with other facts, these results of Kantorovich, Smale and Nesterov-Nemirovskii explain the uncontested success of Newton's method in the construction of efficient algorithms for the iterative resolution of several classes of nonlinear equations, and particularly in the designing of polynomial-time algorithms for optimization problems.

Generalizations of Kantorovich's result and Smale's α-theorem to Riemannian manifolds were recently established by Ferreira and Svaiter in [9] and by Dedieu, Priouret and Malajovich in [5], respectively. To our best knowledge, the theory of self-concordancy has not been extended to a Riemannian setting yet. The main motivation for enlarging the usual Euclidean setting to Riemannian manifolds stems essentially from the necessity to be able to deal with (equality) nonlinear constraints, especially in minimization problems; see, for instance, the works by Adler *et al.* [1], da Cruz Neto *et al.* [6], Edelman *et al.* [8], Smith [19] and Udriste [20].

The goal of this paper is to establish a general local convergence theorem based on a unifying principle which permits to recover, sometimes with certain improvements, all the above mentioned results. The hypotheses and the proof of our central result (cf. Theorem

3.1) rely on two important features appearing either explicitly or implicitly in most of the local analysis of Newton's method.

First, since the singularity-finding problem is indeed a metric-free problem, the choice of a particular Riemannian structure for implementing Newton's method is a strategy among others. The choice of an adequate distance measure is of primary importance, not only because of its dramatic consequences for obtaining good basic estimations, but also for the well-posedness of the method. Let us also observe that the sequence generated by Newton's method in Riemannian manifold may strongly depend on the metric. This contrasts with the case of $\mathbb{R}^n$ viewed as an Euclidean space, a case for which Newton's iterates never depend on the choice of an inner product.

The second and possibly most crucial aspect in setting up a unified theorem for the local convergence of Newton's method consists in a radial parametrization of Kantorovich's majorant method. This technique was originally developed by Zabrejko and Nguen in [25] (see also [2, 24]) in order to obtain refinements of the Kantorovich theorem in Banach spaces. In fact, this approach is based on the construction of a real-valued function, namely the majorant function, using an adequate local Lipschitz-type radial estimate for the first derivative of X. The majorant method reduces the problems of the existence and the local uniqueness of a singularity as well as the quadratic convergence of Newton's method to the same issues but within a scalar setting. Independently, similar ideas were used by Cominetti in [4] and by Wang and Han in [22] for Newton's method in Banach spaces (see also [21, 23]), both improving as a special case the Smale α-theorem. Cominetti also showed in [4] how to recover, at least partially, a basic result of the Nesterov-Nemirovskii self-concordancy theory but without isolating the role of the majorant function.

The paper is organized as follows. In §2 some basic definitions and results of the theory of Riemannian manifolds are recalled. In §3 the central result of the paper is stated, and its proof is given in §4. In §5 it is shown how the classical results of Kantorovich, Nesterov-Nemirovskii and Smale, as well as some Riemannian counterparts excepting for Nesterov-Nemirovskii, can be recovered by specialization of our central result.

2 Notation and basic notions of Riemannian geometry

Let us recall some basic definitions and properties of Riemannian geometry. Some general references on this subject are [7, 13, 15]. Readers who are not familiar with Riemannian geometry are referred to [8] for a pleasant introduction to some of its main concepts and their use in optimization.

Let (M, g) be a connected and finite-dimensional Riemannian manifold. The space of C^1 vector fields on M is denoted by $\chi(M)$. Classically, we denote by T_pM the tangent space to M at p and by $|\cdot|_p$ the norm on T_pM which is given by $|v|_p = \sqrt{g(p)(v, v)}$.

Norms and distances. The Riemannian distance $d : M \times M \to [0, +\infty)$ is defined by $d(p, q) = \inf\{\int_a^b |\dot{c}(t)|_{c(t)} dt \mid c : [a, b] \to M$ is piecewise smooth, $c(a) = p$ and $c(b) = q\}$. Here,

$\int_a^b |\dot{c}(t)|_{c(t)} dt$ is called the length of c, and we will write it simply $\int_a^b |\dot{c}|$ when no confusion can arise. We shall often identify the curves $c : [0, a] \to M$ on M with their graphs, so that $(t, p) \in c$ simply means that $t \in [0, a]$ with $c(t) = p$. Let $B(p_0, r) = \{p \in M \mid d(p, p_0) < r\}$ and $\overline{B}(p_0, r) = \{p \in M \mid d(p, p_0) \leq r\}$ be respectively the open and closed balls of center p_0 and radius r.

Take $p \in M$ and some integer $k \geq 1$. If $T : T_pM^k \to T_pM$ is a linear mapping its norm is defined by

$$||T||_p = \sup\{|T(u_1, \ldots, u_k)|_p \mid u_i \in T_pM, |u_i|_p \leq 1\}.$$

Covariant derivatives and parallel transport. Denote by ∇ the Riemannian (or Levi-Civita) connection on (M, g). For each pair of continuously differentiable vector fields X, Y, the vector field $\nabla_Y X$ stands for the covariant derivative of X with respect to Y. Given a vector field X on M and $p \in M$, define

$$X'(p)v := \nabla_v X(p) = (\nabla_Y X)(p), \ v \in T_pM. \tag{1}$$

where Y is any vector field on M satisfying $v = Y(p)$. The map $X'(p) : T_pM \to T_pM$ is well defined and linear, and we will call it the covariant derivative of X at p.

Let $c : [a, b] \to M$ be a smooth curve and V a vector field along c, that is, a differentiable mapping such that $V(t) \in T_{c(t)}M$ for all $t \in [a, b]$. The covariant derivative of V along c is denoted by $\nabla_{\dot{c}} V$ and defines a vector field along c. The vector field V is called parallel along c when $\nabla_{\dot{c}} V = 0$; in particular, as ∇ is Riemannian, $|V(t)|_{c(t)}$ is constant. Given $v \in T_{c(a)}M$, there exists a unique vector field V parallel along c such that $V(a) = v$, and the parallel transport of v along c from $c(a)$ to $c(b)$ is defined by $P_{c,a,b}v = V(b)$. The linear map $P_{c,a,b} : T_{c(a)}M \to T_{c(b)}M$ is an isometry. The extension of this definition to a piecewise smooth curve c is straightforward. It is direct to verify that $P_{c,a,b}^{-1} = P_{c,b,a}$ and $P_{c,a,b} = P_{c,t,b} \circ P_{c,a,t}$. Recall that a $(1, q)$-tensor can be viewed as a linear[1] mapping from $\chi(M)^q$ in $\chi(M)$. Likewise the tensors can be covariantly derivated (e.g. [15]), if T is $(1, q)$-tensor its derivative is a $(1, q + 1)$-tensor denoted T' which is given by the following formula

$$T'(X_1, \ldots, X_q, X) = [T(X_1, \ldots, X_q)]'(X) - \sum_{i=1}^q T(X_1, \ldots, X_i'(X), \ldots, X_q),$$

where the X_i's and X are vector fields on M. As usual higher order derivatives are defined recursively by $T^{(k+1)} = [T^{(k)}]'$, $k \in \mathbb{N}$. The parallel transport of vectors can be extended to tensors as follows. The curve c being chosen as above, take a $(1, q)$-tensor T on $T_{c(b)}M$, the parallel transport of T along c from $c(b)$ to $c(a)$ is defined by

$$[\mathbb{P}_{c,b,a}T](u_1, \ldots, u_q) := P_{c,b,a}[T(P_{c,a,b}u_1, \ldots, P_{c,a,b}u_q)],$$

[1] The linearity refers here to the structure of $C^1(M)$-module of space of C^1 vector fields $\chi(M)$.

where $u_1, \ldots, u_q \in T_{c(a)}M$. Parallel transportation provides tensors with a fundamental theorem of calculus, namely,

$$\mathbb{P}_{c,t,0}T(c(t)) = T(c(0)) + \int_0^t [\mathbb{P}_{c,s,0}T'(c(s))\,\dot{c}(s)]ds, \qquad (2)$$

which reduces to

$$P_{c,t,0}X(c(t)) = X(c(0)) + \int_0^t P_{c,s,0}X'(c(s))\,\dot{c}(s)ds,$$

if $q = 0$ (see, for instance, [9]).

Geodesic curves and the exponential map. Recall that a C^2 curve γ is a geodesic if $\dot{\gamma}$ is parallel along γ so that its speed $|\dot{\gamma}(t)|_{\gamma(t)}$ is constant. A curve $c : [0,a] \to M$ is said to be piecewise geodesic if there exists a partition $0 = t_0 < t_1 \ldots < t_N = a$ of $[0,a]$ such that the restriction of c to each interval of the form $[t_i, t_{i+1}]$, $i \in \{0, \ldots, N-1\}$ is a geodesic curve.

From now on, (M, d) is assumed to be a complete metric space. By the Hopf-Rinow theorem, the latter is equivalent to the geodesic completeness of (M, g), i.e., for any $p \in M$ and $v \in T_pM$ there exists a unique geodesic γ with $\gamma(0) = p$ and $\dot{\gamma}(0) = v$ such that $\gamma(t)$ is defined for all $t \in \mathbb{R}$. In addition, we have that for any $p, q \in M$ there exists a geodesic γ joining p and q whose length is equal to $d(p, q)$. Such a curve in M is referred to as a *minimizing geodesic*, joining p and q.

The exponential map at p, $\exp_p : T_pM \to M$ is defined by setting $\exp_p[v] = \gamma(1)$, where $\gamma : \mathbb{R} \to M$ is the geodesic with $\gamma(0) = p$ and $\dot{\gamma}(0) = v$, and $\exp_p[tv] = \gamma(t)$ for all $t \in \mathbb{R}$ due to the uniqueness of geodesics under initial conditions. We always have $d(p, \exp_p[v]) \leq |v|_p$, and the equality holds if the geodesic γ restricted to $[0, 1]$ is minimizing.

3   Riemannian Newton's method and local convergence

Let X be a C^1 vector field defined on a connected, complete and finite-dimensional Riemannian manifold (M, g). Consider the following problem:

$$\text{find } p^* \in M \text{ satisfying } X(p^*) = 0 \in T_{p^*}M. \qquad (3)$$

Such a point p^* is referred to as a singularity of X.

First, assume that there exists $p_0 \in M$ such that the covariant derivative $X'(p_0)$ of X at p_0 (see (1)) is invertible, i.e.,

$$X'(p_0) \in GL(T_{p_0}M). \qquad (4)$$

Starting at p_0, the Riemannian-Newton (or R-Newton) method associated with (3) writes

$$(\mathcal{N}) \qquad p_{k+1} = \exp_{p_k}[-X'(p_k)^{-1}X(p_k)],$$

where $\exp_p : T_pM \to M$ is the exponential map at p and $X'(p) : T_pM \to T_pM$ is defined by (1). This natural Riemannian version of Newton's method for solving systems of nonlinear equations has been considered by several authors [1, 5, 6, 8, 19, 20].

