TAIWANESE JOURNAL OF MATHEMATICS
Vol. 13, No. 2B, pp. 633-656, April 2009
This paper is available online at http://www.tjm.nsysu.edu.tw/

EXTENDED NEWTON'S METHOD FOR MAPPINGS ON RIEMANNIAN MANIFOLDS WITH VALUES IN A CONE

Jin-Hua Wang, Shuechin Huang and Chong Li*

Abstract. Robinson's generalized Newton's method for nonlinear functions with values in a cone is extended to mappings on Riemannian manifolds with values in a cone. When $\mathcal{D}f$ satisfies the L-average Lipschitz condition, we use the majorizing function technique to establish the semi-local quadratic convergence of the sequences generated by the extended Newton's method. As applications, we also obtain Kantorovich's type theorem, Smale's type theorem under the γ-condition and an extension of the theory of Smale's approximate zeros.

1. Introduction

In a Banach space, systems of nonlinear equalities and inequalities appear in a wide variety of problems in applied mathematics. These systems play a central role in the model formulation design and analysis of numerical techniques employed in solving problems arising in optimization, complementarity, and variational inequalities (see [5, 8, 7, 27, 31]). Newton's method is a well-known and very powerful technique for solving nonlinear systems of equations. One of the most important results on Newton's method is Kantorovich's theorem [17, 18], which provides a simple and clear criterion, based on the knowledge of the first derivative around the initial point, ensuring the existence, uniqueness of the solution of the equation and the quadratic convergence of Newton's method. Another important result on Newton's method is Smale's point estimate theory (i.e., α-theory and γ-theory) in

Received December 1, 2008.
2000 *Mathematics Subject Classification*: 65H05, 65J15.
Key words and phrases: Extended Newton's method, Riemannian manifolds, L-average Lipschitz condition.
The research of the first and third authors were partially supported by the National Natural Science Foundation of China (grants 10671175; 10731060). The second author was partially supported by a grant from the National Science Council of Taiwan, R.O.C.
*Corresponding author.

[2, 32], where the notion of an approximate zero was introduced and the rules to judge an initial point to be an approximate zero were established, depending only on the information of the analytic nonlinear operator at initial point x_0 and a solution x^*, respectively. There are a lot of works on the weakness and/or extension of the continuity made on the first derivatives of the operators, see for example, [10, 11, 14, 15, 37] and references therein. It should be noted that Wang introduced in [37] the notions of Lipschitz conditions with L-average to unify both Kantorovich's and Smale's criteria.

Newton's method has been extended to solve nonlinear systems of equalities and inequalities in finite-dimensional space (see [5, 27]). Furthermore, Robinson generalized in [31] Newton's method to solve the inclusion problem

$$f(x) \in K, \tag{1.1}$$

where K is a nonempty closed convex cone in a Banach space Y, and f is a function from a reflexive Banach space X into Y. The usual Newton's method corresponds to the special case when K is the degenerate cone $\{0\} \subset Y$. The extended Newton's method to solve (1.1) presented in [31] is as follows:

Algorithm A. Let $x_0 \in X$ be given. For $k = 0, 1, \cdots$, having $x_0, x_1, \cdots, x_k$, determine x_{k+1} as follows. If $\Delta(x_k) \neq \emptyset$, choose $w_k \in \Delta(x_k)$ such that

$$\|w_k\| := \min\{\|w\| \mid w \in \Delta(x_k)\},$$

and set $x_{k+1} = x_k + w_k$, where, for each $x \in X$, $\Delta(x)$ is defined by

$$\Delta(x) := \{w \in X \mid f(x) + f'(x)w \in K\}.$$

In [31], the Kantorovich type theorem was established for the Algorithm A to solve (1.1). Furthermore, the Gauss-Newton method has been extended to solve convex composite optimization (see [3, 4, 20, 24] and the reference therein). Especially, Li and Ng [20] considered an extension of the Gauss-Newton method to convex composite optimization and use the majorizing function technique to establish the semi-local linear/quadratic convergence of the sequences generated by the generalized Gauss-Newton method, which includes the Kantorovich type theorem and Smale's type theorem as special cases.

Recently, there has been an increased interest in studying numerical algorithms on manifolds for there are a lot of numerical problems posed in manifolds arising in many natural contexts, see for example [1, 19, 25, 33, 34]. In particular, in a Riemannian manifold framework, Newton's method is still a main method to solve these problems. An analogue of the well-known Kantorovich's theorem was given in [12] for Newton's method for vector fields on Riemannian manifolds while extensions of the famous Smale's α-theory and γ-theory in [32] to analytic vector fields and analytic mappings on Riemannian manifolds were respectively done in [6]. In the recent paper [22], the convergence criteria in [6] were improved by using the notion of the γ-condition to the vector fields and mappings on Riemannian manifolds. The radii of uniqueness balls of singular points of vector fields satisfying the γ-conditions were studied in [35], while the local behavior of Newton's method on Riemannian manifolds in paper [21]. Furthermore, in [23], a convergence criterion of Newton's method and the radii of the uniqueness balls of the singular points for sections on Riemannian manifolds, which is independent of the curvatures, are established under the assumption that the covariant derivatives of the sections satisfy a kind of the L-average Lipschitz condition.

