IMA Journal of Numerical Analysis Advance Access published January 30, 2012

IMA Journal of Numerical Analysis Page 1 of 18
doi:10.1093/imanum/drr048

Local convergence of Newton's method under a majorant condition in Riemannian manifolds

ORIZON P. FERREIRA
Instituto de Matemática e Estatística, Universidade Federal de Goiás, Campus II, Caixa Postal 131, CEP 74001-970, Goiânia, GO, Brazil
orizon@mat.ufg.br

AND

ROBERTO C. M. SILVA*
Departamento de Matemática, Instituto de Ciências Exatas, Universidade Federal do Amazonas, Campus Universitário Artur Virgílio Filho, CEP 69077-000, Manaus, AM, Brazil
*Corresponding author: rmesquita@ufam.edu.br robertocristovao@gmail.com

[Received on 1 June 2011; revised on 1 October 2011]

A local convergence analysis of Newton's method for finding a singularity of a differentiable vector field defined on a complete Riemannian manifold, based on the majorant principle, is presented in this paper. This analysis provides a clear relationship between the majorant function, which relaxes the Lipschitz continuity of the derivative, and the vector field under consideration. It also allows us to obtain the optimal convergence radius and the biggest range for the uniqueness of the solution and to unify some previously unrelated results.

Keywords: Newton's method; majorant principle; local convergence; Riemannian manifold.

1. Introduction

Newton's method and its variants are powerful tools for solving nonlinear equations in Banach spaces. Besides its practical applications, Newton's method is also a powerful theoretical tool with a wide range of applications in pure mathematics (see Nash, 1956; Moser, 1961; Wayne, 1996; Blum *et al.*, 1997; Krantz & Parks, 2002). Newton's method has been extended to Riemannian manifolds with many different purposes (see Smith, 1994; Udriste, 1994; Adler *et al.*, 2002; Absil *et al.*, 2008). In particular, it has been also extended to Lie groups (see Owren & Welfert, 2000; Li *et al.*, 2009; Wang & Li, 2011). In the last few years, a couple of papers have dealt with the issue of local and semilocal convergence analysis of Newton's method for finding a singularity of a differentiable vector field defined on a complete Riemannian manifold (see Adler *et al.*, 2002; Ferreira & Svaiter, 2002; Dedieu *et al.*, 2003; Jiang *et al.*, 2004; Li & Wang, 2006, 2008; Alvarez *et al.*, 2008; Wang & Li, 2011).

A semilocal analysis of Kantorovich type in a Riemannian context was given in Ferreira (2009) and a generalization of that result was established in Alvarez *et al.* (2008), Li & Wang (2008) and Wang & Li (2011). Extensions to Riemannian manifolds of Smale's γ-theory and local and semilocal analyses of Newton's method under the γ-condition were given in Dedieu *et al.* (2003), Li & Wang (2006, 2008) and Wang & Li (2006). As far as we know, a local analysis of Newton's method in a Riemannian context under a majorant type condition is new.

In Ferreira (2009) (see also Ferreira, 2011), under a majorant condition, the local convergence, quadratic rate and an estimate of the best possible convergence radius of Newton's method in a linear space were established, as well as uniqueness of the solution for the nonlinear equation in question. Besides improving the convergence theory this analysis permits us to obtain important special cases as an application. It is worth pointing out that the majorant condition used here is equivalent to Wang's condition (see Wang, 2000; Wang & Li, 2003); in the Euclidean context, the derivative of a majorant function is always convex.

The aim of this paper is to present a local convergence analysis of Newton's method for finding a singularity of a differentiable vector field defined on a complete Riemannian manifold under a majorant condition, which extends to Riemannian manifolds the results of Ferreira (2009). In our analysis, the classical Lipschitz condition is relaxed using a majorant function. The analysis presented provides a clear relationship between the majorant function and the vector field under consideration. Also, as in Ferreira (2009), it allows us to obtain the biggest range for the uniqueness of singularity and the optimal convergence radius for the method with respect to the majorant function. Moreover, several unrelated previous results pertaining to Newton's method are unified (see Smale, 1986; Nesterov & Nemirovskii, 1994; Blum *et al.*, 1997; Dedieu *et al.*, 2003; Wang & Li, 2003; Huang, 2004) now in the Riemannian context.

The organization of the paper is as follows. In Section 2, some notation and one basic result used in the paper are presented. In Section 3 the main result is stated and in Section 4 some properties of the majorant function are established and the main relationships between the majorant function and the vector field used in the paper are presented. In Section 5, the uniqueness of the solution and the optimal convergence radius are obtained. In Section 6 the main result is proved and three applications of this result are given in Section 7. Some final remarks are made in Section 8.

2. Notation and auxiliary results

In this section we recall some notation, definitions and basic properties of Riemannian manifolds used throughout the paper. They can be found in many introductory books on Riemannian Geometry, for example, in Do Carmo (1992) and Lang (1995).

Throughout the paper, $\mathscr{M}$ is a smooth manifold and $C^1(\mathscr{M})$ is the class of all continuously differentiable functions on $\mathscr{M}$. The space of vector fields on $\mathscr{M}$ is denoted by $\mathscr{X}(\mathscr{M})$, the tangent space of $\mathscr{M}$ at p by $T_p\mathscr{M}$ and the *tangent bundle* of $\mathscr{M}$ by $T\mathscr{M} = \bigcup_{x\in\mathscr{M}} T_x\mathscr{M}$. Let $\mathscr{M}$ be endowed with a Riemannian metric $\langle\cdot,\cdot\rangle$ and with corresponding norm denoted by $\|\cdot\|$, so that $\mathscr{M}$ is now a *Riemannian manifold*. Let us recall that the metric can be used to define the length of a piecewise C^1 curve $\zeta : [a,b] \to \mathscr{M}$ joining p to q, i.e., such that $\zeta(a) = p$ and $\zeta(b) = q$, by $l(\zeta) = \int_a^b \|\zeta'(t)\|\mathrm{d}t$. Minimizing this length functional over the set of all such curves, we obtain a distance $d(p,q)$, which induces the original topology on $\mathscr{M}$. The open and closed balls of radius $r > 0$ centred at p are defined as

$$B_r(p) := \{q \in \mathscr{M} : d(p,q) < r\} \quad \text{and} \quad \overline{B}_r(p) := \{q \in \mathscr{M} : d(p,q) \leqslant r\}.$$

respectively. Also, the metric induces a map $f \in C^1(\mathscr{M}) \mapsto \operatorname{grad} f \in \mathscr{X}(\mathscr{M})$, which associates to each f its *gradient* via the rule $\langle \operatorname{grad} f, X\rangle = df(X)$ for all $X \in \mathscr{X}(\mathscr{M})$. The chain rule generalizes to this setting in the usual way: $(f\circ\zeta)'(t) = \langle \operatorname{grad} f(\zeta(t)), \zeta'(t)\rangle$ for all curves $\zeta \in C^1$. Let ζ be a curve joining the points p and q in $\mathscr{M}$ and let ∇ be a Levi-Civita connection associated to $(\mathscr{M}, \langle, \rangle)$. For each $t \in [a, b]$, ∇ induces an isometry, relative to $\langle \cdot, \cdot \rangle$,

$$\begin{aligned} P_{\zeta,a,t} \colon T_{\zeta(a)}\mathscr{M} &\longrightarrow T_{\zeta(t)}\mathscr{M} \\ \upsilon &\longmapsto P_{\zeta,a,t}\, \upsilon = V(t), \end{aligned} \tag{2.1}$$

where V is the unique vector field on ζ such that $\nabla_{\zeta'(t)} V = 0$ and $V(a) = \upsilon$, the so-called *parallel translation* along ζ from $\zeta(a)$ to $\zeta(t)$. Note also that

$$P_{\zeta,b_1,b_2} \circ P_{\zeta,a,b_1} = P_{\zeta,a,b_2}, \qquad P_{\zeta,b,a} = P_{\zeta,a,b}{}^{-1}.$$

A vector field V along ζ is said to be *parallel* if $\nabla_{\zeta'} V = 0$. If ζ' itself is parallel, then we say that ζ is a *geodesic*. The geodesic equation $\nabla_{\zeta'}\zeta' = 0$ is a second-order nonlinear ordinary differential equation, so the geodesic ζ is determined by its position p and velocity υ at p. It is easy to check that $\|\zeta'\|$ is constant. We say that ζ is *normalized* if $\|\zeta'\| = 1$. A geodesic $\zeta : [a, b] \to \mathscr{M}$ is said to be *minimal* if its length is equal to the distance between its end points, i.e., $l(\zeta) = d(\zeta(a), \zeta(b))$.

A Riemannian manifold is *complete* if its geodesics are defined for any values of t. The Hopf–Rinow theorem asserts that if this is the case then any pair of points, say p and q, in $\mathscr{M}$ can be joined by a (not necessarily unique) minimal geodesic segment. Moreover, $(\mathscr{M}, d)$ is a complete metric space and bounded and closed subsets are compact.