In order to state a general local convergence result for R-Newton's method, we will assume a Lipschitz-type continuity of X' on a neighborhood of p_0, based on the following definition.

Definition 3.1 *Let $\mathcal{G}_2(p_0, r)$ be the class of all the piecewise geodesic curves $c : [0, T] \to M$ which satisfy the following conditions:*

(a) $c(0) = p_0$ and the length of c is no greater than r.

(b) There exists $\tau \in (0, T]$ such that $c_{|[0,\tau]}$ is a minimizing geodesic and $c_{|[\tau,T]}$ is a geodesic.

We suppose that for some $R > 0$ there exists a continuous and nondecreasing function $\ell : [0, R] \to [0, +\infty)$ satisfying the following property: for every $r \in [0, R]$ and $c \in \mathcal{G}_2(p_0, r)$,

$$\|X'(p_0)^{-1}[\mathbb{P}_{c,b,0}X'(c(b)) - \mathbb{P}_{c,a,0}X'(c(a))]\|_{p_0} \le \ell(r) \int_a^b |\dot{c}|, \quad 0 \le a \le b, \quad (5)$$

We call ℓ a *Lipschitz-type radial function for X' around p_0*. Without loss of generality we may assume that $\ell(r) > 0$ for all $r \in (0, R]$.

Remark 3.1 It is important to observe that ℓ can often be estimated in practice. For instance, assume that X is C^2 and consider $\ell(r) \ge \max\{\|X'(p_0)^{-1}[\mathbb{P}_{c,t,0}X''(c(t))]\|_{p_0}\}$, where the maximum is taken over all piecewise geodesic curves belonging to $\mathcal{G}_2(p_0, r)$. It is straightforward to see that (2) implies that ℓ complies with (5).

Condition (5) is a scaled Riemannian analogue of the property used by Zabrejko and Nguen in [25] (see also [2, 24]) to prove a local convergence result for Newton's method in Banach spaces, based on a radial parametrization of the original "majorant method" developed by Kantorovich [12]. Here, scaling means that the inverse of $X'(p_0)$ is incorporated in the distance between covariant derivatives of X, an idea that has been already used in the Banach space context (see [4, 23]).

The striking feature of the majorant method is that the analysis of (3) and ($\mathcal{N}$) reduces to the analysis of an appropriate scalar equation together with the corresponding scalar Newton iteration. Following the idea of [25], we set

$$\beta := |X'(p_0)^{-1}X(p_0)|_{p_0},$$

and introduce the real function $\phi : [0, R] \to \mathbb{R}$ defined by

$$\begin{cases} \phi''(r) = \ell(r), \ \forall r \in (0, R), \\ \phi'(0) = -1, \\ \phi(0) = \beta. \end{cases}$$

A straightforward integration yields

$$\phi(r) = \beta - r + \int_0^r (r-s)\ell(s)ds, \ r \in [0, R].$$

(6)

The following scalar Newton sequence is at the heart of the Zabrejko-Nguen refinement of Kantorovich's majorant method (our central result, namely Theorem 3.1, shows indeed that under mild conditions Newton's iteration $(\mathcal{N})$ inherits of most of its properties):

$$\begin{cases} r_0 = 0, \\ r_{k+1} = r_k - \phi'(r_k)^{-1}\phi(r_k). \end{cases}$$

(7)

Let us consider the following assumption:

the function ϕ given by (6) has a unique zero r^* in $[0, R]$ with $\phi(R) \leq 0$.

(8)

Since $r^* \geq \beta$, we may suppose that $r^* > 0$, otherwise p_0 is a singularity of X and there is nothing to do. Under (8) it is elementary to establish the following result (see, for instance, [25, Proposition 3] or Lemma 4.1(i)-(ii) below).

Lemma 3.1 *Under* (8), *the scalar sequence* $\{r_k\}$ *generated by* (7) *is well defined with* $\phi'(r_k) < 0$ *for all* $k \in \mathbb{N}$, *hence monotonically increasing and convergent towards* r^*.

The central result of this paper may be now stated as follows:

Theorem 3.1 *Under* (4)-(8) *we have the following results.*
(i) The vector field X admits a unique singularity p^* in $\overline{B}(p_0, R)$ which belongs to $\overline{B}(p_0, r^*)$. If $\phi'(r^*) < 0$ then $X'(p^*) \in GL(T_{p^*}M)$.
(ii) The sequence $\{p_k\}$ defined by $(\mathcal{N})$ is well defined, that is, $X'(p_k) \in GL(T_{p_k}M)$ for every $k \in \mathbb{N}$.
(iii) For every $k \in \mathbb{N}$, p_k belongs to $\overline{B}(p_0, r_k)$ and the following estimate holds

$$d(p_{k+1}, p_k) \leq |X'(p_k)^{-1}X(p_k)|_{p_k} \leq r_{k+1} - r_k.$$

(9)

As a consequence, p_k converges to p^* as $k \to +\infty$, and for every $k \in \mathbb{N}$,

$$d(p^*, p_k) \leq r^* - r_k.$$

(10)

(iv) For every $k \in \mathbb{N}$,

$$\frac{|X'(p_{k+1})^{-1}X(p_{k+1})|_{p_{k+1}}}{r_{k+2} - r_{k+1}} \leq \left(\frac{|X'(p_k)^{-1}X(p_k)|_{p_k}}{r_{k+1} - r_k}\right)^2.$$

(11)

Consequently, for all $k \geq k_0 \geq 0$,

$$d(p^*, p_k) \leq (r^* - r_k)\left(\frac{|X'(p_{k_0})^{-1}X(p_{k_0})|_{p_{k_0}}}{r_{k_0+1} - r_{k_0}}\right)^{2^{k-k_0}}.$$

(12)

(v) *Setting*

$$q = \frac{1 - \sqrt{1 - 2\lambda}}{1 + \sqrt{1 - 2\lambda}} \in (0, 1] \tag{13}$$

for

$$\lambda = \frac{r^{*2}\phi''(r^*)^2 - 2r^*\phi''(r^*)\phi'(r^*)}{2[r^*\phi''(r^*) - \phi'(r^*)]^2} \in (0, \frac{1}{2}], \tag{14}$$

then for all $k \in \mathbb{N}$, $r_k \geq r^*(1 - q^{2^k - 1})(1 - q^{2^k})^{-1}$ if $q < 1$, and $r_k \geq r^*(1 - 2^{-k})$ if $q = 1$. In particular, if $\lambda < 1/2$ then for every $k \in \mathbb{N}$,

$$r^* - r_k \leq r^* \frac{1 - q}{1 - q^{2^k}} q^{2^k - 1}. \tag{15}$$

Remark 3.2 By (10) together with Lemma 4.1(iii) below, we have in particular that $d(p^*, p_0) \leq r^* \leq 2\beta$. Since $r_0 = 0$ and $r_1 = \beta$, then the second estimation in (9) is an equality for $k = 0$, and we have $d(p^*, p_1) \leq r^* - \beta \leq \beta$. If the second inequality in (9) is strict for some $k_0 \geq 1$, then (12) gives an *a posteriori* correction improving, for all $k \geq k_0$, the *a priori* estimate (10).

Remark 3.3 In (v), when $r^* = R$ then $\phi''(r^*)$ stands the left derivative of ϕ' at R. On the other hand, notice that by Lemma 3.1, $\phi'(r^*) \leq 0$. It is easy to see that $\lambda = 1/2$ iff $\phi'(r^*) = 0$, which gives the linear rate of convergence $d(p^*, p_k) \leq r^* 2^{-k}$, $\forall k \geq 0$. Otherwise, we obtain the quadratic rate of convergence given by (15). As we will see in the proof, these are worst-case estimates based on a quadratic approximation of the majorant function ϕ (see (23)). Sharper estimates may be obtained by a direct analysis of Newton's method applied to ϕ.

Remark 3.4 The continuity of the Lipschitz-type radial function ℓ is not necessary for Theorem 3.1(i)-(iv). However, this is a simplifying assumption that holds in interesting special cases (see, for instance, Section 5).

Remark 3.5 Theorem 3.1 remains valid for a C^1 vector field $X : \Omega \subset M \to TM$ which is defined only on an open subset Ω of M provided that $\overline{B}(p_0, R) \subset \Omega$.

The next section is devoted to the proof of Theorem 3.1.

4   Proof of Theorem 3.1

The proof of Theorem 3.1 is divided into three parts, namely the existence and the convergence of Newton's sequence, the local uniqueness of the singularity, and the estimates.

4.1 Existence and convergence

4.1.1 Preliminary results

From now on, we assume that the hypothesis of Theorem 3.1 hold, with $\ell(r) > 0$ for all $r \in (0, R]$. We first need to establish a few properties of the function ϕ to enlighten its links with Newton's method in X. This is precisely the purpose of the following lemmas.

Lemma 4.1 Let ϕ be the function given by (6). Then:
(i) ϕ is strictly convex hence ϕ' is increasing on $[0, R]$, and $\phi'(r) < 0$ for all $r \in [0, r^*)$.
(ii) If $r \in [0, r^*)$ and $r^+ = r - \phi'(r)^{-1}\phi(r)$ then $r^+ \in (r, r^*)$.
(iii) For every $r \in [0, R]$, $\beta - r \leq \phi(r) \leq \beta + r(\phi'(r) - 1)/2$. Consequently, $\beta \leq r^* \leq 2\beta$.
(iv) For each $a \in (0, R)$, the function $\varphi_a(r) = \phi'(r + a) - \phi'(r) = \int_r^{r+a} \ell(s)ds > 0$ is nondecreasing on $[0, R - a]$.

Proof. (i) The strict convexity of ϕ on $[0, R]$ follows directly from $\phi''(r) = \ell(r) > 0$ for $r \in (0, R)$. Next, assume that $\phi'(\tilde{r}) = 0$ for some $\tilde{r} \in [0, r^*)$. By convexity, $\tilde{r}$ is a minimum for ϕ on $[0, R]$, hence $0 \leq \phi(\tilde{r}) \leq \phi(R) \leq 0$. Therefore, $\tilde{r}$ is a zero of ϕ, which contradicts the uniqueness of r^*. Since $\phi'(0) = -1$, this proves $\phi' < 0$ on $[0, r^*)$.
(ii) By (i), $r^+ > r$. On the other hand, since ϕ is strictly convex, we have that $\phi(r) + \phi'(r)(r^* - r) < \phi(r^*) = 0$, which amounts to $r^+ < r^*$.
(iii) Of course, $\phi(r) \geq \beta - r$. On the other hand, as $\phi'(r) = -1 + \int_0^r \ell(s)ds$, we have that $2\phi(r) = 2\beta + r(\phi'(r)-1) + \omega(r)$ for $\omega(r) = \int_0^r (r-2s)\ell(s)ds$. But $\omega'(r) = \int_0^r [\ell(s)-\ell(r)]ds \leq 0$, hence $\omega(r) \leq \omega(0) = 0$ and thus $2\phi(r) \leq 2\beta + r(\phi'(r) - 1)$. Since $\phi'(r^*) \leq 0$ by (i), we deduce that $\beta - r^* \leq 0 \leq \beta - r^*/2$.
(iv) This follows directly from $\varphi'_a(r) = \ell(r + a) - \ell(r) \geq 0$. $\qquad\square$

The first estimate of the next result is an infinitesimal version of (5) which extends [25, Proposition 1] to the Riemannian setting.