The purpose of the present paper is to extend Robinson's generalized Newton's method for solving (1.1) to the inclusion problem on Riemannian manifolds, i.e., finding a point $p^* \in M$ such that

$$f(p^*) \in K, \tag{1.2}$$

where K is a nonempty closed convex cone in $\mathbb{R}^n$ and f is a C^1 map from a manifold M into $\mathbb{R}^n$. The extended Newton's method for the inclusion problem (1.2) is defined as follows.

Algorithm 1.1 Let $p_0 \in M$ be given. For $k = 0, 1, \cdots$, having $p_0, p_1, \cdots, p_k$, determine p_{k+1} as follows. If $\Lambda(p_k) \neq \emptyset$, choose $v_k \in \Lambda(p_k)$ such that

$$\|v_k\| := \min\{\|v\| \mid v \in \Lambda(p_k)\} \text{ and set } p_{k+1} = \exp_{p_k} v_k, \tag{1.3}$$

where, for each $p \in M$, $\Lambda(p)$ is defined by

$$\Lambda(p) := \{v \in T_pM \mid f(p) + \mathcal{D}f(p)v \in K\}. \tag{1.4}$$

In the special case when $K = \{0\}$, Algorithm 1.1 reduces to Newton's method on Riemannian manifolds which has been extensively explored in [6, 12, 21, 22, 23, 35] and references therein.

The fact that the linear operator $\mathcal{D}f(p_k)$ is continuous and the assumption that K is closed and convex, imply that $\Lambda(p_k)$ is a closed convex set. This, together with the fact that $T_{p_k}M$ is finite-dimensional, means that there exists $v_k \in \Lambda(p_k)$ such that $\|v_k\| = \min\{\|v\| \mid v \in \Lambda(p_k)\}$, whenever $\Lambda(p_k) \neq \emptyset$.

When $\mathcal{D}f$ satisfies the L-average Lipschitz condition, we use the majorizing function technique to establish the semi-local quadratic convergence of the sequences generated by the extended Newton's method. Applying the results to the special cases, we obtain the well-known Kantorovich type theorem, theorem under the γ-condition and Smale type theorem, respectively. Furthermore, an extension of the theory of Smale's approximate zeros to the setting of inclusion problems on Riemannian manifolds is provided.

2. NOTIONS AND PRELIMINARIES

The notations and notions about smooth manifolds used in the present paper are standard, see for example [9].

Let M be a complete connected m-dimensional Riemannian manifold with the Levi-Civita connection ∇ on M. Let $p \in M$, and let T_pM denote the tangent space at p to M. Let $\langle\cdot,\cdot\rangle$ be the scalar product on T_pM with the associated norm $\|\cdot\|_p$, where the subscript p is sometimes omitted. For any two distinct elements $p,\, q \in M$, let $c : [0,1] \to M$ be a piecewise smooth curve connecting p and q. Then the arc-length of c is defined by $l(c) := \int_0^1 \| c'(t) \| \,\mathrm{d}t$, and the Riemannian distance from p to q by $d(p,q) := \inf_c l(c)$, where the infimum is taken over all piecewise smooth curves $c : [0,1] \to M$ connecting p and q. Thus, by the Hopf-Rinow Theorem (see [9]), (M,d) is a complete metric space and the exponential map at p, $\exp_p : T_pM \to M$ is well-defined on T_pM.

Recall that a geodesic c in M connecting p and q is called a minimizing geodesic if its arc-length equals its Riemannian distance between p and q. Clearly, a curve $c : [0,1] \to M$ is a minimizing geodesic connecting p and q if and only if there exists a vector $v \in T_pM$ such that $\|v\| = d(p,q)$ and $c(t) = \exp_p(tv)$ for each $t \in [0,1]$.

Let $c : \mathbb{R} \to M$ be a C^∞ curve and let $P_{c,\cdot,\cdot}$ denote the parallel transport along c, which is defined by

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

where V is the unique C^∞ vector field satisfying $\nabla_{c'(t)}V = 0$ and $V(c(a)) = v$. Then, for any $a, b \in \mathbb{R}$, $P_{c,c(b),c(a)}$ is an isometry from $T_{c(a)}M$ to $T_{c(b)}M$. Note that, for any $a,\, b,\, b_1,\, b_2 \in \mathbb{R}$,

$$P_{c,c(b_2),c(b_1)} \circ P_{c,c(b_1),c(a)} = P_{c,c(b_2),c(a)} \quad \text{and} \quad P^{-1}_{c,c(b),c(a)} = P_{c,c(a),c(b)}.$$