The *exponential map* at p, $\exp_p : T_p\mathscr{M} \to \mathscr{M}$ is defined by $\exp_p \upsilon = \zeta_\upsilon(1)$, where ζ_υ is the geodesic defined by its position p and velocity υ at p and $\zeta_\upsilon(t) = \exp_p t\upsilon$ for any value of t. For $p \in \mathscr{M}$, let

$$r_p := \sup \left\{ r > 0 : \exp_{p|_{B_r(o_p)}} \text{ is a diffeomorphism} \right\},$$

where o_p denotes the origin of $T_p\mathscr{M}$ and $B_r(o_p) := \{\upsilon \in T_p\mathscr{M} : \|\upsilon - o_p\| < r\}$. Note that if $0 < \delta < r_p$ then $\exp_p B_\delta(o_p) = B_\delta(p)$. The number r_p is called the *injectivity radius* of $\mathscr{M}$ at p.

DEFINITION 2.1 Let $p \in \mathscr{M}$ and r_p be the radius of injectivity at p. Define the quantity

$$K_p := \sup \left\{ \frac{d(\exp_q u, \exp_q \upsilon)}{\|u - \upsilon\|} : q \in B_{r_p}(p), \;\; u, \upsilon \in T_q\mathscr{M},\; u \neq \upsilon, \;\; \|\upsilon\| \leqslant r_p,\; \|u - \upsilon\| \leqslant r_p \right\}.$$

REMARK 2.2 The quantity K_p measures how fast the geodesics spread apart in $\mathscr{M}$. In particular, when $u = 0$ or more generally when u and υ are on the same line through 0,

$$d(\exp_q u, \exp_q \upsilon) = \|u - \upsilon\|.$$

So, $K_p \geqslant 1$ for all $p \in \mathscr{M}$. Now, when $\mathscr{M}$ has a non-negative sectional curvature, the geodesics spread apart less than the rays (Do Carmo, 1992, Chapter 5) so that

$$d(\exp_q u, \exp_q \upsilon) \leqslant \|u - \upsilon\|.$$

As a consequence $K_p = 1$ for all $p \in \mathscr{M}$. Finally, it is worth mentioning that radii less than r_p could be used as well (although this would require additional notation, such as $K_p(\rho)$ for r_p). In this case, the measure by which geodesics spread apart might decrease, thereby providing slightly stronger results, so long as the radius is not too much less than r_p.

Let X be a C^1 vector field on $\mathscr{M}$. The covariant derivative of X determined by the Levi-Civita connection ∇ defines at each $p \in \mathscr{M}$ a linear map $\nabla X(p) : T_p\mathscr{M} \to T_p\mathscr{M}$ given by

$$\nabla X(p)v := \nabla_Y X(p), \tag{2.2}$$

where Y is a vector field such that $Y(p) = v$.

DEFINITION 2.3 Let $\mathscr{M}$ be a complete Riemannian manifold and $Y_1, \ldots, Y_n$ be vector fields on $\mathscr{M}$. Then, the nth covariant derivative of X with respect to $Y_1, \ldots, Y_n$ is defined inductively by

$$\nabla^2_{\{Y_1,Y_2\}} X := \nabla_{Y_2}\nabla_{Y_1} X, \qquad \nabla^n_{\{Y_i\}_{i=1}^n} X := \nabla_{Y_n}(\nabla_{Y_{n-1}} \cdots \nabla_{Y_1} X).$$

DEFINITION 2.4 Let $\mathscr{M}$ be a complete Riemannian manifold, and $p \in \mathscr{M}$. Then, the nth covariant derivative of X at p is the nth multilinear map $\nabla^n X(p) : T_p\mathscr{M} \times \cdots \times T_p\mathscr{M} \to T_p\mathscr{M}$ defined by

$$\nabla^n X(p)(v_1, \ldots, v_n) := \nabla^n_{\{Y_i\}_{i=1}^n} X(p),$$

where $Y_1, \ldots, Y_n$ are vector fields on $\mathscr{M}$ such that $Y_1(p) = v_1, \ldots, Y_n(p) = v_n$.

We remark that Definition 2.4 only depends on the n-tuple of vectors $(v_1, \ldots, v_n)$ since the covariant derivative is tensorial in each vector field Y_i.

DEFINITION 2.5 Let $\mathscr{M}$ be a complete Riemannian manifold, and $p \in \mathscr{M}$. The norm of an nth multilinear map $A : T_p\mathscr{M} \times \cdots \times T_p\mathscr{M} \to T_p\mathscr{M}$ is defined by

$$\|A\| = \sup\left\{\|A(v_1, \ldots, v_n)\| : v_1, \ldots, v_n \in T_p\mathscr{M},\ \|v_i\| = 1,\ i = 1, \ldots, n\right\}.$$

In particular, the norm of the nth covariant derivative of X at p is given by

$$\|\nabla^n X(p)\| = \sup\left\{\|\nabla^n X(p)(v_1, \ldots, v_n)\| : v_1, \ldots, v_n \in T_p\mathscr{M},\ \|v_i\| = 1,\ i = 1, \ldots, n\right\}.$$

LEMMA 2.6 Let Ω be an open subset of $\mathscr{M}$, X a C^1 vector field defined on Ω and $\zeta : [a, b] \to \Omega$ a C^∞ curve. Then

$$P_{\zeta,t,a} X(\zeta(t)) = X(\zeta(a)) + \int_a^t P_{\zeta,s,a} \nabla X(\zeta(s))\, \zeta'(s)\, \mathrm{d}s, \quad t \in [a, b].$$

Proof. See Ferreira & Svaiter (2002). □

LEMMA 2.7 Let Ω be an open subset of $\mathscr{M}$, X a C^2 vector field defined on Ω and $\zeta : [a, b] \to \Omega$ a C^∞ curve. Then, for all $Y \in \mathscr{X}(\mathscr{M})$ we have that

$$P_{\zeta,t,a} \nabla X(\zeta(t))\, Y(\zeta(t)) = \nabla X(\zeta(a)) Y(\zeta(a)) + \int_a^t P_{\zeta,s,a} \nabla^2 X(\zeta(s))(Y(\zeta(s)), \zeta'(s))\, \mathrm{d}s, \quad t \in [a, b].$$

Proof. See Li & Wang (2006). □

LEMMA 2.8 (Banach's lemma.) Let B be a linear operator and let I_p be the identity operator in T_pM. If $\|B - I_p\| < 1$, then B is invertible and $\|B^{-1}\| \leqslant 1/\left(1 - \|B - I_p\|\right)$.

Proof. Under the hypothesis, it is easily shown that $B^{-1} = \sum_{i=0}^{\infty}(B - I_p)^i$ and hence $\|B^{-1}\| \leqslant \sum_{i=0}^{\infty} \|B - I_p\|^i = 1/(1 - \|(B - I_p)\|)$. □

3. Local analysis for Newton's method

In this section, our goal is to state and prove a local theorem for Newton's method. First, we will prove some preliminary results regarding the majorant function, which relaxes the Lipschitz continuity of the derivative. Then we will show that Newton's method is well defined and converges. We will also prove the uniqueness of the solution, and the optimal convergence radius will be established. The statement of the main theorem of the paper is as follows.

THEOREM 3.1 Let $\mathscr{M}$ be a Riemannian manifold, $\Omega \subseteq \mathscr{M}$ an open set and $X : \Omega \to T\mathscr{M}$ a continuously differentiable vector field. Let $p_* \in \Omega$, $R > 0$ and $\kappa := \sup\{t \in [0, R) : B_t(p_*) \subset \Omega\}$. Suppose that $X(p_*) = 0$, $\nabla X(p_*)$ is invertible and there exists an $f : [0,\ R) \to \mathbb{R}$ continuously differentiable such that

$$\left\| \nabla X(p_*)^{-1}[P_{\zeta,1,0}\, \nabla X(p) - \ P_{\zeta,\tau,0} \nabla X(\zeta(\tau)) P_{\zeta,1,\tau}] \right\| \leqslant f'\left(d(p_*, p)\right) - f'\left(\tau d(p_*, p)\right) \tag{3.1}$$

for all $\tau \in [0, 1]$, $p \in B_\kappa(p_*)$, where $\zeta : [0, 1] \to \mathscr{M}$ is a minimizing geodesic from p_* to p, and

h1) $f(0) = 0$ and $f'(0) = -1$,

h2) f' is strictly increasing.