Lemma 4.2 Let $c \in \mathcal{G}_2(p_0, R)$. Then for any $a \leq b$, we have

$$\|X'(p_0)^{-1}[\mathbb{P}_{c,b,0}X'(c(b)) - \mathbb{P}_{c,a,0}X'(c(a))]\|_{p_0} \leq \phi'(\textstyle\int_0^b |\dot{c}|) - \phi'(\textstyle\int_0^a |\dot{c}|). \tag{16}$$

As a consequence, if $\phi'(\int_0^t |\dot{c}|) < 0$ then $X'(c(t)) \in GL(T_{c(t)}M)$ and moreover

$$\|[\mathbb{P}_{c,t,0}X'(c(t))^{-1}]X'(p_0)\|_{p_0} \leq -\phi'\left(\int_0^t |\dot{c}|\right)^{-1}. \tag{17}$$

Proof. Set $\Delta_0 = \|X'(p_0)^{-1}[\mathbb{P}_{c,b,0}X'(c(b)) - \mathbb{P}_{c,a,0}X'(c(a))]\|_{p_0}$. Given $N \in \mathbb{N}$ with $N \geq 1$, define $t_i = a + \frac{i}{N}(b - a)$ and $y_i = c(t_i)$ for all $i \in \{0, \ldots, N\}$. By (5), we have

$$\Delta_0 \leq \sum_{i=0}^{N-1} \|X'(p_0)^{-1}[\mathbb{P}_{c,t_{i+1},0}X'(y_{i+1}) - \mathbb{P}_{c,t_i,0}X'(y_i)]\|_{p_0} \leq \sum_{i=0}^{N-1} \ell\left(\int_0^{t_{i+1}} |\dot{c}|\right)\int_{t_i}^{t_{i+1}} |\dot{c}|.$$

Letting $N \to \infty$, we obtain $\Delta_0 \leq \int_a^b \ell(\int_0^\tau |\dot c|)|\dot c|d\tau = \int_{r(a)}^{r(b)} \ell(s)ds$ with $r(t) = \int_0^t |\dot c|$, which proves (16).

Taking $a = 0$ and $b = t$, (16) yields $\|X'(p_0)^{-1}\mathbb{P}_{c,t,0}X'(c(t)) - I\|_{p_0} \leq \phi'(\int_0^t |\dot c|) + 1$, where $I : T_{p_0}M \to T_{p_0}M$ is the identity mapping. If $\phi'(\int_0^t |\dot c|) < 0$ then this estimate is strictly lower than 1 and so $X'(p_0)^{-1}\mathbb{P}_{c,t,0}X'(c(t))$ is invertible. Therefore $X'(c(t)) \in GL(T_{c(t)}M)$ and moreover $\|[X'(p_0)^{-1}\mathbb{P}_{c,t,0}X'(c(t))]^{-1}\|_{p_0} \leq 1/(-\phi'(\int_0^t |\dot c|))$ which amounts to (17). $\square$

The following result about several fundamental properties of Newton's iteration is a generalization of [9, Lemma 3.7].

Lemma 4.3 *Given $r \in [0, r^*)$ and $p \in \overline{B}(p_0, r)$ such that $|X'(p)^{-1}X(p)|_p \leq -\phi'(r)^{-1}\phi(r)$, set $\tau(\theta) = r - \theta\phi'(r)^{-1}\phi(r)$ and $\gamma(\theta) = \exp_p[-\theta X'(p)^{-1}X(p)]$ for $\theta \in [0, 1]$. Then for every $\theta \in (0, 1]$,*
   (i) *$\tau(\theta) \in (r, r^*)$ and $\gamma(\theta) \in \overline{B}(p_0, \tau(\theta))$,*
   (ii) *$X'(\gamma(\theta)) \in GL(T_{\gamma(\theta)}M)$,*
   (iii) *$|X'(\gamma(\theta))^{-1}X(\gamma(\theta))|_{\gamma(\theta)} \leq -\phi'(\tau(\theta))^{-1}\phi(\tau(\theta))$.*

Proof. (i) Fix $\theta \in (0, 1]$. By Lemma 4.1(i)-(ii), $r = \tau(0) < \tau(\theta) \leq \tau(1) = r^+ < r^*$. On the other hand, as γ is the geodesic with $\gamma(0) = p$ and $\dot\gamma(0) = -X'(p)^{-1}X(p)$, we have $d(\gamma(\theta), p) \leq \theta|X'(p)^{-1}X(p)|_p$ and $d(\gamma(\theta), p_0) \leq \theta|X'(p)^{-1}X(p)|_p + d(p, p_0) \leq -\theta\phi'(r)^{-1}\phi(r) + r = \tau(\theta)$, which proves $\gamma(\theta) \in \overline{B}(p_0, \tau(\theta))$.

In order to prove (ii), let us denote by $c : [0, 2] \to M$ the curve obtained by concatenation of a minimizing geodesic joining p_0 and p, defined on $[0, 1]$, and γ. This is a piecewise geodesic curve with $c(0) = p_0$, $c(1 + \theta) = \gamma(\theta)$ for all $\theta \in [0, 1]$, whose length is lower than r^* by (i). To establish the invertibility of $X'(\gamma(\theta))$, let us estimate the following norm

$$\|X'(p)^{-1}\mathbb{P}_{\gamma,\theta,0}X'(\gamma(\theta)) - I_{T_pM}\|_p$$
$$= \|X'(p)^{-1}P_{c,0,1}X'(p_0)X'(p_0)^{-1}P_{c,1,0}[\mathbb{P}_{\gamma,\theta,0}X'(\gamma(\theta)) - X'(p)]\|_p$$
$$\leq \|[\mathbb{P}_{c,1,0}X'(p)^{-1}]X'(p_0)\|_{p_0}\|X'(p_0)^{-1}[\mathbb{P}_{c,1+\theta,0}X'(c(1+\theta)) - \mathbb{P}_{c,1,0}X'(p)]\|_{p_0}.$$

By Lemma 4.2 and the monotonicity of ϕ', it ensues that

$$\|[\mathbb{P}_{c,1,0}X'(p)^{-1}]X'(p_0)\|_{p_0} \leq -\phi'(d(p_0, p))^{-1} \leq -\phi'(r)^{-1},$$

and

$$\|X'(p_0)^{-1}[\mathbb{P}_{c,1+\theta,0}X'(c(1+\theta)) - \mathbb{P}_{c,1,0}X'(p)]\|_{p_0} \leq \phi'(\tau(\theta)) - \phi'(r).$$

Consequently, $\|X'(p)^{-1}\mathbb{P}_{\gamma,\theta,0}X'(\gamma(\theta)) - I_{T_pM}\|_p \leq 1 - \phi'(r)^{-1}\phi'(\tau(\theta)) < 1$. It follows that $X'(\gamma(\theta))$ is invertible with in addition

$$|[\mathbb{P}_{\gamma,\theta,0}X'(\gamma(\theta))^{-1}]X'(p)\|_p \leq \phi'(\tau(\theta))^{-1}\phi'(r). \tag{18}$$

Let us deal with (iii). Recalling (2), we notice that

$$|X'(\gamma(\theta))^{-1}X(\gamma(\theta))|_{\gamma(\theta)} = |X'(\gamma(\theta))^{-1}P_{\gamma,0,\theta}[X(p) + \int_0^\theta P_{\gamma,s,0}X'(\gamma(s))\dot\gamma(s)ds]|_{\gamma(\theta)}.$$

Since $\dot{\gamma}$ is parallel along γ, $\dot{\gamma}(s) = -P_{\gamma,0,s}X'(p)^{-1}X(p)$. Setting $\Theta = |X'(\gamma(\theta))^{-1}X(\gamma(\theta))|_{\gamma(\theta)}$, it follows

$$
\begin{aligned}
\Theta &= |X'(\gamma(\theta))^{-1}P_{\gamma,0,\theta}[X(p) - \int_0^\theta \mathbb{P}_{\gamma,s,0}X'(\gamma(s))X'(p)^{-1}X(p)ds]|_{\gamma(\theta)} \\
&= |X'(\gamma(\theta))^{-1}P_{\gamma,0,\theta}[(1-\theta)X(p) + \int_0^\theta[X'(p) - \mathbb{P}_{\gamma,s,0}X'(\gamma(s))]X'(p)^{-1}X(p)ds]|_{\gamma(\theta)}
\end{aligned}
$$

Therefore

$$
|X'(\gamma(\theta))^{-1}X(\gamma(\theta))|_{\gamma(\theta)} \le (1-\theta)|X'(\gamma(\theta))^{-1}P_{\gamma,0,\theta}X(p)|_{\gamma(\theta)} + R(\theta)
$$

with

$$
R(\theta) = |X'(\gamma(\theta))^{-1}P_{\gamma,0,\theta}\int_0^\theta[X'(p) - \mathbb{P}_{\gamma,s,0}X'(\gamma(s))]X'(p)^{-1}X(p)ds|_{\gamma(\theta)}.
$$

By (18), this proves that

$$
|X'(\gamma(\theta))^{-1}X(\gamma(\theta))|_{\gamma(\theta)} \le -(1-\theta)\phi'(\tau(\theta))^{-1}\phi(r) + R(\theta). \tag{19}
$$

Let us estimate $R(\theta)$. Since, by Lemma 4.2 and recalling that ϕ' is increasing, we have $\|[\mathbb{P}_{c,1+\theta,0}X'(\gamma(\theta))^{-1}]X'(p_0)\|_{p_0} \le -\phi'(\tau(\theta))^{-1}$, it follows that

$$
\begin{aligned}
R(\theta) &\le -\phi'(\tau(\theta))^{-1}|X'(p_0)^{-1}P_{c,1,0}\int_0^\theta[X'(p) - \mathbb{P}_{\gamma,s,0}X'(\gamma(s))]X'(p)^{-1}X(p)ds|_{p_0} \\
&\le -\phi'(\tau(\theta))^{-1}\int_0^\theta\|X'(p_0)^{-1}[\mathbb{P}_{c,1,0}X'(p) - \mathbb{P}_{c,1+s,0}X'(c(1+s))]\|_{p_0}ds|X'(p)^{-1}X(p)|_p \\
&\le \phi'(\tau(\theta))^{-1}\phi'(r)^{-1}\phi(r)\int_0^\theta\|X'(p_0)^{-1}[\mathbb{P}_{c,1,0}X'(p) - \mathbb{P}_{c,1+s,0}X'(c(1+s))]\|_{p_0}ds.
\end{aligned}
$$