In particular, we write $P_{q,p}$ for $P_{c,q,p}$ in the case when c is a minimizing geodesic connecting p and q. Let $C^1(TM)$ denote the set of all the C^1-vector fields on M and $C^i(M)$ the set of all C^i-mappings from M to $\mathbb{R}$ ($i = 0, 1$, where C^0-mappings mean continuous mappings), respectively. Let $f : M \to \mathbb{R}^n$ be a C^1 mapping such that

$$f = (f_1, f_2, \cdots, f_n)$$

with $f_i \in C^1(M)$ for each $i = 1, 2, \cdots, n$. Let ∇ be the Levi-Civita connection on M, and let $X \in C^1(TM)$. Following [6] (see also [23]), the derivative of f along the vector field X is defined by

$$\nabla_X f = (\nabla_X f_1, \nabla_X f_2, \cdots, \nabla_X f_n) = (X(f_1), X(f_2), \cdots, X(f_n)).$$

Thus, the derivative of f is a mapping $\mathcal{D}f : (C^1(TM)) \to C^0(M)$ defined by

$$\mathcal{D}f(X) = \nabla_X f \quad \text{for each } X \in C^1(TM). \tag{2.1}$$

We use $\mathcal{D}f(p)$ to denote the derivative of f at p. Let $v \in T_pM$. Taking $X \in C^1(TM)$ such that $X(p) = v$, and any nontrivial smooth curve $c : (-\varepsilon, \varepsilon) \to M$ with $c(0) = p$ and $c'(0) = v$, one has that

$$\mathcal{D}f(p)v := \mathcal{D}f(X)(p) = \nabla_X f(p) = \left(\frac{\mathrm{d}}{\mathrm{d}t}(f \circ c)(t)\right)_{t=0}, \tag{2.2}$$

which only depends on the tangent vector v.

Another important link of the present study relates to the concept of convex process, which was introduced by Rockafellar [28, 29] for convexity problems (see also Robinson [30]).

Definition 2.1. Let E be a Banach space, and let $W : E \to 2^{\mathbb{R}^n}$ be a set-valued mapping. W is called a convex process from E to $\mathbb{R}^n$ if it satisfies

(i) $W(x + y) \supseteq Wx + Wy$ for all $x, y \in E$;
(ii) $W\lambda x = \lambda Wx$ for all $\lambda > 0$, $x \in E$;
(iii) $0 \in W0$.

As usual, the domain, range, and inverse of a convex process W are respectively denoted by $\mathrm{D}(W), \mathrm{R}(W), W^{-1}$; i.e.,

$$\begin{aligned} \mathrm{D}(W) &:= \{x \in E \mid Wx \neq \emptyset\}, \\ \mathrm{R}(W) &:= \cup\{Wx \mid x \in \mathrm{D}(W)\}, \\ W^{-1}y &:= \{x \in E \mid y \in Wx\}. \end{aligned}$$

Obviously, W^{-1} is a convex process from $\mathbb{R}^n$ to E. The norm of a convex process W is defined by

$$\|W\| := \sup\{\|Wx\| \mid x \in \mathrm{D}(W),\ \|x\| \leq 1\},$$

where, following [30, 20] , for a set G in a Banach space, $\|G\|$ denotes its distance to the origin, that is,

$$\|G\| := \inf\{\|a\| \mid \quad a \in G\}.$$

The convex process W is said to be normed if $\|W\| < +\infty$. The following proposition is taken from [31] and will be useful.

Lemma 2.1. *Let $p_0 \in M$ be such that W_{p_0} carries $T_{p_0}M$ onto $\mathbb{R}^n$. Then the following assertions hold:*

(i) $W_{p_0}^{-1}$ *is normed.*

(ii) *If* Q *is a linear transformation from* $T_{p_0}M$ *to* $\mathbb{R}^n$ *such that* $\|W_{p_0}^{-1}\|\|Q\| < 1$, *then the convex process* $\overline{W}$ *defined by*

$$\overline{W} := W_{p_0} + Q$$

carries $T_{p_0}M$ *onto* $\mathbb{R}^n$. *Furthermore,* $\overline{W}^{-1}$ *is normed and*

$$\|\overline{W}^{-1}\| \leq \frac{\|W_{p_0}^{-1}\|}{1 - \|W_{p_0}^{-1}\|\|Q\|}.$$

3. Convergence Criterion

Let Z be a Banach space or a Riemannian manifold. We use $\mathbf{B}_Z(p,r)$ and $\overline{\mathbf{B}_Z(p,r)}$ to denote respectively the open metric ball and the closed metric ball at p with radius r, that is,

$$\mathbf{B}_Z(p,r) := \{q \in Z|\ d(p,q) < r\} \quad \text{and} \quad \overline{\mathbf{B}_Z(p,r)} := \{q \in Z|\ d(p,q) \leq r\}.$$