Let $\nu := \sup\{t \in [0, R) : f'(t) < 0\}$, $\rho := \sup\{\delta \in (0, \nu) : [f(t)/f'(t) - t]/t < 1/K_{p_*},\ t \in (0, \delta)\}$ and

$$r := \min\left\{\kappa,\ \rho,\ r_{p_*}\right\}.$$

Then the sequences with starting points $p_0 \in B_r(p_*) \setminus \{p_*\}$ and $t_0 = d(p_*, p_0)/K_{p_*}^{1/\mu}$ with $0 \leqslant \mu \leqslant 1$, namely,

$$p_{k+1} = \exp_{p_k}(-\nabla X(p_k)^{-1} X(p_k)) \quad \text{and} \quad t_{k+1} = |t_k - f(t_k)/f'(t_k)|, \quad k = 0, 1, \ldots, \tag{3.2}$$

respectively, are well defined; $\{t_k\}$ is strictly decreasing, is contained in $(0, r)$ and converges to 0; $\{p_k\}$ is contained in $B_r(p_*)$ and converges to the point p_* which is the unique zero of X in $B_\sigma(p_*)$, where $\sigma := \sup\{t \in (0, \kappa) : f(t) < 0\}$ and we have that

$$\lim_{k\to\infty} \left[d(p_*,\ p_{k+1})\big/d(p_*,\ p_k)\right] = 0, \qquad \lim_{k\to\infty} [t_{k+1}/t_k] = 0. \tag{3.3}$$

Moreover, if $K_{p_*} = 1$, $f(\rho)/(\rho f'(\rho)) - 1 = 1$ and $\rho < \min\{\kappa, r_{p_*}\}$, then $r = \rho$ is the best possible convergence radius.
If, additionally,

h3) the function $(0,\ \nu) \ni t \mapsto [f(t)/f'(t) - t]/t^{\mu+1}$ is strictly increasing,

then the sequence $\{t_{k+1}/t_k^{\mu+1}\}$ is strictly decreasing and

$$d(p_*,\ p_{k+1}) \leqslant K_{p_*}\left[t_{k+1}/t_k^{\mu+1}\right] d(p_*,\ p_k)^{\mu+1}, \qquad k = 0, 1, \ldots. \tag{3.4}$$

REMARK 3.2 If f has a convex derivative f', then **h3** holds with $\mu = 1$. The proof follows by using an argument similiar to the one used to prove Ferreira (2009, Proposition 2.6). In this case, the Newton sequence converges with a quadratic rate.

Now, we will give some examples of majorant functions satisfying conditions **h1**, **h2** and **h3**.

EXAMPLE 3.3 The following continuously differentiable functions satisfy **h1**, **h2** and **h3**:

i) $f : [0, +\infty) \to \mathbb{R}$ such that $f(t) = t^{1+\mu} - t$;

ii) $f : [0, +\infty) \to \mathbb{R}$ such that $f(t) = e^{-t} + t^2 - 1$.

Letting $0 < \mu < 1$, the derivative of the first function is not convex, nor that of the second.

From now on, we assume that the hypotheses of Theorem 3.1 hold, with the exception of **h3**, which will be considered to hold only when explicitly stated.

4. Preliminary results

In this section, we will prove all statements in Theorem 3.1 regarding the sequence $\{t_k\}$ associated to the majorant function. The main relationships between the majorant function and the vector field will also be established, as well as the results in Theorem 3.1 related to the uniqueness of the solution and the optimal convergence radius.

4.1 *The scalar sequence*

Next, we will prove the statements in Theorem 3.1 involving $\{t_k\}$. We begin by proving that the constants κ, ν and σ are positive.

PROPOSITION 4.1 The constants κ, ν and σ are positive and $t - f(t)/f'(t) < 0$ for all $t \in (0, \nu)$.

Proof. Since Ω is open and $p_* \in \Omega$, we can immediately conclude that $\kappa > 0$. As f' is continuous in 0 with $f'(0) = -1$, there exists $\delta > 0$ such that $f'(t) < 0$ for all $t \in (0, \delta)$. So, $\nu > 0$. Now, because $f(0) = 0$ and f' is continuous in 0 with $f'(0) = -1$, there exists $\delta > 0$ such that $f(t) < 0$ for all $t \in (0, \delta)$. Hence $\sigma > 0$.

Assumption **h2** implies that f is strictly convex. So, using the strict convexity of f and the first equality in assumption **h1** we have $f(t) - tf'(t) < f(0) = 0$ for $t \in (0, R)$. If $t \in (0, \nu)$ then $f'(t) < 0$, which combined with the last inequality yields the desired inequality. □

According to **h2** and the definition of ν, we have $f'(t) < 0$ for all $t \in [0, \nu)$. Therefore, Newton's iteration map for f is well defined in $[0, \nu)$. Let us call it n_f,

$$\begin{aligned} n_f : [0, \nu) &\to (-\infty, 0], \\ t &\mapsto t - f(t)/f'(t). \end{aligned} \tag{4.1}$$

Since $f'(t) \neq 0$ for all $t \in [0, \nu)$ the Newton iteration map n_f is a continuous function.

PROPOSITION 4.2 $\lim_{t\to 0} |n_f(t)|/t = 0$. As a consequence, $\rho > 0$ and $|n_f(t)| < t/K_{p_*}$ for all $t \in (0, \rho)$.

Proof. Using Definition 4.1, Proposition 4.1, $f(0) = 0$, and the definition of ν, a simple algebraic manipulation gives

$$\frac{|n_f(t)|}{t} = [f(t)/f'(t) - t]/t = \frac{1}{f'(t)} \frac{f(t) - f(0)}{t - 0} - 1, \qquad t \in (0, \nu). \tag{4.2}$$

Since $f'(0) \neq 0$, the first statement follows by taking the limit in (4.2), as t goes to 0. Since $\lim_{t\to 0} |n_f(t)|/t = 0$, the first equality in (4.2) implies that there exists $\delta > 0$ such that

$$0 < [f(t)/f'(t) - t]/t < 1/K_{p_*}, \qquad t \in (0, \delta).$$

So, we conclude that ρ is positive. Therefore, the first equality in (4.2) together with the definition of ρ implies that $|n_f(t)|/t = [f(t)/f'(t) - t]/t < 1/K_{p_*}$ for all $t \in (0, \rho)$, as required. □

Using (4.1), it is easy to see that the sequence $\{t_k\}$ is equivalently defined as

$$t_0 = d(p_*, p_0)/K_{p_*}^{1/\mu}, \qquad t_{k+1} = |n_f(t_k)|, \qquad k = 0, 1, \ldots. \tag{4.3}$$

COROLLARY 4.3 The sequence $\{t_k\}$ is well defined, strictly decreasing and contained in $(0, \rho)$. Moreover, $\{t_k\}$ converges to 0 with a superlinear rate, i.e., $\lim_{k\to\infty} t_{k+1}/t_k = 0$. If, additionally, **h3** holds then the sequence $\{t_{k+1}/t_k^{\mu+1}\}$ is strictly decreasing.

Proof. Since $K_{p_*} \geqslant 1$ we have $0 < t_0 = d(p_*, p_0)/K_{p_*}^{1/\mu} \leqslant d(p_*, p_0) < r \leqslant \rho$. So, using $K_{p_*} \geqslant 1$ together with Proposition 4.2 and an induction argument, we conclude that $t_{k+1} = |n_f(t_k)| < t_k$ for $k = 0.1, \ldots$. Hence, $\{t_k\}$ is well defined, strictly decreasing and contained in $(0, \rho)$. So, we have proved the first statement of the corollary.

Since $\{t_k\} \subset (0, \rho)$ is strictly decreasing it converges. So, $\lim_{k\to\infty} t_k = t_*$ and $0 \leqslant t_* < \rho$, which together with (4.3) and the continuity of n_f imply that $0 \leqslant t_* = |n_f(t_*)|$. But, if $t_* \neq 0$ then Proposition 4.2 implies $|n_f(t_*)| < t_*$, hence $t_* = 0$. Now, $\lim_{k\to\infty} t_k = 0$. Thus, the definition of $\{t_k\}$ in (4.3) and the first statement in Proposition 4.2 imply that $\lim_{k\to\infty} t_{k+1}/t_k = \lim_{k\to\infty} |n_f(t_k)|/t_k = 0$ and the second statement is proved.

Since $\{t_k\}$ is strictly decreasing, the last statement is an immediate consequence of **h3**. □

4.2 *Relationship between the majorant function and the vector field*

Next, we will present the main relationships between the majorant function f and the vector field X.

LEMMA 4.4 If $d(p_*, p) < \min\{\kappa, \nu\}$, then $\nabla X(p)$ is invertible and

$$\|\nabla X(p)^{-1} P_{\zeta,0,1} \nabla X(p_*)\| \leqslant 1/|f'(d(p_*, p))|,$$

where $\zeta : [0, 1] \to \mathscr{M}$ is a minimizing geodesic from p_* to p. In particular, $\nabla X(p)$ is invertible for all $p \in B_r(p_*)$, where r is as defined in Theorem 3.1.