Again, Lemma 4.2 yields

$$
R(\theta) \le \phi'(\tau(\theta))^{-1}\phi'(r)^{-1}\phi(r)\int_0^\theta\left[\phi'(\int_0^{1+s}|\dot{c}|) - \phi'(\int_0^1|\dot{c}|)\right]ds,
$$

hence with lemma 4.1(iv)

$$
\begin{aligned}
R(\theta) &\le \phi'(\tau(\theta))^{-1}\phi'(r)^{-1}\phi(r)\left[\int_0^\theta\phi'(r - s\phi'(r)^{-1}\phi(r))ds - \theta\phi'(r)\right] \\
&= -\phi'(\tau(\theta))^{-1}[\phi(\tau(\theta)) - (1-\theta)\phi(r)].
\end{aligned}
$$

It follows from (19) that $|X'(\gamma(\theta))^{-1}X(\gamma(\theta))|_{\gamma(\theta)} \le -\phi'(\tau(\theta))^{-1}\phi(\tau(\theta))$, which achieves the proof. $\qquad\square$

4.1.2 Proof of (i)-(iii): existence and convergence

Let us first prove by induction that, for all $k \geq 0$, $p_k \in B(p_0, r_k)$ and $X'(p_k)$ is invertible with $|X'(p_k)^{-1}X(p_k)|_{p_k} \leq r_{k+1} - r_k$. The case where $k = 0$ follows from the assumptions of Theorem 3.1. Let $k \geq 1$ and assume that the result holds for all $i \in \{0, \ldots, k-1\}$. We have

$$d(p_k, p_0) \leq \sum_{i=0}^{k-1} d(p_{i+1}, p_i) \leq \sum_{i=0}^{k-1} |X'(p_i)^{-1}X(p_i)|_{p_i} \leq r_k,$$

and since $|X'(p_{k-1})^{-1}X(p_{k-1})|_{p_{k-1}} \leq r_k - r_{k-1}$, Lemma 4.3 (for $\theta = 1$) yields the conclusion. The estimate (9) follows immediately. Since $\{r_k\}$ converges to r^* by Lemma 3.1, we deduce that $\{p_k\}$ is a Cauchy sequence in the complete manifold (M, d), hence it has a limit p^*, which is a singularity of X. Indeed, we have that $|X(p_k)|_{p_k} \leq \|X'(p_k)\|_{p_k}(r_{k+1} - r_k)$ and letting $k \to \infty$ we obtain $X(p^*) = 0$. The estimate (10) and the fact that $p^* \in \overline{B}(p_0, r^*)$ are elementary consequences of (9). If $\phi'(r^*) < 0$ then, by Lemma 4.2, we deduce that $X'(p^*)$ is invertible.

4.2 Uniqueness

We have proved that $\{p_k\}$ is well defined and convergent towards a singularity p^* of X. Next, we extend to our more general setting the proof of the local uniqueness of p^* that is given in [9, §3.2] for the special case of the Riemannian Kantorovich theorem. It is worth pointing out that this extension is only technical; the key arguments are the same of [9].

4.2.1 Auxiliary results

The following two results extend [9, Lemma 3.8] and [9, Corollary 3.9] to our framework.

Lemma 4.4 *Let $0 \leq r < r^*$ and $q \in \overline{B}(p_0, r)$ be such that $|X'(q)^{-1}X(q)|_q \leq -\phi'(r)^{-1}\phi(r)$. Suppose that $q^* \in \overline{B}(p_0, r^*)$ satisfies $X(q^*) = 0$ and $r + d(q, q^*) = r^*$. Then $d(p_0, q) = r$. Furthermore, defining $r^+ = r - \phi'(r)^{-1}\phi(r)$ and $q^+ = \exp_q[-X'(q)^{-1}X(q)]$, we have*

$$r^+ + d(q^+, q^*) = r^*.$$

Proof. Let $\gamma : [0, 1] \to M$ be a minimizing geodesic joining q and q^*. We denote $v = \dot{\gamma}(0)$, which satisfies $|v|_q = d(q, q^*)$. Thus $d(q, q^*) \leq |X'(q)^{-1}X(q) + v|_q + |X'(q)^{-1}X(q)|_q \leq |X'(q)^{-1}X(q) + v|_q - \phi'(r)^{-1}\phi(r)$. On the other hand, since $X(q^*) = 0$, by (2) we have $X(q) = -\int_0^1 \mathbb{P}_{\gamma,s,0}X'(\gamma(s))v\,ds$. Therefore

$$|X'(q)^{-1}X(q) + v|_q = |X'(q)^{-1}\int_0^1 [\mathbb{P}_{\gamma,s,0}X'(\gamma(s)) - X'(q)]v\,ds|_q.$$

Let denote by $c : [0,2] \to M$ the concatenation of $\tilde{\gamma}$, a minimizing geodesic joining p_0 and q, with γ. We have

$$|X'(q)^{-1}X(q) + v|_q$$
$$\leq \|[\mathbb{P}_{\tilde{\gamma},1,0}X'(q)^{-1}]X'(p_0)\|_{p_0}|X'(p_0)^{-1}P_{\tilde{\gamma},1,0}\int_0^1[\mathbb{P}_{\gamma,s,0}X'(\gamma(s)) - X'(q)]vds|_{p_0}$$
$$\leq \|[\mathbb{P}_{c,1,0}X'(q)^{-1}]X'(p_0)\|_{p_0}\int_0^1\|X'(p_0)^{-1}[\mathbb{P}_{c,1+s,0}X'(c(s)) - \mathbb{P}_{c,1,0}X'(q)]\|_{p_0}|v|_qds.$$

It follows from Lemma 4.2 that

$$|X'(q)^{-1}X(q) + v|_q \leq -\phi'(d(p_0,q))^{-1}\int_0^1[\phi'(d(p_0,q) + s|v|_q) - \phi'(d(p_0,q))]|v|_qds.$$

As $d(p_0,q) \leq r$, Lemma 4.1 yields $\phi'(d(p_0,q)) \leq \phi'(r) < 0$ and $\varphi_{a_s}(d(p_0,q)) \leq \varphi_{a_s}(r)$ for $a_s = s|v|_q$, hence

$$|X'(q)^{-1}X(q) + v|_q \leq -\phi'(r)^{-1}\int_0^1(\phi'(r + s|v|_q) - \phi'(r))|v|_qds = |v|_q + \phi'(r)^{-1}\phi(r),$$

where we have used that $r + |v|_q = r^*$ which implies $\phi(r + |v|_q) = 0$. But we have seen that $|v|_q \leq |X'(q)^{-1}X(q) + v|_q - \phi'(r)^{-1}\phi(r)$. This proves that all these inequalities are equalities. In particular, as $|v|_q \neq 0$, it follows that

$$\phi'(d(p_0,q)) = \phi'(r), \tag{20}$$
$$|X'(q)^{-1}X(q) + v|_q = |v|_q - |X'(q)^{-1}X(q)|_q, \tag{21}$$
$$|X'(q)^{-1}X(q)|_q = -\phi'(r)^{-1}\phi(r). \tag{22}$$

By the injectivity of ϕ' on $[0, r^*)$ (cf. Lemma 4.1), it follows from (20) that $d(p_0,q) = r$. From (21), we deduce that $X'(q)^{-1}X(q) = -\lambda v$ for some $0 < \lambda < 1$. Thus $q^+ = \exp_q[\lambda v] = \gamma(\lambda)$, therefore $d(q,q^+) = |X'(q)^{-1}X(q)|_q$ and $d(q^+,q^*) = d(q,q^*) - d(q,q^+) = r^* - r - |X'(q)^{-1}X(q)|_q = r^* - r^+$, where we have used (22). $\qquad\square$

Lemma 4.5 *Let $q^* \in \overline{B}(p_0,r^*)$ with $X(q^*) = 0$. If there exist $\tilde{r},\tilde{q}$ such that $0 \leq \tilde{r} < r^*$, $\tilde{q} \in \overline{B}(p_0,\tilde{r})$, $|X'(\tilde{q})^{-1}X(\tilde{q})|_{\tilde{q}} \leq -\phi'(\tilde{r})^{-1}\phi(\tilde{r})$, $\tilde{r} + d(\tilde{q},q^*) = r^*$, then $d(p_0,q^*) = r^*$.*

Proof. Let $\{\tau_k\}$ and $\{q_k\}$ be defined by $\tau_0 = \tilde{r}$, $\tau_{k+1} = \tau_k - \phi'(\tau_k)^{-1}\phi(\tau_k)$ and $q_0 = \tilde{q}$, $q_{k+1} = \exp_{q_k}[-X'(q_k)^{-1}X(q_k)]$ respectively. As for $\{r_k\}$ and $\{p_k\}$, it is possible to prove that τ_k is increasing and tends to r^*, q_k converges to some $\tilde{q}^* \in \overline{B}(p_0,r^*)$, and for all k, $|X'(q_k)^{-1}X(q_k)|_{q_k} \leq -\phi'(\tau_k)^{-1}\phi(\tau_k)$. Moreover, by Lemma 4.4, we know that for all k, $\tau_k + d(q_k,q^*) = r^*$ and $d(p_0,q_k) = \tau_k$. Combining all these informations leads to $q^* = \tilde{q}^*$ and so $d(p_0,q^*) = r^*$. $\qquad\square$

4.2.2 Proof of the uniqueness of the singularity

Let us now prove the uniqueness of p^* in the ball $\overline{B}(p_0,R)$. We begin by establishing the uniqueness of the singularity in $\overline{B}(p_0,r^*)$. Let $q^* \in \overline{B}(p_0,r^*)$ be such that $X(q^*) = 0$. In order to prove that necessarily $q^* = p^*$ we consider two cases.

Case 1 If $d(p_0, q^*) < r^*$, we show by induction that for all k, $d(p_k, q^*) + r_k < r^*$. Indeed, the initialization just needs $r_0 = 0$. If the property is true for a fixed k, we set $\gamma(\theta) = \exp_{p_k}[-\theta X'(p_k)^{-1} X(p_k)]$ and $\psi(\theta) = d(\gamma(\theta), q^*) + r_k + \theta(r_{k+1} - r_k)$. We know that $\psi(0) < r^*$. If there exists $\tilde{\theta}$ such that $\psi(\tilde{\theta}) = r^*$, then Lemma 4.3 ensures that Lemma 4.5 applies with $\tilde{r} = r_k - \tilde{\theta}\phi'(r_k)^{-1}\phi(r_k)$ and $\tilde{q} = \gamma(\tilde{\theta})$. But its conclusion contradicts $d(p_0, q^*) < r^*$. Thus, by continuity of ψ, one has $\psi(1) < r^*$, that is $d(p_{k+1}, q^*) + r_{k+1} < r^*$.