We often omit the subscript Z if no confusion occurs. Let $p \in Z$ and G be a subset of Z. Then the distance of p to G is denoted by $d(p,G)$ and defined by

$$d(p,G) := \inf\{d(p,q) \mid q \in G\}.$$

Let L be a positive valued nondecreasing integrable function on $[0,+\infty)$. The notion of Lipschitz condition with the L average for operators from Banach spaces to Banch spaces was first introduced in [37] by Wang for the study of Smale's point estimate theory, where the terminology "the center Lipschitz condition in the inscribed sphere with L average" was used (see [37]). Recently, this notion has been extended and applied to sections on Riemannian manifolds in [23]. Note that a mapping $f: M \to \mathbb{R}^n$ is a special example of sections. Definition 3.1 below is a slight modification of the corresponding one in [23] for mappings on Riemannian manifolds. Let f be a C^1 mapping, and let $p \in M$. We define a set-valued mapping W_p from T_pM into $\mathbb{R}^n$ by

$$W_p v := \mathcal{D}f(p)v - K \quad \text{for each } v \in T_pM. \tag{3.1}$$

Since K is a closed convex cone, W_p is a convex process. Obviously, for each $p \in M$, $\mathrm{D}(W_p) = T_pM$. The inverse of W_p is

$$W_p^{-1}y := \{v \in T_pM|\quad \mathcal{D}f(p)v \in y + K\} \quad \text{for each } y \in \mathbb{R}^n. \tag{3.2}$$

W_p is said to carry T_pM onto $\mathbb{R}^n$ if $\mathrm{D}(W_p^{-1}) = \mathbb{R}^n$. By Lemma 2.1, if W_p carries T_pM onto $\mathbb{R}^n$ then W_p^{-1} is normed. In the remainder of this paper, we always assume that $p_0 \in M$ such that W_{p_0} carries $T_{p_0}M$ onto $\mathbb{R}^n$.

Definition 3.1. *Let $r > 0$. $\mathcal{D}f$ is said to satisfy*

(i) *the center L-average Lipschitz condition on $\mathbf{B}(p_0, r)$, if for any point $p \in \mathbf{B}(p_0, r)$ and any geodesic c connecting p_0, p with $l(c) < r$, we have*

$$\|W_{p_0}^{-1}\| \cdot \|\mathcal{D}f(p)P_{c,p,p_0} - \mathcal{D}f(p_0)\| \le \int_0^{l(c)} L(u)\mathrm{d}u. \tag{3.3}$$

(ii) *the L-average Lipschitz condition on $\mathbf{B}(p_0, r)$, if for any two points $p, q \in \mathbf{B}(p_0, r)$ and any geodesic c connecting p, q with $d(p_0, p) + l(c) < r$, we have*

$$\|W_{p_0}^{-1}\| \cdot \|\mathcal{D}f(q)P_{c,q,p} - \mathcal{D}f(p)\| \le \int_{d(p_0,p)}^{d(p_0,p)+l(c)} L(u)\mathrm{d}u. \tag{3.4}$$

Obviously, the L-average Lipschitz condition implies the center L-average Lipschitz condition. Let $r_0 > 0$ be such that

$$\int_0^{r_0} L(u)\mathrm{d}u = 1. \tag{3.5}$$

Then we have the following lemma.

Lemma 3.1. *Let $0 < r \le r_0$. Suppose that $\mathcal{D}f$ satisfies the center L-average Lipschitz condition on $\mathbf{B}(p_0, r)$. Then, for each point $p \in \mathbf{B}(p_0, r)$ and any geodesic c connecting p_0, p with $l(c) < r$, W_p carries T_pM onto $\mathbb{R}^n$ and*

$$\|W_p^{-1}\| \le \frac{\|W_{p_0}^{-1}\|}{1 - \int_0^{l(c)} L(u)\mathrm{d}u}. \tag{3.6}$$

Proof. Since $\mathcal{D}f$ satisfies the center L-average Lipschitz condition on $\mathbf{B}(p_0, r)$, it follows that

$$\|W_{p_0}^{-1}\| \cdot \|\mathcal{D}f(p)P_{c,p,p_0} - \mathcal{D}f(p_0)\| \le \int_0^{l(c)} L(u)\mathrm{d}u < \int_0^{r_0} L(u)\mathrm{d}u = 1. \tag{3.7}$$

For each $v \in T_pM$, the convex process

$$W_p v = \mathcal{D}f(p)v - K = [\mathcal{D}f(p_0)P_{c,p_0,p} + (\mathcal{D}f(p) - \mathcal{D}f(p_0)P_{c,p_0,p})]v - K. \tag{3.8}$$