Proof. Let $I_{p_*} : T_{p_*}\mathscr{M} \to T_{p_*}\mathscr{M}$ be the identity operator, $p \in B_\kappa(p_*)$ and $\zeta : [0, 1] \to \mathscr{M}$ be a minimizing geodesic from p_* to p. Since $P_{\zeta,0,0} = I_{p_*}$ and $P_{\zeta,0,1}$ is an isometry we obtain that

$$\left\| \nabla X(p_*)^{-1} P_{\zeta,1,0} \nabla X(p) P_{\zeta,0,1} - I_{p_*} \right\| = \left\| \nabla X(p_*)^{-1} [P_{\zeta,1,0} \nabla X(p) - P_{\zeta,0,0} \nabla X(p_*) P_{\zeta,1,0}] \right\|.$$

Since $d(p_*, p) < \nu$ we have $f'(d(p_*, p)) < 0$. Thus, using the last equation, (3.1) and **h1** we conclude that

$$\|\nabla X(p_*)^{-1} P_{\zeta,1,0} \nabla X(p) P_{\zeta,0,1} - I_{p_*}\| \leqslant f'(d(p_*, p)) - f'(0) < -f'(0) = 1.$$

Hence, it follows from the first part of Lemma 2.8 that $\nabla X(p_*)^{-1} P_{\zeta,1,0} \nabla X(p) P_{\zeta,0,1}$ is invertible, as well as $\nabla X(p)$. Moreover, because $P_{\zeta,0,1}$ is an isometry, the second part of Lemma 2.8 and the above inequality imply that

$$\|\nabla X(p)^{-1} P_{\zeta,0,1} \nabla X(p_*)\| \leqslant \frac{1}{1-(f'(d(p_*,p))-f'(0))} = \frac{1}{|f'(d(p_*,p))|},$$

where in the last equality we used $f'(0)=-1$ and $f'<0$ in $[0,\nu)$. As $r \leqslant \nu$, the last statement is proved. □

Lemma 4.4 guarantees, in particular, that $\nabla X(p)$ is invertible for all $p \in B_r(p_*)$ and, consequently, the Newton iteration map is well defined. Let us denote by N_X the Newton iteration map for X in that region:

$$\begin{aligned} N_X : B_r(p_*) &\to \mathscr{M} \\ p &\mapsto \exp_p(-\nabla X(p)^{-1} X(p)). \end{aligned} \tag{4.4}$$

One can apply a *single* Newton iteration to any $p \in B_r(p_*)$ to obtain $N_X(p)$ which may not belong to $B_r(p_*)$, and may not even belong to the domain of X. So, this is enough to guarantee the well-definedness of only one iteration. To ensure that the Newton iterations may be repeated indefinitely, we need some additional results.

Newton's iteration at a point happens to be a zero of the linearization of X at such a point. So, we study the linearization error at a point in $B_\kappa(p_*)$:

$$E_X(p, p*) := X(p*) - P_{\alpha,0,1}\left[X(p) + \nabla X(p)\alpha'(0)\right], \qquad p \in B_\kappa(p_*), \tag{4.5}$$

where $\alpha : [0,1] \to \mathscr{M}$ is a minimizing geodesic from p to $p*$. We will bound this error by the error in the linearization on the majorant function f

$$e_f(t,u) := f(u) - \left[f(t) + f'(t)(u-t)\right], \qquad t,\ u \in [0,R). \tag{4.6}$$

LEMMA 4.5 If $d(p_*,p) \leqslant \kappa$ then $\|\nabla X(p_*)^{-1} E_X(p,p_*)\| \leqslant e_f(d(p_*,p),0)$.

Proof. Let $p \in B_\kappa(p_*)$ and let $\zeta : [0,1] \to \mathscr{M}$ be a minimizing geodesic from p_* to p. Let $\alpha(u) = \zeta(1-u)$. Using Lemma 2.6 and $P_{\alpha,0,1} = {P_{\alpha,1,0}}^{-1}$ it follows that

$$X(p_*) = P_{\alpha,0,1} X(p) + \int_0^1 P_{\alpha,u,1}(\nabla X(\alpha(u))\alpha'(u))\,\mathrm{d}u.$$

Since α is a geodesic, α' is parallel and $\alpha'(u) = P_{\alpha,0,u}\alpha'(0)$. So, the latter equation implies

$$X(p_*) - P_{\alpha,0,1}[X(p) + \nabla X(p)\alpha'(0)] = \int_0^1 \left[P_{\alpha,u,1}\nabla X(\alpha(u))P_{\alpha,0,u} - P_{\alpha,0,1}\nabla X(p)\right]\alpha'(0)\,\mathrm{d}u.$$

Since $\alpha(u) = \zeta(1-u)$, we have $P_{\alpha,u,1} = P_{\zeta,1-u,0}$, $P_{\alpha,0,u} = P_{\zeta,1,1-u}$ and $P_{\alpha,0,1} = P_{\zeta,1,0}$. Hence, the last equation becomes

$$X(p_*) - P_{\alpha,0,1}[X(p) + \nabla X(p)\alpha'(0)] = \int_0^1 \left[P_{\zeta,1,0}\nabla X(p) - P_{\zeta,1-u,0}\nabla X(\zeta(1-u))P_{\zeta,1,1-u}\right]\zeta'(1)\,\mathrm{d}u.$$

Therefore, using the last equality, (3.1), $\|\zeta'(1)\| = d(p_*,p)$ and (4.5) it is easy to conclude that

$$\|\nabla X(p_*)^{-1} E_X(p,p_*)\| \leqslant \int_0^1 [f'(d(p_*,p)) - f'((1-u)d(p_*,p))]d(p_*,p)\,\mathrm{d}u.$$

Since $f(0)=0$, performing the above integral and using (4.6) the desired result follows. □

LEMMA 4.6 If $d(p_*, p) < r$ then $d(p_*, N_X(p)) \leqslant K_{p_*}|n_f(d(p_*, p))|$. As a consequence, $N_X(B_r(p_*)) \subset B_r(p_*)$.

Proof. Since $X(p_*) = 0$, the inequality is trivial for $p = p_*$. Now, assume that $0 < d(p_*, p) \leqslant r$. Lemma 4.4 implies that $\nabla X(p)$ is invertible. Let $\alpha : [0, 1] \to \mathscr{M}$ be a minimizing geodesic from p to p_*. Thus, because $X(p_*) = 0$, the definition of $E_X(p, p_*)$ in (4.5) and direct manipulation yields

$$-\nabla X(p)^{-1} P_{\alpha,1,0} E_X(p, p_*) = \nabla X(p)^{-1} X(p) + \alpha'(0).$$

Using the above equation, Lemmas 4.4 and 4.5, it is easy to conclude that

$$\begin{aligned}\|\nabla X(p)^{-1} X(p) + \alpha'(0)\| &\leqslant \| - \nabla X(p)^{-1} P_{\alpha,1,0} \nabla X(p_*)\| \|\nabla X(p_*)^{-1} E_F(p, p_*)\| \\ &\leqslant e_f(d(p_*, p), 0)/|f'(d(p_*, p))|.\end{aligned}$$

On the other hand, taking into account that $f(0) = 0$, the definitions of e_f and n_f imply that

$$e_f(d(p_*, p), 0)/|f'(d(p_*, p))| = -n_f(d(p_*, p)) = |n_f(d(p_*, p))|.$$

Since $K_{p_*} \geqslant 1$ and $d(p_*, p) < r$, combining the above two equations, the second part of Proposition 4.2 and the definitions of r and r_{p_*} in Theorem 3.1 we obtain that

$$\|\nabla X(p)^{-1} X(p) + \alpha'(0)\| \leqslant |n_f(d(p_*, p))| < d(p_*, p)/K_{p_*} \leqslant r_{p_*}, \quad \|\alpha'(0)\| = d(p_*, p) \leqslant r_{p_*}.$$

Thus, letting $p = p_*, q = p, v = \alpha'(0), u = -\nabla X(p)^{-1} X(p)$ in Definition 2.1 and taking into account the last equation, we conclude that

$$d(p_*, N_X(p)) \leqslant K_{p_*} \|\nabla X(p)^{-1} X(p) + \alpha'(0)\| \leqslant K_{p_*} |n_f(d(p_*, p))|.$$

So, the first statement is proved.

Take $p \in B_r(p_*)$. Since $d(p_*, p) < r$ and $r \leqslant \rho$, the first part of the lemma and the second part of Proposition 4.2 imply that $d(p_*, N_X(p)) \leqslant K_{p_*}|n_f(d(p_*, p))| < d(p_*, p)$ and the result follows. □

LEMMA 4.7 If **h3** holds and $d(p_*, p) \leqslant t < r$ then $d(p_*, N_X(p)) \leqslant [K_{p_*}|n_f(t)|/t^{\mu+1}]\, d(p_*, p)^{\mu+1}$.

Proof. The inequality is trivial for $p = p_*$. Since $0 < d(p_*, p) \leqslant t$, then assumption **h3** and (4.1) give $|n_f(d(p_*, p))|/d(p_*, p)^{\mu+1} \leqslant |n_f(t)|/t^{\mu+1}$. So, using Lemma 4.6 the result follows. □

5. Uniqueness and optimal convergence radius

Next, we will obtain the uniqueness of the solution and the optimal convergence radius.

LEMMA 5.1 The point p_* is the unique zero of X in $B_\sigma(p_*)$.