Case 2 If $d(p_0, q^*) = r^*$, then, arguing by induction, Lemmas 4.3 and 4.4 show that for all k, $d(p_k, q^*) + r_k = r^*$.

Therefore in any case, the following inequality holds

$$d(p_k, q^*) + r_k \leq r^*, \forall k \geq 0.$$

Since $r_k \to r^*$ and $p_k \to p^*$, we have $q^* = p^*$. In other words, p^* is the unique singularity of X on $\overline{B}(p_0, r^*)$.

The proof is complete if $R = r^*$. Otherwise, $\phi(R) < 0$ and we take $q^* \in \overline{B}(p_0, R)$, a singularity of X, i.e. $X(q^*) = 0$. Denote by $\gamma : [0, 1] \to M$ a minimizing geodesic joining p_0 to q^*, and $v = \dot{\gamma}(0)$. Since $P_{\gamma,1,0}X(q^*) = 0$, we have $|X'(p_0)^{-1}X(p_0) + v|_{p_0} = |X'(p_0)^{-1}[P_{\gamma,1,0}X(q^*) - X(p_0) - X'(p_0)v]|_{p_0}$. The left hand side term is greater than $|v|_{p_0} - \phi(0)$, and the right one is smaller than $\phi(|v|_{p_0}) - \phi(0) + |v|_{p_0}$. So $\phi(|v|_{p_0}) \geq 0$. As $|v|_{p_0} \leq R$, necessarily $|v|_{p_0} \leq r^*$. This means $q^* \in \overline{B}(p_0, r^*)$ and then $q^* = p^*$.

4.3 Estimates

4.3.1 Proof of (iv)

The estimates in Theorem 3.1(iv) can be obtained by a simple adaptation of some arguments of [23]. First, given $k \geq 0$, notice that similar computations to those in the proof of Lemma 4.3 yield

$$|X'(p_0)^{-1}P_{c,2,0}X(p_{k+1})|_{p_0} \leq |X'(p_k)^{-1}X(p_k)|_{p_k} \int_0^1 [\phi'(\int_0^{1+t}|\dot{c}|) - \phi'(\int_0^1 |\dot{c}|)]dt,$$

where $c : [0, 2] \to M$ is the concatenation of a minimizing geodesic curve joining p_0 to p_k with $\gamma(t) = \exp_{p_k}[-tX'(p_k)^{-1}X(p_k)]$, $t \in [0, 1]$. Therefore $\int_0^1 |\dot{c}| = d(p_k, p_0)$ and $\int_0^{1+t}|\dot{c}| = d(p_k, p_0) + t\beta_k$ for $\beta_k = |X'(p_k)^{-1}X(p_k)|_{p_k}$. Hence

$$
\begin{aligned}
|X'(p_0)^{-1}P_{c,2,0}X(p_k)|_{p_0} &\leq \beta_k \int_0^1 [\phi'(d(p_k, p_0) + t\beta_k) - \phi'(d(p_{k-1}, p_0))] \, dt \\
&\leq \int_0^1 \int_0^{t\beta_k} \ell(d(p_k, p_0) + r) \, dr \, \beta_k dt \\
&= \int_0^{\beta_k} \ell(d(p_k, p_0) + r)(\beta_k - r) \, dr.
\end{aligned}
$$

Next, following the proof of [23, Proposition 3.2], let us consider the auxiliary function $\psi_a(s) = \frac{1}{s^2}\int_0^s \ell(a+r)(s-r)dr$ for $s \in [0, R-a]$. It is direct to verify that ψ_a is nondecreasing. Since $\beta_k \leq r_{k+1} - r_k$ and $d(p_k, p_0) \leq r_k$ by (9), we deduce that

$$|X'(p_0)^{-1}P_{c,2,0}X(p_k)|_{p_0} \leq \beta_k{}^2\psi_{d(p_k,p_0)}(r_{k+1}-r_k) \leq \beta_k{}^2\psi_{r_k}(r_{k+1}-r_k).$$

But

$$\phi(r_{k+1}) = \beta - r_{k+1} + \int_0^{r_{k+1}}\ell(s)(r_{k+1}-s)ds = \int_{r_k}^{r_{k+1}}\ell(s)(r_{k+1}-s)ds$$
$$= \int_0^{r_{k+1}-r_k}\ell(r_k+r)(r_{k+1}-r_k-r)dr = (r_{k+1}-r_k)^2\psi_{r_k}(r_{k+1}-r_k).$$

Thus

$$|X'(p_0)^{-1}P_{c,2,0}X(p_{k+1})|_{p_0} \leq \phi(r_{k+1})(\beta_k/(r_{k+1}-r_k))^2.$$

Finally, by Lemma 4.2 we obtain

$$
\begin{aligned}
|X'(p_{k+1})^{-1}X(p_{k+1})|_{p_{k+1}} &\leq \|[\mathbb{P}_{c,2,0}X'(p_{k+1})^{-1}]X'(p_0)\|_{p_0}|X'(p_0)^{-1}P_{c,2,0}X(p_{k+1})|_{p_0} \\
&\leq -\phi'(d(p_k,p_0)+\beta_k)^{-1}\phi(r_{k+1})(\beta_k/(r_{k+1}-r_k))^2 \\
&\leq -\phi'(r_{k+1})^{-1}\phi(r_{k+1})(\beta_k/(r_{k+1}-r_k))^2,
\end{aligned}
$$

which proves (11). This implies for all $k \geq k_0 \geq 0$ and $n \geq 0$, $d(p_{k+n+1}, p_{k+n}) \leq |X'(p_{k+n})^{-1}X(p_{k+n})|_{p_{k+n}} \leq (r_{k+n+1}-r_{k+n})(\beta_{k_0}/(r_{k_0+1}-r_{k_0}))^{2^{k-k_0}}$. Summing for all $n \geq 0$, we obtain (12).

4.4   Proof of (v)

Let us proceed with the proof of the last part of Theorem 3.1. Define $Q : [0, R] \to \mathbb{R}$ as the following quadratic polynomial function

$$Q(r) = \phi'(r^*)(r - r^*) + \tfrac{1}{2}\phi''(r^*)(r - r^*)^2, \tag{23}$$

whose smallest root is r^* (recall that $\phi''(r^*) = \ell(r^*) > 0$ and $\phi'(r^*) \leq 0$). Newton's method for solving the equation $Q(r) = 0$, starting at $\rho_0 := 0$, generates the sequence

$$\rho_{k+1} = \rho_k - Q'(\rho_k)^{-1}Q(\rho_k). \tag{24}$$

It is easy to verify that $\{\rho_k\}$ is well defined, monotonically increasing and convergent to r^*. Furthermore, it is well known that in this case the solution of (24) has the closed form

$$\rho_k = \begin{cases} r^*(1 - q^{2^k-1})(1 - q^{2^k})^{-1} & \text{if} \quad q \in (0,1), \\ r^*(1 - 2^{-k}) & \text{if} \quad q = 1, \end{cases}$$

where q is given by (13); see, for instance, [10, 16, 23, 25]. Therefore, Theorem 3.1(vi) is a direct consequence of the following result.

Lemma 4.6 *Let $\{r_k\}$ and $\{\rho_k\}$ be respectively defined by (7) and (24). Then for every $k \geq 0$, $r_k \geq \rho_k$.*

Proof. We argue by induction. The property is immediate for $k = 0$ because $r_0 = \rho_0 = 0$. Now, assume that $r_k \geq \rho_k$ for some $k \geq 0$. By using (24) one has

$$
\begin{aligned}
r_{k+1} - \rho_{k+1} &= r_{k+1} - r_k + \rho_k - \rho_{k+1} + r_k - \rho_k \\
&= -\phi'(r_k)^{-1}\phi(r_k) + Q'(\rho_k)^{-1}Q(\rho_k) + r_k - \rho_k \\
&= Q'(\rho_k)^{-1}[Q(\rho_k) + Q'(\rho_k)(r_k - \rho_k)] - \phi'(r_k)^{-1}\phi(r_k).
\end{aligned}
$$

By convexity, $Q(\rho_k) + Q'(\rho_k)(r_k - \rho_k) \leq Q(r_k)$, and $Q'(\rho_k) \leq Q'(r_k)$ because $\rho_k \leq r_k < r^*$. Since Q' is negative on $[0, r^*]$, we deduce that

$$
r_{k+1} - \rho_{k+1} \geq Q'(r_k)^{-1}Q(r_k) - \phi'(r_k)^{-1}\phi(r_k).
$$

As $\phi'' = \ell$ is nondecreasing on $[0, R]$, the function ϕ' is convex and thus

$$
\phi(r_k) = -\int_{r_k}^{r^*} \phi'(t)dt \geq -\frac{\phi'(r_k) + \phi'(r^*)}{2}(r^* - r_k).
$$

It ensues that

$$
-\phi'(r_k)^{-1}\phi(r_k) \geq \frac{1}{2}(r^* - r_k)[1 + \phi'(r_k)^{-1}\phi'(r^*)].
$$

On the other hand, a straightforward computation gives $Q'(r_k)^{-1}Q(r_k) = \frac{1}{2}(r^* - r_k)[-1 - Q'(r_k)^{-1}\phi'(r^*)].$] By convexity of ϕ', we have $Q'(r_k) = \phi'(r^*) + \phi''(r^*)(r_k - r^*) \leq \phi'(r_k) < 0$. As $\phi'(r^*) \leq 0$, we get

$$
Q'(r_k)^{-1}Q(r_k) \geq -\frac{1}{2}(r^* - r_k)[1 + \phi'(r_k)^{-1}\phi'(r^*).]
$$

Therefore, we obtain

$$
r_{k+1} - \rho_{k+1} \geq Q'(r_k)^{-1}Q(r_k) - \phi'(r_k)^{-1}\phi(r_k) \geq 0,
$$

which proves the result. $\qquad\qquad\square$

5 Special cases

5.1 Riemannian Kantorovich's theorem

The well known Kantorovich's theorem for Newton's method in Banach spaces [12] gives a criterion for (quadratic) convergence which is verifiable at the starting point provided a local Lipschitz constant is known. An extension of Kantorovich's theorem to finite-dimensional and complete Riemannian manifolds has been obtained by Ferreira and Svaiter in [9]. We will see that the latter can be viewed as a special case of Theorem 3.1. To this end, let us introduce the following definition of Lipschitz continuity for tensors (for a related notion see [9, Definition 2.2]).