Since $P_{c,p_0,p}$ is an isomorphism from T_pM to $T_{p_0}M$ (see [16, p.30]), $W_{p_0} \circ P_{c,p_0,p}$ is a convex process from T_pM into $\mathbb{R}^n$ and $\mathcal{D}f(p) - \mathcal{D}f(p_0)P_{c,p_0,p}$ is a linear transformation from T_pM to $\mathbb{R}^n$. Moreover, $\|W_{p_0} \circ P_{c,p_0,p}\| = \|W_{p_0}\|$ and $\|\mathcal{D}f(p) - \mathcal{D}f(p_0)P_{c,p_0,p}\| = \|\mathcal{D}f(p)P_{c,p,p_0} - \mathcal{D}f(p_0)\|$ thanks to the fact that $P_{c,p_0,p}$ is an isometry (see [9, p.56]). Thus, by (3.7) and (3.8), Lemma 2.1 is applicable to concluding that W_p carries T_pM onto $\mathbb{R}^n$ and

$$\|W_p^{-1}\| \leq \frac{\|W_{p_0}^{-1}\|}{1-\|W_{p_0}^{-1}\|\|\mathcal{D}f(p)P_{c,p,p_0}-\mathcal{D}f(p_0)\|} \leq \frac{\|W_{p_0}^{-1}\|}{1-\int_0^{d(p_0,p)} L(u)\mathrm{d}u}.$$

The proof is complete. ∎

Let $b>0$ be such that

$$b=\int_0^{r_0} L(u)u\mathrm{d}u \tag{3.9}$$

and let $\xi>0$. For our main theorem, we define the majorizing function ϕ by

$$\phi(t)=\xi-t+\int_0^t L(u)(t-u)\mathrm{d}u \quad \text{for each } t\geq 0. \tag{3.10}$$

Thus

$$\phi'(t)=-1+\int_0^t L(u)\mathrm{d}u \quad \text{and} \quad \phi''(t)=L(t) \quad \text{for a.e. } t\geq 0. \tag{3.11}$$

Let t_n denote the Newton sequence for ϕ with initial point $t_0=0$ generated by

$$t_{n+1}=t_n-\phi'(t_n)^{-1}\phi(t_n) \quad \text{for each } n=0,1,\cdots. \tag{3.12}$$

In particular, by (3.10), (3.11) and (3.12), one has

$$t_1-t_0=\xi. \tag{3.13}$$

The following lemmas will play a key role in the present paper, which are known in [37] (see also [20]).

Lemma 3.2. *Suppose that $0<\xi\leq b$. Then the following assertions hold:*

(i) *ϕ is strictly decreasing on $[0,r_0]$ and strictly increasing on $[r_0,+\infty)$ with*

$$\phi(\xi)>0, \quad \phi(r_0)=\xi-b\leq 0, \quad \phi(+\infty)\geq\xi>0. \tag{3.14}$$

Moreover, if $\xi<b$, ϕ has two zeros, denoted respectively by r_1 and r_2, such that

$$\xi<r_1<\frac{r_0}{b}\xi<r_0<r_2, \tag{3.15}$$

and if $\xi=b$, then ϕ has a unique zero r_1 in $(\xi,+\infty)$ (in fact $r_1=r_0$).

(ii) *The sequence $\{t_n\}$ generated by* (3.12) *is strictly increasing and converges to r_1.*

(iii) *The convergence of* $\{t_n\}$ *is of quadratic rate if* $\xi < b$, *and linear if* $\xi = b$.

Lemma 3.3. *Let* $0 \le c < +\infty$. *Define*

$$\chi(t) = \frac{1}{t^2}\int_0^t L(c+u)(t-u)\mathrm{d}u \quad \textit{for each } 0 \le t < +\infty.$$

Then χ *is increasing on* $[0, +\infty)$.

Now we are ready to prove the main theorem.

Theorem 3.1. *Suppose that*

$$\xi := \|W_{p_0}^{-1}\| d(f(p_0), K) \le b. \tag{3.16}$$

and that $\mathcal{D}f$ *satisfies the* L*-average Lipschitz condition on* $\mathbf{B}(p_0, r_1)$. *Then the sequence* $\{p_k\}$ *generated by Algorithm 1.1 with initial point* p_0 *is well-defined and converges to a solution* $p^* \in \overline{\mathbf{B}(p_0, r_1)}$ *of the inclusion problem* (1.2). *Moreover, if* $\{v_k\}$ *and* $\{p_k\}$ *are sequences generated by* (1.3), *then the following assertions hold for each* $n = 1, 2, \cdots$:

$$d(p_k, p_{k-1}) \le \|v_{k-1}\| \le t_k - t_{k-1}, \tag{3.17}$$

$$\|v_k\| \le (t_{k+1} - t_k)\left(\frac{\|v_{k-1}\|}{t_k - t_{k-1}}\right)^2 \tag{3.18}$$

and

$$d(p_{k-1}, p^*) \le r_1 - t_{k-1}. \tag{3.19}$$

Proof.
Below, we verify the following implication:

$$(3.17) \text{ holds for all } k = 1, \cdots, n \Longrightarrow (3.18) \text{ holds for } k = n. \tag{3.20}$$

Note that for each $j = 1, \cdots, n$,

$$d(p_j, p_0) \le \sum_{i=1}^{j} d(p_i, p_{i-1}) \le \sum_{i=1}^{j}(t_i - t_{i-1}) = t_j < r_1.$$