Proof. Assume that $q \in B_\sigma(p_*)$ and $X(q) = 0$. Let $\alpha : [0, 1] \to \mathscr{M}$ be a minimizing geodesic from p_* to q. Since $X(p_*) = 0$ and $X(q) = 0$, using Lemma 2.6 it is easy to see that

$$\alpha'(0) = -\int_0^1 \nabla X(p_*)^{-1} \left[P_{\alpha,u,0} \nabla X(\alpha(u)) - P_{\alpha,0,0} \nabla X(\alpha(0)) P_{\alpha,u,0}\right] P_{\alpha,0,u}\, \alpha'(0)\, \mathrm{d}u. \tag{5.1}$$

Let $\zeta(t) = \alpha(tu)$. Since $P_{\zeta,1,0} = P_{\alpha,u,0}$ we have that

$$P_{\alpha,u,0} \nabla X(\alpha(u)) - P_{\alpha,0,0} \nabla X(\alpha(0)) P_{\alpha,u,0} = P_{\zeta,1,0} \nabla X(\zeta(1)) - P_{\zeta,0,0} \nabla X(\alpha(0)) P_{\zeta,1,0}.$$

Thus, using the last equality and condition (3.1) with $q = \zeta(1) = \alpha(u)$ and $\tau = 0$, in view of $f_0(0) = -1$, it is easy to conclude that

$$\left\| \nabla X(p_*)^{-1} \left[P_{\alpha,u,0} \nabla X(\alpha(u)) - P_{\alpha,0,0} \nabla X(\alpha(0)) P_{\alpha,u,0} \right] \right\| \leqslant f'(d(p_*, \alpha(u))) + 1.$$

Since $\alpha : [0,1] \to \mathscr{M}$ is the minimizing geodesic from p_* to q we have that $\|\alpha'(0)\| = d(p_*, q)$, $d(p_*, \alpha(u)) = u d(p_*, q)$. Hence, combining the last inequality with (5.1) and taking into account that $f(0) = 0$ and that parallel transport is an isometry, we conclude that

$$d(p_*, q) \leqslant \int_0^1 \left[f'(u d(p_*, q)) + 1 \right] d(p_*, q)\, \mathrm{d}u = f(d(p_*, q)) + d(p_*, q).$$

Therefore, $f(d(p_*, p)) \geqslant 0$. Now, since f is strictly convex, we shall have $f < 0$ in $(0, \sigma)$, i.e., 0 is the unique zero of f in $[0, \sigma)$. Hence, because $d(p_*, q) < \sigma$ and $f(d(p_*, q)) \geqslant 0$ we have $d(p_*, q) = 0$, i.e., $q = p_*$. So, p_* is the unique zero of X in $B_\sigma(p_*)$. □

LEMMA 5.2 If $K_{p_*} = 1$, $f(\rho)/(\rho f'(\rho)) - 1 = 1$ and $\rho < \min\{\kappa, r_{p_*}\}$, then $r = \rho$ is the optimal convergence radius.

Proof. Let $\mathscr{M} = \mathbb{R}$. The curvature of $\mathscr{M}$ is equal to zero, $r_p = \infty$ and $K_t = 1$ for all $t \in \mathscr{M}$. Define the function $h : (-\kappa, \kappa) \to \mathbb{R}$ by

$$h(t) = \begin{cases} -f(-t), & t \in (-\kappa, 0], \\ f(t), & t \in [0, \kappa). \end{cases}$$

It is straightforward to show that $X = h$, $\Omega = (-\kappa, \kappa)$ and $p_* = 0$ satisfy all the assumptions of Theorem 3.1. If $K_{p_*} = 1$, $\rho < \kappa$ and $f(\rho)/(\rho f'(\rho)) - 1 = 1$, then it is easy to conclude that Newton's method for solving $h(t) = 0$, with starting point $t_0 = \rho < \kappa$, produces the cycle $\{(-\rho)^k\}$. Hence, the Newton sequence is well defined and it does not converge. Therefore, the lemma is proved. □

6. The Newton sequence

In this section, we will prove the statements in Theorem 3.1 involving the Newton sequence $\{p_k\}$. First, note that the first equation in (3.2) together with (4.4) implies that the sequence $\{p_k\}$ satisfies

$$p_{k+1} = N_X(p_k), \qquad k = 0, 1, \ldots, \tag{6.1}$$

which is indeed an equivalent definition of this sequence.

PROPOSITION 6.1 The sequence $\{p_k\}$ is well defined, is contained in $B_r(p_*)$ and converges to the point p_*, the unique zero of X in $B_\sigma(p_*)$, and

$$\lim_{k \to \infty} \left[d(p_*, p_{k+1}) / d(p_*, p_k) \right] = 0. \tag{6.2}$$

If, additionally, **h3** holds, then the sequences $\{p_k\}$ and $\{t_k\}$ satisfy

$$d(p_*, p_{k+1}) \leqslant K_{p_*} \left[t_{k+1} / t_k^{\mu+1} \right] d(p_*, p_k)^{\mu+1}, \qquad k = 0, 1, \ldots. \tag{6.3}$$

Proof. Since $p_0 \in B_r(p_*)$ and $r \leqslant \nu$, combining (6.1), the inclusion $N_X(B_r(p_*)) \subset B_r(p_*)$ in Lemma 4.6, and Lemma 4.4, it is easy to conclude that by an induction argument the sequence $\{p_k\}$ is well defined and remains in $B_r(p_*)$.

Now, we are going to prove that $\{p_k\}$ converges to p_*. Since $d(p_*, p_k) < r \leqslant \rho$ for $k = 0, 1, \ldots,$ we obtain from (6.1), Lemma 4.6 and Proposition 4.2 that

$$d(p_*, p_{k+1}) = d(p_*, N_X(p_k)) \leqslant K_{p_*}|n_f(d(p_*, p_k))| < d(p_*, p_k), \qquad k = 0, 1, \ldots. \tag{6.4}$$

So, $\{d(p_*, p_k)\}$ is strictly decreasing and convergent. Let $\ell_* = \lim_{k\to\infty} d(p_*, p_k)$. Since $\{d(p_*, p_k)\}$ rests in $(0, \rho)$ and is strictly decreasing, we have $0 \leqslant \ell_* < \rho$. Thus, the continuity of n_f in $[0, \rho)$ and (6.4) implies that $0 \leqslant \ell_* = K_{p_*}|n_f(\ell_*)|$. On the other hand, we have concluded from Proposition 4.2 that, if $\ell_* > 0$, then $|n_f(\ell_*)| < \ell_*/K_{p_*}$. Hence, we must have $\ell_* = 0$. Therefore, the convergence of $\{p_k\}$ to p_* is proved. The uniqueness of p_* in $B_\sigma(p_*)$ was proved in Lemma 5.1.

For proving the equality in (6.2) note that equation (6.4) implies

$$\left[d(p_*, p_{k+1})/d(p_*, p_k)\right] \leqslant K_{p_*}\left[|n_f(d(p_*, p_k))|/d(p_*, p_k)\right], \qquad k = 0, 1, \ldots.$$

Since $\lim_{k\to\infty} d(p_*, p_k) = 0$, the desired equality follows from the first statement in Proposition 4.2.

Now we will show (6.3). First, we will prove by induction that the sequences $\{t_k\}$ and $\{p_k\}$ defined, in (6.1) and (4.3), respectively, satisfy

$$d(p_*, p_k) \leqslant t_k/K_{p_*}^{1/\mu}, \qquad k = 0, 1, \ldots. \tag{6.5}$$

Since $t_0 = d(p_*, p_0)/K_{p_*}^{1/\mu}$, the above inequality holds for $k = 0$. Now, assume for induction that $d(p_*, p_k) \leqslant t_k/K_{p_*}^{1/\mu}$. Using (6.1), Lemma 4.7, the induction assumption and (4.3), we obtain that

$$d(p_*, p_{k+1}) = d(p_*, N_X(p_k)) \leqslant K_{p_*}\frac{|n_f(t_k)|}{t_k^{\mu+1}}\, d(p_*, p_k)^{\mu+1} \leqslant |n_f(t_k)|/K_{p_*}^{1/\mu} = t_{k+1}/K_{p_*}^{1/\mu},$$

and the proof by induction is complete. Since $K_{p_*} \geqslant 1$, in particular, equation (6.5) implies that

$$d(p_*, p_k) \leqslant t_k, \qquad k = 0, 1, \ldots.$$

Therefore, it is easy to see that the desired inequality follows by combination of the last inequality, Lemma 4.7 , (6.1) and (4.3). □

The proof of Theorem 3.1 follows by the combination of Corollary 4.3, Lemma 4.6, Lemma 5.2 and Proposition 6.1.

7. Special cases

In this section, we will present three special cases of Theorem 3.1.

7.1 *Convergence result under a Hölder-like condition*

In this section, we will present the convergence theorem for Newton's method under a Hölder-like condition, which extends to the Riemannian context the result that has appeared in Huang (2004) and Wang & Li (2003).