Definition 5.1 *A (1,k)-tensor T on M is said to be L-Lipschitz continuous on a subset S of M, if for all geodesic curve $\gamma : [0,1] \to M$ with endpoints in S, we have*

$$\|\mathbb{P}_{\gamma,1,0}T(\gamma(1)) - T(\gamma(0))\|_{\gamma(0)} \leq L \int_0^1 |\dot{\gamma}(t)|dt.$$

Theorem 5.1 (R-Kantorovich's theorem) *Given $p_0 \in M$ such that $X'(p_0) \in GL(T_{p_0}M)$, set $a = \|X'(p_0)^{-1}\|_{p_0}$ and $\beta = |X'(p_0)^{-1}X(p_0)|_{p_0}$. Assume that $2a\beta L \leq 1$ for a constant L such that X' is L-Lipschitz continuous on the closed ball $\overline{B}(p_0, R_0)$ for some $R_0 \geq r^*$ where*

$$r^* = \frac{1 - \sqrt{1 - 2a\beta L}}{aL}.$$

Then the sequence $\{p_k\}$ generated by Newton's method starting at p_0 is well defined and convergent to a singularity p^* of X.

If $2a\beta L = 1$, i.e. $r^* = 2\beta$, then p^* is the unique singularity of X in $\overline{B}(p_0, 2\beta)$.

If $2a\beta L < 1$ then $X'(p^*) \in GL(T_{p^*}M)$ and p^* is the unique singularity of X in $\overline{B}(p_0, R)$ for any $R \in [r^*, R_0]$ such that $R < \frac{1 + \sqrt{1 - 2a\beta L}}{aL}$.

In any case $d(p^*, p_0) \leq r^* \leq 2\beta$, and for all $k \in \mathbb{N}$, one has $d(p^*, p_k) \leq r^* - r_k = r^*\frac{1-q}{1-q^{2^k}}q^{2^k-1}$, where q is given by (13) for $\lambda = a\beta L$, and $\{r_k\}$ is the sequence generated by Newton's method, starting at $r_0 = 0$, applied to the scalar function

$$\phi(r) = \beta - r + \frac{aL}{2}r^2. \tag{25}$$

Proof. Assume without loss of generality that $\beta > 0$. Take $c \in \mathcal{G}_2(p_0, r)$. Take (t, y) and (t', y') in c with $t \leq t'$. We have

$$\begin{aligned}
\|X'(p_0)^{-1}[\mathbb{P}_{c,t',0}X'(y') - \mathbb{P}_{c,t,0}X'(y)]\|_{p_0} &\leq \|X'(p_0)^{-1}\|_{p_0}\|\mathbb{P}_{c,t',t}X'(y') - X'(y)\|_y \\
&\leq aL \int_t^{t'} |\dot{c}|,
\end{aligned}$$

where we have used the isometry property of the parallel transport and Definition 5.1. Define $\ell(r) = aL$, if $r \in [0, R]$, and let us verify that $\phi := \phi_\ell$ complies with the requirements of Theorem 3.1. Indeed, in this case ϕ is given by (25) and we have that $\phi(R) \leq 0$. The roots of ϕ are $r^{\pm} = \frac{1 \pm \sqrt{1 - 2a\beta L}}{aL}$. We have $r^- \leq 2\beta \leq \frac{1}{aL} \leq r^+$ with equality iff $2a\beta L = 1$. In any case, $r^* = r^-$ is the unique root of ϕ on $[0, R]$, with $\phi'(r^*) = -1 + aLr^*$, which is negative when $2a\beta L < 1$. Finally, straightforward computations in (14) yield $\lambda = a\beta L$. The result follows thus by a direct application of Theorem 3.1. $\qquad\square$

5.2 An Euclidean case: Nesterov-Nemirovskii self-concordancy

In their pioneering work [14], Nesterov and Nemirovskii developed a general theory of the computational complexity of interior-point methods for convex optimization, based on the

notion of self-concordant functions. See [17] for a simplified yet comprehensive presentation of this theory. Inspired by the analysis in [4], we next show that Theorem 3.1 yields a local convergence result of Newton's method for the minimization of self-concordant functions which is a slight variant of a key result in [14].

Suppose that $M = \mathbb{R}^n$ with the usual identification $T_x\mathbb{R}^n \cong \mathbb{R}^n$. Let us denote by $\langle \cdot, \cdot \rangle$ the Euclidean product in $\mathbb{R}^n$. Recall that a convex function $f \in C^3(\Omega; \mathbb{R})$ defined on an nonempty, open and convex set $\Omega \subset \mathbb{R}^n$ is said to be a-self-concordant for $a > 0$ if

$$\forall x \in \Omega, \ \forall h \in \mathbb{R}^n, \ |f'''(x)[h, h, h]| \leq 2a^{-1/2}(f''(x)[h, h])^{3/2}. \tag{26}$$

From now on, following [14], we assume that $f : \Omega \to \mathbb{R}$ is a strong, non-degenerate and a-self-concordant function, $i.e.$,

1. $f(y_k) \to \infty$ whenever $\{y_k\}$ converges to a point in the boundary $\partial\Omega$.

2. $H_x := f''(x)$ is positive definite for all $x \in \Omega$.

3. f satisfies (26).

In particular, f is strictly convex on Ω and $x^* \in \Omega$ is the unique minimizer of f on Ω iff $f'(x^*) = 0$. Now, set

$$\langle h_1, h_2 \rangle_{H_x} = a^{-1} f''(x)[h_1, h_2]$$

and

$$|h|_{H_x} = \sqrt{\langle h, h \rangle_{H_x}}.$$

The Dikin ellipsoid of center $x \in \Omega$ and radius $r > 0$ is defined by

$$D_r(x) = \{y \in \mathbb{R}^n \mid |y - x|_{H_x} \leq r\}.$$

It corresponds to the closed ball of center x and radius r when $\mathbb{R}^n$ is endowed with the metric structure induced by the scalar product $\langle \cdot, \cdot \rangle_{H_x}$. By [14, Theorem 1.1.1] (see also [17, Theorem 2.5.18]), if $r < 1$ then $D_r(x) \subset \Omega$ and for any $y \in D_r(x)$ we have

$$\forall h \in \mathbb{R}^n, \ |h|_{H_y} \leq \frac{1}{1 - r}|h|_{H_x}. \tag{27}$$

Lemma 5.1 *Fix $x_0 \in \Omega$ and let $\mathbb{R}^n$ be endowed with the metric structure induced by $\langle \cdot, \cdot \rangle_{H_{x_0}}$. Then for any $R \in (0, 1)$, $X = f'$ satisfies (5) with $\ell : [0, R] \to [0, +\infty)$ being given by*

$$\ell(r) = \frac{2}{(1 - r)^3}.$$

Proof. It suffices to show that $\|f''(x_0)^{-1}f'''(y)\|_{H_{x_0}} \leq \ell(r)$ for $r \in [0, R]$ and $y \in \overline{B}(x_0, r) = D_r(x_0)$ (see Remark 3.1). Using the notation $\langle f'''(y)h_1h_2, h_3 \rangle = f'''(y)[h_1, h_2, h_3]$, by definition we have

$$
\begin{aligned}
\|f''(x_0)^{-1}f'''(y)\|_{H_{x_0}} &= \sup\{|f''(x_0)^{-1}f'''(y)h_1h_2|_{H_{x_0}} \mid |h_i|_{H_{x_0}} \leq 1, i = 1, 2\} \\
&= \sup\{\langle f''(x_0)^{-1}f'''(y)h_1h_2, h_3 \rangle_{H_{x_0}} \mid |h_i|_{H_{x_0}} \leq 1, i = 1, 2, 3\} \\
&= a^{-1}\sup\{f'''(y)[h_1, h_2, h_3] \mid |h_i|_{H_{x_0}} \leq 1, i = 1, 2, 3\}.
\end{aligned}
$$

On the other hand, by a general lemma on symmetric trilinear forms [14, Proposition 8.1.1] (see [11] for an alternative proof of that lemma), it follows from the self-concordance property (26) that $f'''(y)[h_1, h_2, h_3] \leq 2a|h_1|_{H_y}|h_2|_{H_y}|h_3|_{H_y}$. Hence,

$$
\|f''(x_0)^{-1}f'''(y)\|_{H_{x_0}} \leq 2\sup\left\{|h_1|_{H_y}|h_2|_{H_y}|h_3|_{H_y} \mid |h_i|_{H_{x_0}} \leq 1, i = 1, 2, 3\right\} \leq \frac{2}{(1-r)^3},
$$

where in the last inequality we have used (27). This proves our claim. $\qquad\square$

Theorem 5.2 (Self-concordant minimization) *Let $x_0 \in \Omega$ and $\beta = |f''(x_0)^{-1}f'(x_0)|_{H_{x_0}}$. If $\beta \leq (\sqrt{2} - 1)^2 = 3 - 2\sqrt{2} \approx 0.17157$ then f admits a unique minimizer x^* which belongs to $D_{r^*}(x_0)$ for*

$$
r^* = \frac{1}{4}[\beta + 1 - \sqrt{\beta^2 - 6\beta + 1}],
$$

and in particular $|x^* - x_0|_{H_{x_0}} \leq r^* \leq 2\beta$. The sequence generated by Newton's method $x_{k+1} = x_k - f''(x_k)^{-1}f'(x_k)$ starting at x_0 is well defined, contained in $D_{r^*}(x_0)$ and convergent to x^*.

For all $k \geq 0$, $|x^* - x_k|_{H_{x_0}} \leq r^* - r_k$ where $\{r_k\}$ is the sequence generated by Newton's method, starting at $r_0 = 0$, applied to the scalar function

$$
\phi(r) = \beta - 2r + \frac{r}{1-r}. \tag{28}
$$

The sequence $\{r_k\}$ converges to r^* which is the smallest zero of ϕ in $[0, 1)$. Furthermore,

$$
r_k = \frac{1 - \nu^{2^k - 1}}{1 - \nu^{2^k - 1}\eta}r^*, \tag{29}
$$

where

$$
\nu = \frac{1 - \beta - \sqrt{\beta^2 - 6\beta + 1}}{1 - \beta + \sqrt{\beta^2 - 6\beta + 1}}, \quad \eta = \frac{1 + \beta - \sqrt{\beta^2 - 6\beta + 1}}{1 + \beta + \sqrt{\beta^2 - 6\beta + 1}}.
$$

Proof. Motivated by Lemma 5.1, consider $\ell(r) = 2/(1-r)^3$, $r \in [0, R]$ for some $R < 1$. A direct computation shows that the corresponding function ϕ defined by (6) is given by (28). Finding a zero of this function amounts to solving $2r^2 - (\beta + 1)r + \beta = 0$. This equation has a real root iff $\Delta := \beta^2 - 6\beta + 1 = (\beta - 3 + 2\sqrt{2})(\beta - 3 - 2\sqrt{2}) \geq 0$, which is the case because $\beta \leq (\sqrt{2} - 1)^2 = 3 - 2\sqrt{2}$. The roots are then $r^\pm = [1 + \beta \pm \sqrt{\Delta}]/4$, and we choose

$r^* = r^-$, the smallest one. When $\beta > 0$ (otherwise x_0 is the solution), it is clear that $r^* > 0$. Furthermore, it is direct to verify that $r^* \leq 1 - 1/\sqrt{2}$ and $\phi(1 - 1/\sqrt{2}) \leq 0$ with equalities iff $\beta = (\sqrt{2} - 1)^2$. Hence, it is possible to choose $R \in [r^*, 1 - 1/\sqrt{2}]$ such that the assumptions of Theorem 3.1 hold.