Hence, $p_n \in \mathbf{B}(p_0, r_1)$. Thus, Lemma 3.1 is applicable to concluding that W_{p_n} carries $T_{p_n}M$ onto $\mathbb{R}^n$ and

$$\|W_{p_n}^{-1}\| \le \frac{\|W_{p_0}^{-1}\|}{1 - \int_0^{d(p_0,p_n)} L(u)\mathrm{d}u} \le \frac{\|W_{p_0}^{-1}\|}{1 - \int_0^{t_n} L(u)\mathrm{d}u}. \tag{3.21}$$

Therefore, $\Lambda(p_n) \neq \emptyset$ and so there exists $v_n \in \Lambda(p_n)$ such that

$$\|v_n\| = \min\{\|v\| \mid v \in \Lambda(p_n)\}. \tag{3.22}$$

Then, p_{n+1} exists. Now consider the problem of finding $v \in T_{p_n}M$ such that

$$f(p_n) + \mathcal{D}f(p_n)v \in f(p_{n-1}) + \mathcal{D}f(p_{n-1})v_{n-1} + K. \tag{3.23}$$

Since $v_{n-1} \in \Lambda(p_{n-1})$, we have $f(p_{n-1}) + \mathcal{D}f(p_{n-1})v_{n-1} \in K$. Thus,

$$\Lambda(p_n) = \{v \in T_{p_n}M \mid f(p_n) + \mathcal{D}f(p_n)v \in f(p_{n-1}) + \mathcal{D}f(p_{n-1})v_{n-1} + K\}.$$

This together with (3.22) implies that

$$v_n \in W_{p_n}^{-1}[-f(p_n) + f(p_{n-1}) + \mathcal{D}f(p_{n-1})v_{n-1}]. \tag{3.24}$$

Hence, it follows that

$$\|v_n\| \leq \|W_{p_n}^{-1}\|\|-f(p_n) + f(p_{n-1}) + \mathcal{D}f(p_{n-1})v_{n-1}\|. \tag{3.25}$$

Let c_n be the geodesic connecting p_{n-1}, p_n such that $c_n(t) = \exp_{p_{n-1}}(tv_{n-1})$ for each $t \in [0,1]$. Noting that

$$\begin{aligned}
&-f(p_n) + f(p_{n-1}) + \mathcal{D}f(p_{n-1})v_{n-1} \\
=\ &-[f(p_n) - f(p_{n-1}) - \mathcal{D}f(p_{n-1})v_{n-1}] \\
=\ &-\int_0^1 [\mathcal{D}f(c_n(\tau))P_{c_n,c_n(\tau),p_{n-1}} - \mathcal{D}f(p_{n-1})]v_{n-1}\mathrm{d}\tau,
\end{aligned}$$

one has from (3.4) that

$$\begin{aligned}
&\|W_{p_0}^{-1}\|\|-f(p_n) + f(p_{n-1}) + \mathcal{D}f(p_{n-1})v_{n-1}\| \\
\leq &\int_0^1 \|W_{p_0}^{-1}\|\|\mathcal{D}f(c_n(\tau))P_{c_n,c_n(\tau),p_{n-1}} - \mathcal{D}f(p_{n-1})\|\|v_{n-1}\|\mathrm{d}\tau \\
\leq &\int_0^1 \int_{d(p_{n-1},p_0)}^{d(p_{n-1},p_0)+\tau\|v_{n-1}\|} L(u)\mathrm{d}u\|v_{n-1}\|\mathrm{d}\tau \\
= &\int_0^{\|v_{n-1}\|} L(d(p_{n-1},p_0)+u)(\|v_{n-1}\| - u)\mathrm{d}u \\
\leq &\int_0^{\|v_{n-1}\|} L(t_{n-1}+u)(\|v_{n-1}\| - u)\mathrm{d}u.
\end{aligned} \tag{3.26}$$

Since (3.17) holds for $k = n$, we obtains from Lemma 3.3 that

$$\begin{aligned}
&\frac{\int_0^{\|v_{n-1}\|} L(t_{n-1}+u)(\|v_{n-1}\| - u)\mathrm{d}u}{\|v_{n-1}\|^2} \\
\leq &\frac{\int_0^{t_n - t_{n-1}} L(t_{n-1}+u)(t_n - t_{n-1} - u)\mathrm{d}u}{(t_n - t_{n-1})^2}.
\end{aligned} \tag{3.27}$$

Thus, combining (3.21), (3.25), (3.26) and (3.27) yields that

$$\|v_n\| \leq \frac{\int_0^{t_n - t_{n-1}} L(t_{n-1}+u)(t_n - t_{n-1} - u)\mathrm{d}u}{1 - \int_0^{t_n} L(u)\mathrm{d}u} \frac{\|v_{n-1}\|^2}{(t_n - t_{n-1})^2}. \tag{3.28}$$