THEOREM 7.1 Let $\mathscr{M}$ be a Riemannian manifold, $\Omega \subseteq \mathscr{M}$ an open set and $X : \Omega \to T\mathscr{M}$ a continuously differentiable vector field. Take $p_* \in \Omega$ and let $\kappa := \sup\{t \in [0, R) : B_t(p_*) \subset \Omega\}$. Suppose that $X(p_*) = 0$, $\nabla X(p_*)$ is invertible and there exists a constant $L > 0$ and $0 \leqslant \mu \leqslant 1$ such that

$$\left\| \nabla X(p_*)^{-1} [P_{\zeta,1,0}\, \nabla X(p) - \; P_{\zeta,\tau,0} \nabla X(\zeta(\tau)) P_{\zeta,1,\tau}] \right\| \leqslant L(1 - \tau^{\mu}) d(p_*, p)^{\mu} \tag{7.1}$$

for all $\tau \in [0, 1]$, $p \in B_\kappa(p_*)$, where $\zeta : [0, 1] \to \mathscr{M}$ is a minimizing geodesic from p_* to p. Let

$$r := \min\left\{\kappa, [(\mu + 1)/(L((1 + K_{p_*})\mu + 1))]^{1/\mu}, r_{p_*}\right\}.$$

Then, the sequences with starting point $p_0 \in B_r(p_*) \setminus \{p_*\}$ and $t_0 = d(p_*, p_0)/K_{p_*}^{1/\mu}$ namely

$$p_{k+1} = \exp_{p_k}(-\nabla X(p_k)^{-1} X(p_k)) \quad \text{and} \quad t_{k+1} = \frac{L\mu t_k^{\mu+1}}{(\mu + 1)[1 - L\, t_k^{\mu}]}, \qquad k = 0, 1, \ldots,$$

respectively, are well defined; $\{t_k\}$ is strictly decreasing, is contained in $(0, r)$ and converges to 0; $\{p_k\}$ is contained in $B_r(p_*)$ and converges to p_* which is the unique zero of X in $B_{[(\mu+1)/L]^{1/\mu}}(p_*)$ and there holds

$$d(p_*, p_{k+1}) \leqslant \frac{K_{p_*} L\mu}{(\mu + 1)[1 - Lt_k^{\mu}]} d(p_*, p_k)^{\mu+1}, \qquad k = 0, 1, \ldots.$$

Moreover, if $K_{p_*} = 1$ and $[(\mu + 1)/(L(2\mu + 1))]^{1/\mu} < \min\{\kappa, r_{p_*}\}$, then

$$r = [(\mu + 1)/(L(2\mu + 1))]^{1/\mu}$$

is the best possible convergence radius.

Proof. We can immediately prove that X, p_* and $f : [0, \kappa) \to \mathbb{R}$, defined by $f(t) = Lt^{\mu+1}/(\mu+1) - t$, satisfy the inequality (3.1) and the conditions **h1**, **h2** and **h3** in Theorem 3.1. In this case, it is easy to see that ρ and ν, as defined in Theorem 3.1, satisfy

$$\rho = [(\mu + 1)/(L((1 + K_{p_*})\mu + 1))]^{1/\mu} \leqslant \nu = [1/L]^{1/\mu},$$

and, as a consequence, $r := \min\{\kappa, [(\mu+1)/(L((1+K_{p_*})\mu+1))]^{1/\mu}, r_{p_*}\}$. Moreover, if $K_{p_*} = 1$, then the above equality becomes $\rho = [(\mu + 1)/(L(2\mu + 1))]^{1/\mu}$, and in this case, $f(\rho)/(\rho f'(\rho)) - 1 = 1$, $f(0) = f([(\mu + 1)/L]^{1/\mu}) = 0$ and $f(t) < 0$ for all $t \in (0, [(\mu + 1)/L]^{1/\mu})$. Therefore, the result follows by invoking Theorem 3.1. □

7.2 *Convergence result under Smale's condition*

In this section, we will present a local convergence theorem for Newton's method under Smale's condition. This is of Dedieu *et al.* (2003, Theorem 1.1) (see also Wang & Li, 2006, Theorem 3.1) which generalizes to the Riemannian context Smale (1986, corollary to Proposition 3, p. 195) (see also Blum *et al.* (1997), Proposition 1, p. 157 and Remark 1, p. 158).

THEOREM 7.2 Let $\mathscr{M}$ be an analytic Riemannian manifold, $\Omega \subseteq \mathscr{M}$ an open set and $X : \Omega \to T\mathscr{M}$ an analytic vector field. Take $p_* \in \Omega$ and let $\kappa := \sup\{t \in [0, R) : B_t(p_*) \subset \Omega\}$. Suppose that $X(p_*) = 0$, $\nabla X(p_*)$ is invertible and

$$\gamma := \sup_{n>1} \left\| \frac{\nabla X(p_*)^{-1} \nabla^n X(p_*)}{n!} \right\|^{1/(n-1)} < +\infty. \tag{7.2}$$

Let

$$r := \min \left\{ \kappa, \frac{K_{p*} + 4 - \sqrt{K_{p*}^2 + 8K_{p*} + 8}}{4\gamma}, r_{p_*} \right\}.$$

Then, the sequences with starting point $p_0 \in B_r(p_*) \setminus \{p_*\}$ and $t_0 = d(p_*, p_0)/K_{p_*}$, namely,

$$p_{k+1} = \exp_{p_k}(-\nabla X(p_k)^{-1} X(p_k)) \quad \text{and} \quad t_{k+1} = (\gamma t_k^2)/[2(1-\gamma t_k)^2 - 1], \quad k = 0, 1, \ldots,$$

respectively, are well defined; $\{t_k\}$ is strictly decreasing, is contained in $(0, r)$ and converges to 0; and $\{p_k\}$ is contained in $B_r(p_*)$ and converges to the point p_*, which is the unique zero of X in $B_{1/(2\gamma)}(p_*)$. Moreover, $\{t_{k+1}/t_k^2\}$ is strictly decreasing, $t_{k+1}/t_k^2 < \gamma/[2(1-\gamma\, d(p_*, p_0)/K_{p_*})^2 - 1]$ for $k = 0, 1, \ldots,$ and

$$d(p_*, p_{k+1}) \leqslant \frac{\gamma}{2(1-\gamma t_k)^2 - 1}\, d(p_*, p_k)^2 \leqslant \frac{\gamma}{2(1-\gamma\, d(p_0, p_*)/K_{p_*})^2 - 1}\, d(p_*, p_k)^2, k = 0, 1, \ldots.$$

If, additionally, $K_{p_*} = 1$, $(5-\sqrt{17})/(4\gamma) < \min\{\kappa, r_{p_*}\}$, then $r = (5-\sqrt{17})/(4\gamma)$ is the best possible convergence radius.

We need the following result to prove the above theorem.

LEMMA 7.3 Let $\mathscr{M}$ be an analytic Riemannian manifold, $\Omega \subseteq \mathscr{M}$ an open set and $X : \Omega \to T\mathscr{M}$ an analytic vector field. Suppose that $p_* \in \Omega$, $\nabla X(p_*)$ is invertible and that $B_{1/\gamma}(p_*) \subset \Omega$, where γ is defined in (7.2). Then, for all $p \in B_{1/\gamma}(p_*)$,

$$\|\nabla X(p_*)^{-1} P_{\zeta,1,0} \nabla^2 X(p)\| \leqslant (2\gamma)/(1-\gamma\, d(p_*, p))^3,$$

where $\zeta : [0, 1] \to \mathscr{M}$ is a minimizing geodesic from p_* to p.

Proof. The proof follows the pattern of Alvarez *et al.* (2008, Lemma 5.3). □

The next result gives an alternative condition for checking condition (3.1), whenever the vector field under consideration is twice continuously differentiable.

LEMMA 7.4 Let $\mathscr{M}$ be an analytic Riemannian manifold, $\Omega \subseteq \mathscr{M}$ an open set and $X : \Omega \to T\mathscr{M}$ an analytic vector field. Suppose that $p_* \in \Omega$ and $\nabla X(p_*)$ is invertible. If there exists an $f : [0, R) \to \mathbb{R}$ twice continuously differentiable such that

$$\|\nabla X(p_*)^{-1} P_{\alpha,1,0} \nabla^2 X(q)\| \leqslant f''(d(p_*, q)) \quad \forall q \in B_\kappa(p_*), \tag{7.3}$$

where $\alpha : [0, 1] \to \mathscr{M}$ is a minimizing geodesic from p_* to q, then X and f satisfy (3.1).