Notice that, as $\phi'(r) = -(1 - 4r + 2r^2)/(r-1)^2$, we have that $\phi'(r^*) < 0$ if $r^* < 1 - 1/\sqrt{2}$, which amounts to $\beta < (\sqrt{2} - 1)^2$ and $\Delta > 0$, hence $\nu < 1$ and we can ensure that quadratic convergence occurs in that case.

This completes the proof because, in $\mathbb{R}^n$ viewed as an Euclidean space, Newton's iterates does not depend on the choice of the inner product. $\qquad\square$

Remark 5.1 The closed form (29) for r_k is given in [23], where a majorant function analogue to (28) is used to improve the original α-theorem of Smale. See also Theorem 5.3 below.

Remark 5.2 The convergence test provided by the constant β in Theorem 5.2 is not optimal: it can be refined by using the fact that ∇f is not only a vector field but also a gradient, see for instance [14, 17]. On the other hand it is worthwhile to note that the rate of convergence (29) slightly improves classical results on this topic.

Remark 5.3 It would be interesting to develop a theory of self-concordancy for real-valued functions defined on Riemannian manifolds. This is beyond the scope of this paper.

5.3 Riemannian Smale's α-theorem

Another celebrated local convergence result on Newton's method is the so called α-theorem of Smale [18] for analytic functions, which is a very useful tool for the construction and computation complexity analysis of homotopy algorithms for solving nonlinear equations [3]. An extension of Smale's α-theorem to Riemannian manifolds was established by Dedieu, Priouret and Malajovich [5]. As a special case of Theorem 3.1, we will obtain a Riemannian version of Smale's theorem which improves [5, Theorem 1.4] (see Remark 5.4).

From now on, let us assume that both the complete Riemannian manifold (M, g) and the vector field X under consideration are analytic. Following [18], we set

$$\gamma(p) = \sup_{k \geq 2} \left\| \frac{1}{k!} X'(p)^{-1} X^{(k)}(p) \right\|_p^{\frac{1}{k-1}}, \tag{30}$$

for each point $p \in M$ such that $X'(p) \in GL(T_pM)$, and $\gamma(p) = +\infty$ otherwise. By analyticity of X, $\gamma(p) < +\infty$ whenever $X'(p)$ is invertible and the following Taylor formula holds:

$$\forall u \in T_pM, \ |u|_p < \gamma(p)^{-1} \ \Rightarrow \ \mathbb{P}_{\zeta,1,0} X^{(k)}(\zeta(1)) = \sum_{j=0}^{+\infty} \frac{1}{j!} X^{(j+k)}(p)[u]^j, \tag{31}$$

where $\zeta(t) = \exp_p[tu]$. Here, we adopt the convention $0^{-1} = +\infty$.

Lemma 5.2 *Let $p_0 \in M$ be such that $X'(p_0) \in GL(T_{p_0}M)$, $r > 0$ and $p \in \overline{B}(p_0, r)$.*
(i) If $r < \gamma(p_0)^{-1}$ then for all $k \geq 2$,

$$\|X'(p_0)^{-1}[\mathbb{P}_{\zeta,1,0}X^{(k)}(p)]\|_{p_0} \leq \frac{k!\gamma(p_0)^{k-1}}{(1-\gamma(p_0)r)^{k+1}}, \tag{32}$$

where $\zeta : [0,1] \to M$ is the minimizing geodesic curve joining p_0 and p.
(ii) If $r < (1 - 1/\sqrt{2})\gamma(p_0)^{-1}$ then $X'(p) \in GL(T_pM)$ and

$$\gamma(p) \leq \frac{1}{(1-\gamma(p_0)r)\psi(\gamma(p_0)r)}\gamma(p_0) \tag{33}$$

for $\psi(\alpha) = 1 - 4\alpha + 2\alpha^2 = (\alpha - 1 + 1/\sqrt{2})(\alpha - 1 - 1/\sqrt{2})$.

Proof. (i) Let us write $\zeta(t) = \exp_{p_0}[tu]$, $t \in [0,1]$, with $u \in T_{p_0}M$ and $p = \exp_{p_0}[u]$ so that $|u|_{p_0} \leq r$. We obtain from (31) that

$$
\begin{aligned}
\|X'(p_0)^{-1}[\mathbb{P}_{\zeta,1,0}X^{(k)}(p)]\|_{p_0} &\leq \sum_{j=0}^{+\infty} \frac{1}{j!}\|X'(p_0)^{-1}X^{(j+k)}(p_0)[u]^j\|_{p_0} \leq \sum_{j=0}^{+\infty} \frac{(j+k)!}{j!}\gamma(p_0)^{j+k-1}|u|_{p_0}^j \\
&\leq k!\gamma(p_0)^{k-1}\sum_{j=0}^{+\infty} \frac{(j+k)!}{j!k!}(\gamma(p_0)r)^j = \frac{k!\gamma(p_0)^{k-1}}{(1-\gamma(p_0)r)^{k+1}}.
\end{aligned}
$$

(ii) Denoting by $I : T_{p_0}M \to T_{p_0}M$ the identity operator on $T_{p_0}M$, we get

$$
\begin{aligned}
\|X'(p_0)^{-1}[\mathbb{P}_{\zeta,1,0}X'(p)] - I\|_{p_0} &\leq \sum_{j=0}^{+\infty} \frac{1}{j!}\|X'(p_0)^{-1}X^{(1+j)}(p_0)\|_{p_0}r^j - 1 \\
&\leq \sum_{j=0}^{+\infty}(1+j)(\gamma(p_0)r)^j - 1 = \frac{1}{(1-\gamma(p_0)r)^2} - 1.
\end{aligned}
$$

As $\gamma(p_0)r < 1 - 1/\sqrt{2}$, we have that $1/(1-\gamma(p_0)r)^2 - 1 < 1$. We deduce that $X'(p_0)^{-1}\mathbb{P}_{\zeta,t,0}X'(p)$ is invertible, hence $X'(p) \in GL(T_pM)$, and

$$\|[\mathbb{P}_{\zeta,1,0}X'(p)^{-1}]X'(p_0)\|_{p_0} \leq \frac{(1-\gamma(p_0)r)^2}{2(1-\gamma(p_0)r)^2 - 1} = \frac{(1-\gamma(p_0)r)^2}{\psi(\gamma(p_0)r)}.$$

Therefore

$$
\begin{aligned}
\|X'(p)^{-1}X^{(k)}(p)\|_p &= \|X'(p)^{-1}P_{\zeta,0,1}X'(p_0)X'(p_0)^{-1}P_{\zeta,1,0}X^{(k)}(p)\|_p \\
&\leq \|[\mathbb{P}_{\zeta,1,0}X'(p)]^{-1}X'(p_0)\|_{p_0}\|X'(p_0)^{-1}\mathbb{P}_{\zeta,1,0}X^{(k)}(p)\|_{p_0} \\
&\leq \frac{k!\gamma(p_0)^{k-1}}{(1-\gamma(p_0)r)^{k-1}\psi(\gamma(p_0)r)}.
\end{aligned}
$$

From the fact that $0 < \psi(\alpha) \le 1$ when $0 \le \alpha < 1 - 1/\sqrt{2}$, it follows that

$$\gamma(p) \le \sup_{k \ge 2} \frac{\gamma(p_0)}{(1 - \gamma(p_0)r)\psi(\gamma(p_0)r)^{\frac{1}{k-1}}} = \frac{\gamma(p_0)}{(1 - \gamma(p_0)r)\psi(\gamma(p_0)r)},$$

which proves (33). $\qquad\square$

Lemma 5.3 *Let $p_0 \in M$ be such that $X'(p_0) \in GL(T_{p_0}M)$. For any $R \le (1-1/\sqrt{2})\gamma(p_0)^{-1}$, X satisfies* (5) *with*

$$\ell(r) = \frac{2\gamma(p_0)}{(1 - \gamma(p_0)r)^3}, \quad r \in [0, R].$$

Proof. To prove that ℓ complies with (5) it suffices, by Remark 3.1, to establish that for every $r \in [0, R]$ and $c \in \mathcal{G}_2(p_0, r)$ we have that

$$||X'(p_0)^{-1}[\mathbb{P}_{c,a,0}X''(c(a))]|| \le \ell(r), \quad \forall a \in [0, 2]. \tag{34}$$

By continuity, it suffices to establish this property for $r \in [0, R)$.

Fix $c \in \mathcal{G}_2(p_0, r)$ with $r < R$. By definition of $\mathcal{G}_2(p_0, r)$, c is the concatenation of two geodesics $\zeta_i : [0, 1] \to M$ $(i = 0, 1)$ with ζ_0 being minimizing. Let us write $c(t) = \zeta_0(t)$ if $t \in [0, 1]$ and $c(t) = \zeta_1(t - 1)$ if $t \in [1, 2]$. Since $\zeta_0(0) = c(0) = p_0$, applying Lemma 5.2(i) to ζ_0 with $k = 2$, we deduce that (34) holds for all $a \in [0, 1]$.

In order to prove (34) for $a \in]1, 2]$, we argue as follows. First, consider $a \in (1, 2]$ such that $d(c(1), c(a)) < \gamma(c(1))^{-1}$. Since c on $[1, 2]$ is a geodesic curve, we may write $c(a) = \exp_{c(1)}[u_1]$ for some $u_1 \in T_{c(1)}M$ with $|u_1|_{c(1)} < \gamma(c(1))^{-1}$. Notice that $|u_0|_{p_0} + |u_1|_{c(1)} = d(p_0, c(1)) + d(c(1), c(a)) \le r$. By Taylor's formula (31), we have $\mathbb{P}_{c,a,1}X''(c(a)) = \sum_{j=0}^{+\infty} \frac{1}{j!}X^{(j+2)}(c(1))[u_1]^j$, hence

$$\mathbb{P}_{c,a,0}X''(c(a)) = \sum_{j=0}^{+\infty} \frac{1}{j!}[\mathbb{P}_{c,1,0}X^{(j+2)}(c(1))][(P_{c,1,0}u_1)]^j.$$

Since $c(t) = \zeta_0(t) = \exp_{p_0}[tu_0]$, $t \in [0, 1]$, for some $u_0 \in T_{p_0}M$ with $|u_0|_{p_0} = d(p_0, c(1)) \le r < \gamma(p_0)^{-1}$, we have $c(1) = \exp_{p_0}[u_0]$ and we can use (31) again to get

$$\mathbb{P}_{c,a,0}X''(c(a)) = \sum_{j_0,j_1=0}^{+\infty} \frac{1}{j_0!j_1!}X^{(j_0+j_1+2)}(p_0)[u_0^{j_0}, (P_{c,1,0}u_1)^{j_1}].$$