Note that

$$\begin{aligned}
\phi(t_n) &= \phi(t_n) - \phi(t_{n-1}) - \phi'(t_{n-1})(t_n - t_{n-1}) \\
&= \int_0^1 [\phi'(t_{n-1} + \tau(t_n - t_{n-1})) - \phi'(t_{n-1})]\mathrm{d}\tau(t_n - t_{n-1}) \\
&= \int_0^1 \int_{t_{n-1}}^{t_{n-1}+\tau(t_n - t_{n-1})} L(u)\mathrm{d}u \|v_{n-1}\|\mathrm{d}\tau \\
&= \int_0^{t_n - t_{n-1}} L(t_{n-1}+u)((t_n - t_{n-1}) - u)\mathrm{d}u.
\end{aligned}$$

This, together with (3.28), implies that

$$\|v_n\| \leq (-\phi'(t_n)^{-1}\phi(t_n)) \left(\frac{\|v_{n-1}\|}{t_n - t_{n-1}}\right)^2 = (t_n - t_{n-1}) \left(\frac{\|v_{n-1}\|}{t_n - t_{n-1}}\right)^2.$$

Consequently, the implication (3.20) holds. Clearly, if (3.17) holds for each $k = 1, 2, \cdots$, then $\{p_k\}$ is a Cauchy sequence by the monotonicity of $\{t_n\}$ and hence converges to some point $p^* \in \overline{\mathbf{B}(p_0, r_1)}$. Thus, (3.19) is clear. Moreover, for arbitrary k,

$$[f(p_{k+1}) - f(p^*)] - [f(p_{k+1}) - f(p_k) - \mathcal{D}f(p_k)v_k] \in K - f(p^*). \tag{3.29}$$

The continuity assumptions imply that the left-hand side of (3.29) approaches zero, and since $K - f(p^*)$ is closed, we have $0 \in K - f(p^*)$. So $f(p^*) \in K$. Therefore, to complete the proof, we only need to prove that (3.17) and (3.18) hold for each $k = 1, 2, \cdots$. We proceed by mathematical induction. Since W_{p_0} carries $T_{p_0}M$ onto $\mathbb{R}^n$, we have

$$\Lambda(p_0) = \{v \in T_{p_0}M \mid f(p_0) + \mathcal{D}f(p_0)v \in K\} \neq \emptyset.$$

Thus, there exists $v_0 \in \Lambda(p_0)$ such that $\|v_0\| = \min\{\|v\| \mid v \in \Lambda(p_0)\}$. Below we show that $\|v_0\| \leq \xi$. To proceed, let $c \in K$, and let $v \in W_{p_0}^{-1}(c - f(p_0))$. Then by (3.2), we have that

$$\mathcal{D}f(p_0)v \in c - f(p_0) + K \subseteq -f(p_0) + K,$$

and so $f(p_0) + \mathcal{D}f(p_0)v \in K$, that is, $v \in \Lambda(p_0)$. Hence, $W_{p_0}^{-1}(c - f(p_0)) \subseteq \Lambda(p_0)$. Thus,

$$d(0, \Lambda(p_0)) \leq \|W_{p_0}^{-1}(c - f(p_0))\| \leq \|W_{p_0}^{-1}\| \|c - f(p_0)\|.$$

Since this is valid for each $c \in K$, it follows that

$$\|v_0\| = d(0, \Lambda(p_0)) \leq \|W_{p_0}^{-1}\| d(f(p_0), K) = \xi.$$

Hence,

$$d(p_1, p_0) \leq \|v_0\| \leq \xi = t_1 - t_0,$$

i.e., (3.17) holds for $k = 1$. Then, by (3.20), (3.18) holds for $k = 1$. Furthermore, assume that (3.17) and (3.18) hold for all $1 \leq k \leq n$. Then

$$\|v_n\| \leq (t_{n+1} - t_n) \left(\frac{\|v_{n-1}\|}{t_n - t_{n-1}} \right)^2 \leq t_{n+1} - t_n.$$

This shows that (3.17) holds for $k = n + 1$, and hence (3.17) holds for all k with $1 \leq k \leq n + 1$. Thus, (3.20) implies that (3.18) holds for $k = n + 1$. Therefore, (3.17) and (3.18) hold for each $k = 1, 2, \cdots$. The proof is complete. ■

4. Special Cases and Applications

This section is devoted to four applications of the result in the previous section. The first three are concerned with Kantorovich's type theorem, theorem under the γ-condition and Smale's type theorem, respectively. The last one is concerned with an extension of the theory of Smale's approximate zeros to the setting of inclusion problems on Riemannian manifolds.