Proof. Take $\tau \in [0,1]$ and $p \in B_\kappa(p_*)$ and $\zeta : [0,1] \to \mathscr{M}$ to be the minimizing geodesic from p_* to p. Let $\upsilon \in T_p\mathscr{M}$ and let $Y \in \mathscr{X}(\mathscr{M})$ be the vector field on ζ such that $\nabla_{\zeta'(t)}Y = 0$ and $Y(p) = \upsilon$. Thus, from Lemma 2.7 we have that

$$P_{\zeta,1,\tau}\nabla X(p)Y(p) = \nabla X(\zeta(\tau))Y(\zeta(\tau)) + \int_\tau^1 P_{\zeta,s,\tau}\nabla^2 X(\zeta(s))(Y(\zeta(s)),\zeta'(s))\,\mathrm{d}s.$$

Using $Y(p) = \upsilon$ and $Y(\zeta(\tau)) = P_{\zeta,1,\tau}\upsilon$, we obtain after some algebraic manipulation of the last equality that

$$\nabla X(p_*)^{-1}\left[\nabla X(p) - P_{\zeta,\tau,0}\nabla X(s(\tau))P_{\zeta,0,\tau}\right]\upsilon = \int_\tau^1 \nabla X(p_*)^{-1}P_{\zeta,s,0}\nabla^2 X(\zeta(s))(Y(\zeta(s)),\zeta'(s))\,\mathrm{d}s.$$

Since $\|Y(\zeta(s))\| = \|\upsilon\|$ for all $s \in [0,1]$ and υ is arbitrary, it is easy to conclude from Definition 2.5 that

$$\left\|\nabla X(p_*)^{-1}\left[\nabla X(p) - P_{\zeta,\tau,0}\nabla X(s(\tau))P_{\zeta,0,\tau}\right]\right\| \leqslant \int_\tau^1 \|\nabla X(p_*)^{-1}P_{\zeta,s,0}\nabla^2 X(\zeta(s))\|\,\|\zeta'(s)\|\,\mathrm{d}s.$$

Now, as $\|\zeta'(s)\| = d(p_*,p)$ and $sd(p_*,p) = d(p_*,\zeta(s)) < \kappa$ for all $s \in [0,1]$ and f satisfies (7.3) with $\alpha(t) = \zeta(ts)$ and $q = \zeta(s)$, we obtain from the last inequality that

$$\left\|\nabla X(p_*)^{-1}[P_{\zeta,1,0}\,\nabla X(p) - \,P_{\zeta,\tau,0}\nabla X(\zeta(\tau))P_{\zeta,1,\tau}]\right\| \leqslant \int_\tau^1 f''(sd(p_*,p))d(p_*,p)\,\mathrm{d}s.$$

Evaluating the latter integral, the statement follows. □

COROLLARY 7.5 Let $\mathscr{M}$ be an analytic Riemannian manifold, $\Omega \subseteq \mathscr{M}$ an open set and $X : \Omega \to T\mathscr{M}$ an analytic vector field. Take $p_* \in \Omega$ and let $\kappa := \sup\{t \in [0,R) : B_t(p_*) \subset \Omega\}$ and γ be as defined in (7.2). Suppose that $X(p_*) = 0$ and $\nabla X(p_*)$ is invertible. Then

$$\left\|\nabla X(p_*)^{-1}[P_{\zeta,1,0}\,\nabla X(p) - \,P_{\zeta,\tau,0}\nabla X(\zeta(\tau))P_{\zeta,1,\tau}]\right\| \leqslant \frac{1}{(1-\gamma\, d(p_*,p))^2} - \frac{1}{(1-\tau\gamma\, d(p_*,p))^2}$$

for all $\tau \in [0,1]$, $p \in B_{1/\gamma}(p_*)$, where $\zeta : [0,1] \to \mathscr{M}$ is a minimizing geodesic from p_* to p.

Proof. The proof follows by a combination of Lemma 7.4 and Lemma 7.3. □

Proof of Theorem 7.2. Assume that all hypotheses of Theorem 7.2 hold. Consider the real analytic function $f : [0, 1/\gamma) \to \mathbb{R}$ defined by

$$f(t) = \frac{t}{1-\gamma t} - 2t.$$

It is straightforward to show that f is analytic and that

$$f(0) = 0, \quad f'(t) = 1/(1-\gamma t)^2 - 2, \quad f'(0) = -1, \quad f''(t) = (2\gamma)/(1-\gamma t)^3, \quad f^n(0) = n!\gamma^{n-1},$$

for $n \geqslant 2$. It follows from the last equalities that f satisfies **h1**, **h2** and **h3** with $\mu = 1$. Now, since $f'(t) = 1/(1-\gamma t)^2 - 2$, we conclude from Corollary 7.5 that X and f satisfy (3.1) with $R = 1/\gamma$. In this case, it is easy to see that the constants ν, ρ, r and σ, as defined in Theorem 3.1, satisfy

$$\rho = \frac{K_{p*} + 4 - \sqrt{K_{p*}^2 + 8K_{p*} + 8}}{4\gamma} < \nu = \frac{\sqrt{2}-1}{\sqrt{2}\gamma} < \frac{1}{\gamma}, \qquad r := \min\{\kappa, \rho, r_{p_*}\},$$

$\sigma = 1/(2\gamma)$, $f(0) = f(1/(2\gamma)) = 0$ and $f(t) < 0$ for all $t \in (0,\ 1/(2\gamma))$. Moreover, if $K_{p_*} = 1$, then the above equality becomes $\rho = (5 - \sqrt{17})/(4\gamma)$. Also, $\{t_k\}$ is equivalent to $t_{k+1} = |t_k - f(t_k)/f'(t_k)|$, for $k = 0, 1, \ldots$, and

$$t_{k+1}/t_k^2 = \frac{\gamma}{2(1-\gamma\, t_k)^2 - 1} < \frac{\gamma}{2(1-\gamma\, t_0)^2 - 1} = \frac{\gamma}{2(1-\gamma\, d(p_0, p_*)/K_{p_*})^2 - 1}, \quad k = 0, 1, \ldots.$$

Therefore, the result follows by applying Theorem 3.1. □

7.3 *Convergence result under the Nesterov–Nemirovskii condition*

In this section, we show a corresponding theorem to Theorem 3.1 under the Nesterov–Nemirovskii condition (see Nesterov & Nemirovskii, 1994). To prove this theorem, we first need some definitions and results.

First of all, note that $\mathscr{M} := (\mathbb{R}^n, \langle\cdot,\cdot\rangle_{p_*})$ is a Riemannian manifold, $T_p\mathscr{M} \cong \mathbb{R}^n$ and the distance is given by

$$d(p_*, p) = \|p_* - p\|_{p_*}.$$

Therefore, the open ball of radius $r > 0$ centred at p_* (the Dikin ellipsoid of radius $r > 0$ centred at p_*) is

$$W_r(p_*) := \{p_* \in \mathscr{M} : d(p_*, p) < r\}.$$

The gradient and the Hessian of the function g in $\mathscr{M}$ are given by

$$\operatorname{grad} g(p) := g''(p_*)^{-1} g'(p) \quad \text{and} \quad \operatorname{Hess} g(p) := g''(p_*)^{-1} g''(p), \tag{7.4}$$

respectively. Note that the sectional curvature of $\mathscr{M}$ is zero, the geodesics of $\mathscr{M}$ are straight lines, the parallel transport is the identity, the exponential map is given by

$$\exp_p v = p + v \quad \text{and} \quad p \in \mathscr{M},\ \ v \in T_p\mathscr{M}, \tag{7.5}$$

$r_p \equiv +\infty$ and $K_p \equiv 1$.

THEOREM 7.6 Let $\mathbb{R}^n$ be the n-dimensional Euclidean space and $\langle\cdot,\cdot\rangle$ the usual inner product. Let $\Omega \subset \mathbb{R}^n$ be an open convex set and let $g : \Omega \to \mathbb{R}$ be a strictly convex function, three times continuously differentiable on Ω. Take $p_* \in \Omega$ with $g''(p_*)$ nonsingular. Define a new inner product and the associated norm by

$$\langle u, v\rangle_{p_*} := a^{-1}\langle g''(p_*)u, v\rangle, \quad \|u\|_{p_*} := \sqrt{\langle u, u\rangle_{p_*}}, \ \ u, v \in \mathbb{R}^n,$$

respectively. Suppose that g is a-self-concordant, i.e., satisfies

$$|g'''(p)(v, v, v)| \leqslant 2a^{-1/2}(g''(p)(v, v))^{3/2} \quad \forall\ p \in C,\ \ v \in \mathbb{R}^n.$$

Let

$$r := \min\left\{\kappa, (5 - \sqrt{17})/4\right\}, \qquad W_r(p_*) := \{p_* \in \mathscr{M} : \|p_* - p\|_{p_*} < r\}.$$

Then, the sequences with starting point $p_0 \in W_r(p_*) \setminus \{p_*\}$ and $t_0 = d(p_0, p_*)$, namely,

$$p_{k+1} = p_k - g''(p_k)^{-1} g'(p_k) \quad \text{and} \quad t_{k+1} = t_k^2/[2(1-\gamma t_k)^2 - 1], \qquad k = 0, 1, \ldots,$$

respectively, are well defined; $\{t_k\}$ is strictly decreasing, is contained in $(0, r)$ and converges to 0; and $\{p_k\}$ is contained in $W_r(p_*)$ and converges to the point p_* which is the unique zero of g' in $W_{1/2}(p_*)$. Moreover, $\{t_{k+1}/t_k^2\}$ is strictly decreasing, $t_{k+1}/t_k^2 < 1/[2(1-d(p_*, p_0))^2 - 1]$ for $k = 0, 1, \ldots,$ and

$$d(p_*, p_{k+1}) \leqslant \frac{1}{2(1-t_k)^2 - 1} d(p_*, p_k)^2 \leqslant \frac{1}{2(1-d(p_*, p_0))^2 - 1} d(p_*, p_k)^2, \quad k = 0, 1, \ldots.$$

We also need some auxiliary results about self-concordant functions to prove the above theorem. We begin with the well-known result in the theory of self-concordant functions, namely, if $r < 1$ then $W_r(x_*) \subset \Omega$.