Therefore

$$
\begin{aligned}
\| X'(p_0)^{-1}[\mathbb{P}_{c,a,0}X''(c(a))] \|_{p_0} &\leq \sum_{j_0,j_1=0}^{+\infty} \frac{1}{j_0!j_1!} \| X'(p_0)^{-1}X^{(j_0+j_1+2)}(p_0)\|_{p_0} |u_0|_{p_0}^{j_0} |P_{c,1,0}u_1|_{p_0}^{j_1} \\
&\leq \sum_{j_0,j_1=0}^{+\infty} \frac{(j_0+j_1+2)!}{j_0!j_1!} \gamma(p_0)^{j_0+j_1+1} |u_0|_{p_0}^{j_0} |u_1|_{c(1)}^{j_1} \\
&= \sum_{k=0}^{+\infty} \Big(\sum_{j=0}^{k} \frac{(k+2)!}{j!(k-j)!} |u_0|_{p_0}^{j} |u_1|_{c(1)}^{k-j} \Big) \gamma(p_0)^{k+1} \\
&= \sum_{k=0}^{+\infty} (k+2)(k+1)\Big(|u_0|_{p_0} + |u_1|_{c(1)} \Big)^{k} \gamma(p_0)^{k+1} \\
&\leq \sum_{k=0}^{+\infty} (k+2)(k+1)(r\gamma(p_0))^{k}\gamma(p_0) = \frac{2\gamma(p_0)}{(1-r\gamma(p_0))^3}.
\end{aligned}
$$

The previous argument shows that (34) holds for every $a \in (1,2]$ with $d(c(1),c(a)) < \gamma(c(1))^{-1}$. If $d(c(1),c(a)) \geq \gamma(c(1))^{-1}$ then one does not know whether a direct Taylor's expansion at $c(1)$ is valid or not. Nevertheless, as $r < R \leq (1-1/\sqrt{2})\gamma(p_0)^{-1}$, by virtue of (33), there exists a positive constant K depending on r and $\gamma(p_0)$ such that $\gamma(p) \leq K$ for all $p \in \overline{B}(p_0,r)$. Consequently, there exist a finite subdivision $1 = t_1 < \ldots < t_{n+1} = a$ of $[1,a]$ and corresponding tangent vectors $u_i \in T_{c(t_i)}M$ such that $c(t_{i+1}) = \exp_{c(t_i)}[u_i]$ and $|u_i|_{c(t_i)} = d(c(t_{i+1}),c(t_i)) \leq K^{-1} \leq \gamma(c(t_i))^{-1}$ for $i = 1,\ldots,n$. In such a case, successive applications of appropriate parallel transports and Taylor's formulas yield

$$
\mathbb{P}_{c,a,0}X''(c(a)) = \sum_{j_0,\ldots,j_n=0}^{+\infty} \frac{1}{j_0!\ldots j_n!} X^{(j_0+\ldots+j_n+2)}(p_0)[u_0^{j_0}, (P_{c,t_1,0}u_1)^{j_1}\ldots, (P_{c,t_n,0}u_n)^{j_n}].
$$

Then, since $|u_0|_{p_0} + \sum_{i=1}^{n} |u_i|_{c(t_i)} \leq r$, we can use similar arguments to show that (34) holds. We leave the details to the reader. $\qquad\square$

Theorem 5.3 (Riemannian α-theorem) *Let $p_0 \in M$ be such that $X'(p_0) \in GL(T_{p_0}M)$, and set $\alpha = \beta\gamma$ for $\beta = |X'(p_0)^{-1}X(p_0)|_{p_0}$ and $\gamma = \gamma(p_0)$.*
If $\alpha \leq \alpha_0 := (\sqrt{2}-1)^2 = 3 - 2\sqrt{2}$ then the sequence $\{p_k\}$ generated by Newton's method starting at p_0 is well defined, contained in $\overline{B}(p_0,r^*)$ where

$$
r^* = \frac{1}{4\gamma}[1 + \alpha - \sqrt{\alpha^2 - 6\alpha + 1}],
$$

and convergent to some p^*, which is the unique singularity of X on $\overline{B}(p_0,(1-1/\sqrt{2})\gamma^{-1})$. In particular, $d(p^*,p_0) \leq r^* \leq 2\beta$.
For all $k \geq 0$, $d(p^*,p_k) \leq r^* - r_k$ where $\{r_k\}$ is the sequence generated by Newton's method, starting at $r_0 = 0$, applied to the scalar function

$$
\phi(r) = \beta - 2r + \frac{r}{1 - \gamma r}. \tag{35}
$$

The sequence $\{r_k\}$ converges to r^* which is the smallest zero of ϕ in $[0, \gamma^{-1})$. Furthermore, r_k has the closed form

$$r_k = \frac{1 - \nu^{2^k - 1}}{1 - \nu^{2^k - 1}\eta} r^*,$$

for

$$\nu = \frac{1 - \alpha - \sqrt{\alpha^2 - 6\alpha + 1}}{1 - \alpha + \sqrt{\alpha^2 - 6\alpha + 1}}, \quad \eta = \frac{1 + \alpha - \sqrt{\alpha^2 - 6\alpha + 1}}{1 + \alpha + \sqrt{\alpha^2 - 6\alpha + 1}}.$$

Proof. By Lemma 5.3, one can use the same arguments, up to the factor γ, of the proof of Theorem 5.2. We leave the details to the reader. $\qquad\square$

Remark 5.4 Theorem 5.3 is a finite-dimensional Riemannian version of some results of [22, 23] and improves two aspects of the R-α-theorem proved by Dedieu *et al.* in [5]. First, the constant $\alpha_0 = 3 - 2\sqrt{2}$ in Theorem 5.3, which is the same of [22, 23], is better than the analogue constant 0.130716944... in [5, 18] characterized as the unique root of the equation $2\alpha = \psi(\alpha)^2$ in $[0, 1 - 1/\sqrt{2})$. Second, in our approach there is no need of any condition relying on the injectivity radius $\mathbf{r}_{p_0}$ of the exponential map at p_0, while in [5, Theorem 1.4] it is assumed in addition that $\beta \leq s_0 \mathbf{r}_{p_0}$ for a suitable universal constant $s_0 > 0$.

Remark 5.5 We focus on the conditions of the original α-theorem in order to illustrate our approach. However, further improvements of Smale's result relying directly on the quantities $\|X'(p_0)^{-1}X^{(k)}(p_0)\|_{p_0}$, $k \geq 2$, (and not on the upper bound $\gamma(p_0)$) are given in [21, 23] for the Euclidean case. Of course, Riemannian versions of those results can be obtained by specialization of Theorem 3.1; we will not develop this point here.

References

[1] Adler R., Dedieu J.-P., Margulies J., Martens M., Shub M., *Newton method on Riemannian manifolds and a geometric model for the human spine*, IMA J.Numer. Anal., **22**, (2002), pp. 1-32.

[2] Appel J., De Pascale E., Lysenko J.V., Zabrejko P.P., *New results on Newton-Kantorovich approximations with applications to nonlinear integral equations*, Numer. Func. Anal. Optim., **18**, (1997), pp. 1-17.

[3] Blum L., Cucker F., Shub M., Smale S., "Complexity and real computation", Springer-Verlag, New York, 1998.

[4] Cominetti R., *On the convergence of Newton's method*, unpublished technical note.

[5] Dedieu J.-P., Priouret P., Malajovich G., *Newton's method on Riemannian manifolds: covariant alpha theory*, IMA J.Numer. Anal. **23**, (2003), pp. 395-419.

[6] da Cruz Neto J.X., de Lima L.L., Oliveira P.R., *Geodesic algorithms in Riemannian geometry*, Balkan J. Geom. Appl. **3** (1998), pp. 89-100.

[7] Do Carmo M., "Riemannian geometry", Birkhäuser, Boston, 1992.

[8] Edelman A., Arias T.A., Smith T.,*The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix. Anal. Appl. **20** (1998), pp. 303-353.

[9] Ferreira O.P., Svaiter B.F., *Kantorovich's theorem on Newton's method in Riemannian manifolds*, J. of Complexity **18**, (2002), pp. 304-329.

[10] Gragg W.B., Tapia R.A., *Optimal error bounds for the Newton-Kantorovich theorem*, SIAM J. Numer. Anal. 11 (1974), pp. 10-13.

[11] Jarre F., *Interior-point methods for convex programming*, Appl. Math. Optim. 26 (1992), no. 3, pp. 287-311.

[12] Kantorovich L.V., Akilov G.P., "Functional Analysis in Normed Spaces", Pergamon, Oxford, 1964.

[13] Lang S. , "Differential and Riemannian Manifolds", Springer-Verlag, New York, 1995.

[14] Nesterov Y., Nemirovskii A., "Interior-point polynomial algorithms in convex programming", SIAM Studies in Applied Mathematics, 13, Philadelphia, PA, 1994.

[15] Petersen P., "Riemannian geometry", Graduate Texts in Mathematics, 171, Springer-Verlag, New York, 1998.

[16] Pták V., *The rate of convergence of Newton's process*, Numer. Math. 25 (1976), pp. 279-285.

[17] Renegar J., "A mathematical view of interior-point methods in convex optimization", MPS/SIAM Series on Optimization, Philadelphia, PA, 2001.

[18] Smale S., *Newton's method estimates from data at one point*, The Merging of discipilines: New directions in Pure, Applied and Computational Mathematics, R. Ewing, K. Gross & C. Martin Eds., Berlin, Springer, 1986.

[19] Smith S., *Optimization techniques on Riemannian manifolds*, Fields Institute Communications AMS, **3** (1994), pp. 113-146.

[20] Udriste C., "Convex Functions and Optimization Methods on Riemannian Manifolds", Mathematics and Its Applications, Vol. 297, Kluwer Academic Press, Dordrecht, 1994.

[21] Wang D., Zhao F., *The theory of Smale's point estimation and its applications*, in Linear/nonlinear iterative methods and verification of solution (Matsuyama, 1993), J. Comput. Appl. Math. 60 (1995), no. 1-2, pp. 253-269.

[22] Wang X., Han D., *On dominating sequence method in the point estimate and Smale's theorem*, Scientia Sinica Ser. A (1990), pp. 905-913.

[23] Wang X., *Convergence of Newton's method and inverse function theorem in Banach space*, Math. Comp. **68** (1999), pp. 169-186.

[24] Zabrejko P.P., *On the approximate solution of operator solution*, (Russian), in Shilov G.E.: Differentiation of functions: Higher derivatives and variational calculus, Yaroslav. Gosud. Univ., Yaroslavl' 1980, pp. 51-71.

[25] Zabrejko P.P., Nguen D.F., *The majorant method in the theory of Newton-Kantorovich approximations and the Pták error estimates*, Numer. Func. Anal. Optim., **9** (1987), pp. 671-674.