LEMMA 7.7 There holds:

$$\|g''(p_*)^{-1} g'''(p)\|_{p_*} \leqslant \frac{2}{(1-d(p_*, p))^3} \quad \forall\, x \in W_r(p_*),\ r < 1.$$

Proof. The proof follows the same pattern as Alvarez *et al.* (2008, Lemma 5.1). □

Proof of Theorem 7.6. Assume that all hypotheses of Theorem 7.6 hold. Consider the real function $f : [0, 1) \to \mathbb{R}$ defined by

$$f(t) = \frac{t}{1-t} - 2t.$$

It is straightforward to show that

$$f(0) = 0, \quad f'(t) = 1/(1-t)^2 - 2, \quad f'(0) = -1, \quad f''(t) = 2/(1-t)^3, \quad f^n(0) = n!$$

for $n \geqslant 2$. From the last equalities it is easy to see that f satisfies **h1**, **h2** and **h3** with $\mu = 1$. Now, since $f'(t) = 1/(1-t)^2$, setting $\mathscr{M} := (\mathbb{R}^n, \langle .\, . \rangle_{p_*})$ and $X = g'$, it is easy to conclude from Lemma 7.7 that X and f satisfy (3.1) with $R = 1$. In this case, as $K_{p_*} = 1$ and $r_{p_*} = +\infty$ the constants ν, ρ, r and σ, as defined in Theorem 3.1, satisfy

$$\rho = (5 - \sqrt{17})/4 < \nu = \frac{\sqrt{2} - 1}{\sqrt{2}} < 1, \quad r := \min\{\kappa, \rho\},$$

$\sigma = 1/2$, $f(0) = f(1/2) = 0$ and $f(t) < 0$ for all $t \in (0, 1/2)$. The sequence $\{t_k\}$ is equivalent to $t_{k+1} = |t_k - f(t_k)/f'(t_k)|$ for $k = 0, 1, \ldots,$ and

$$t_{k+1}/t_k^2 = \frac{1}{2(1-t_k)^2 - 1} < \frac{1}{2(1-t_0)^2 - 1} = \frac{1}{2(1-d(p_*, p_0))^2 - 1}, \quad k = 0, 1, \ldots.$$

Since $X = g'$, using (7.4) and (7.5), the Newton sequence as defined in (3.2) becomes

$$p_{k+1} = p_k - g''(p_k)^{-1} g'(p_k).$$

Therefore, the result follows by applying Theorem 3.1. □

8. Final remarks

Lemma 7.4 and Lemma 7.7 imply that $f(t) = t/(1-t) - 2t$ is a majorant function to the a-self-concordant function $g : \Omega \to \mathbb{R}$. It is easy to see that f is not self-concordant. So, we cannot apply Lemma 5.2, as before, to conclude that r in Theorem 7.6 is the best possible convergence radius.

The results in Theorem 3.1 are dependent on the injective radius of the exponential map. It would be interesting to establish the convergence radius independently of the injective radius of the exponential map.

Using the theory of Riemannian self-concordant functions (see Jiang *et al.*, 2004), it would be interesting to establish a Riemannian version of Theorem 7.6.

Acknowledgements

The authors would like to thank the two anonymous referees for their constructive comments, which have helped to substantially improve the presentation of the paper.

Funding

Conselho Nacional de Desenvolvimento Cientco e Tecnologico (CNPq) (302024/2008-5, 480101/2008-6 and 473756/2009-9 to O.P.F.); PRONEX–Optimization Fundaˇc~ao de Amparo a Pesquisa do Estado do Rio de Janeiro/CNPq) (to O.P.F.); Fundaˇc~ao de Apoio a Pesquisa da Universidade Federal de Goias (to O.P.F.); Fundaˇc~ao de Amparo a Pesquisa do Estado do Amazonas (to R.C.M.S.); Coordenaˇc~ao de Aperfeiˇcoamento de Pessoal de Nvel Superior (to R.C.M.S.); CNPq (to R.C.M.S.).

References

ABSIL, P.-A., MAHONY, R. & SEPULCHRE, R. (2008) *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton University Press.

ADLER, R., DEDIEU, J.-P., MARGULIES, J., MARTENS, M. & SHUB, M. (2002) Newton method on Riemannian manifolds and a geometric model for the human spine. *IMA J. Numer. Anal.*, **22**, 1–32.

ALVAREZ, F., BOTLE, J. & MUNIER, J. (2008) A unifying local convergence result for Newton's method in Riemannian manifolds. *Found. Comput. Math.*, **8**, 197–226.

BLUM, L., CUCKER, F., SHUB, M. & SMALE, S. (1997) *Complexity and Real Computation*. Berlin, Germany: Springer.

DEDIEU, J.-P., PRIOURET, P. & MALAJOVICH, G. (2003) Newton's method on Riemannian manifolds: COVARIANT alpha theory. *IMA J. Numer. Anal.*, **23**, 395–419.

DO CARMO, M. P. (1992) *Riemannian Geometry*. Boston: Birkhauser.

FERREIRA, O. P. (2009) Local convergence of Newton's method in Banach space from the viewpoint of the majorant principle. *IMA J. Numer. Anal.*, **29**, 746–759.

FERREIRA, O. P. (2011) Local convergence of Newton's method under majorant condition. *J. Comput. Appl. Math.*, **235**, 1515–1522.

FERREIRA, O. P. & SVAITER, B. F. (2002) Kantorovich's theorem on Newton's method in Riemannian manifolds. *J. Complexity*, **18**, 304–329.

HUANG, Z. (2004) The convergence ball of Newton's method and the uniqueness ball of equations under Hölder-type continuous derivatives. *Comput. Math. Appl.*, **47**, 247–251.

JIANG, D. C., MOORE, J. B. & JI, H. B. (2004) Self-concordant functions for optimization on smooth manifolds, *J. Glob. Optim.*, **38**, 437–457.

KRANTZ, S. G. & PARKS, H. R. (2002) *The Implicit Function Theorem: History, Theory, and Applications*. Boston: Birkhauser.

LANG, S. (1995) *Differential and Riemannian Manifolds*. New York: Springer-Verlag.

LI, C. & WANG, J. H. (2006) Newton's method on Riemannian manifolds: Smale's point estimate theory under the γ-condition. *IMA J. Numer. Anal.*, **26**, 228–251.

LI, C. & WANG, J. H. (2008) Newton's method for sections on Riemannian manifolds: GENERALIZED covariant α-theory. *J. Complexity*, **24**, 423–451.

LI, C., WANG, J.H. & DEDIEU, J.-P. (2009) Smale's point estimate theory for Newton's method on Lie groups. *J. Complexity*, **25**, 128–151.

MOSER, J. (1961) A new technique for the construction of solutions of nonlinear differential equations. *Proc. Natl. Acad. Sci. USA*, **47**, 1824–1831.

NASH, J.(1956) The embedding problem for Riemannian manifolds. *Ann. Math. (2)*, **63**, 20–63.

NESTEROV, Y. & NEMIROVSKII, A. (1994) *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM Studies in Applied Mathematics, vol. 13, Philadelphia, PA: SIAM.

OWREN, B. & WELFERT, B. (2000) The Newton iteration on Lie groups. *BIT*, **40**, 121–145.

SMALE, S. (1986) Newton method estimates from data at one point. *The Merging of Disciplines: New Directions in Pure, Applied and Computational Mathematics* (R. Ewing, K. Gross & C. Martin eds). Berlin, Germany: Springer, pp. 185–196.

SMITH, S. (1994) *Optimization Techniques on Riemannian Manifolds*. Fields Institute Communications, vol. 3. Providence, RI: American Mathematical Society, pp. 113–146.

UDRISTE, C. (1994) *Convex Functions and Optimization Methods on Riemannian Manifolds*. Mathematics and its Applications, vol. 297. Dordrecht: Kluwer Academic.

WANG, X. (2000) Convergence of Newton's method and uniqueness of the solution of equations in Banach space. *IMA J. Numer. Anal.*, **20**, 123–134.

WANG, J. H. & LI, C. (2006) Uniqueness of the singular points of vector fields on Riemannian manifolds under the γ-condition. *J. Complexity*, **22**, 533–548.

WANG, J. H. & LI, C. (2011) Kantorovich's theorems of Newton's method for mappings and optimization problems on Lie groups. *IMA J. Numer. Anal.*, **31**, 322–347.

WANG, X. & LI, C. (2003) Convergence of Newton's method and uniqueness of the solution of equations in Banach spaces II. *Acta Math. Sin. Engl. Ser.*, **19**, 405–412.

WAYNE, C. E. (1996) *An Introduction to KAM Theory. Dynamical Systems and Probabilistic Methods in Partial Differential Equation*. Lectures in Applied Mathematics, vol. 31. Providence, RI: American Mathematical Society.