4.1. Kantorovich's type theorem

Throughout this subsection, we assume that the function L is a positive constant function. Recall from [12] that $\mathcal{D}f$ is Lipschitz continuous on $\mathbf{B}(p_0, r)$ with modulus $\lambda > 0$ if for any two points $p, q \in \mathbf{B}(p_0, r)$ and any geodesic $c : [0, 1] \to \mathbf{B}(p_0, r)$ connecting p, q with $c(0) = p, c(1) = q$, one has

$$\|\mathcal{D}f(q)P_{c,q,p} - \mathcal{D}f(p)\| \leq \lambda l(c).$$

Clearly, if $\mathcal{D}f$ is Lipschitz continuous on $\mathbf{B}(p_0, r)$ with modulus λ, then $\mathcal{D}f$ satisfies L-average Lipschitz condition on $\mathbf{B}(p_0, r)$ with Lipschitz constant $L = \lambda \|W_{p_0}^{-1}\|$, i.e., if for any two points $p, q \in \mathbf{B}(p_0, r)$ and any geodesic $c : [0, 1] \to \mathbf{B}(p_0, r)$ connecting p, q with $c(0) = p, c(1) = q$ and $d(p_0, p) + l(c) < r$, one has

$$\|W_{p_0}^{-1}\| \|\mathcal{D}f(q)P_{c,q,p} - \mathcal{D}f(p)\| \leq \lambda \|W_{p_0}^{-1}\| l(c).$$

The corresponding majorizing function ϕ reduces to a quadratic function, i.e.,

$$\phi(t) = \xi - t + \frac{\lambda \|W_{p_0}^{-1}\|}{2} t^2 \quad \text{for each } t \geq 0.$$

Then by (3.9),

$$r_0 = \frac{1}{\lambda\|W_{p_0}^{-1}\|} \quad \text{and} \quad b = \frac{1}{2\lambda\|W_{p_0}^{-1}\|}.$$

Moreover, if $\xi \le \frac{1}{2\lambda\|W_{p_0}^{-1}\|}$, the zeros of ϕ are equal to

$$r_1 = \frac{1-\sqrt{1-2\lambda\|W_{p_0}^{-1}\|\xi}}{\lambda\|W_{p_0}^{-1}\|} \quad \text{and} \quad r_2 = \frac{1+\sqrt{1-2\lambda\|W_{p_0}^{-1}\|\xi}}{\lambda\|W_{p_0}^{-1}\|}. \tag{4.1}$$

The Newton sequence $\{t_n\}$ for ϕ with $t_0 = 0$ (see. [13, 26, 36, 37]) satisfies

$$r_1 - t_n = \frac{(1-\rho)\rho^{2^n-1}}{1-\rho^{2^n}} r_1 \quad \text{for each } n = 0, 1, \cdots, \tag{4.2}$$

where $\rho = \frac{1-\sqrt{1-2\lambda\|W_{p_0}^{-1}\|\xi}}{1+\sqrt{1-2\lambda\|W_{p_0}^{-1}\|\xi}}$.

Thus, the following corollary is a trivial application of Theorem 3.1.

Corollary 4.1. *Suppose that*

$$\xi = \|W_{p_0}^{-1}\| d(f(p_0), K) \le \frac{1}{2\lambda\|W_{p_0}^{-1}\|}$$

and that $\mathcal{D}f$ satisfies the Lipschitz continuous on $\mathbf{B}(p_0, r_1)$ with modulus λ. Then the sequence $\{p_k\}$ generated by Algorithm 1.1 with initial point p_0 is well-defined and converges to a solution $p^* \in \overline{\mathbf{B}(p_0, r_1)}$ of the inclusion problem (1.2). *Moreover, if $\{p_k\}$ is the sequence generated by* (1.3)*, then for each $n = 0, 1, \cdots$,*

$$d(p_n, p^*) \le \frac{(1-\rho)\rho^{2^n-1}}{1-\rho^{2^n}} r_1 \le \rho^{2^n-1} r_1. \tag{4.3}$$

4.2. Theorem under the γ-condition

Let k, κ be positive integers such that $k \le \kappa$. Let f be a C^κ-mapping. Following [23], define inductively the derivative of order k for f. Recall that ∇ is the Levi-Civita connection on M. Let $C^\kappa(TM)$ denote the set of all the C^κ-vector fields on M and $C^\kappa(M)$ the set of all C^κ-mappings from M to $\mathbb{R}$, respectively.

Recall from (2.1) that the map $\mathcal{D}^1 f = \mathcal{D}f : (C^\kappa(TM))^1 \to C^{\kappa-1}(M)$ is defined by

$$\mathcal{D}f(X) = \nabla_X(f) \quad \text{for each } X \in C^\kappa(TM).$$

Define the map $\mathcal{D}^k f : (C^\kappa(TM))^k \to C^{\kappa-k}(M)$ by

$$\mathcal{D}^k f(X_1,\cdots,X_{k-1},X_k) = \nabla_{X_k}(\mathcal{D}^{k-1}f(X_1,\cdots,X_{k-1})) - \sum_{i=1}^{k-1}\mathcal{D}^{k-1}f(X_1,\cdots,\nabla_{X_k}X_i,\cdots,X_{k-1}) \tag{4.4}$$

for each $X_1,\cdots,X_{k-1},X_k \in C^\kappa(TM)$. Then, one can use mathematical induction to prove easily that $\mathcal{D}^k f(X_1,\cdots,X_k)$ is tensorial with respect to each component X_i, that is, k multi-linear map from $(C^\kappa(TM))^k$ to $C^{\kappa-k}(M)$, where the linearity refers to the structure of $C^k(M)$-module. This implies that the value of $\mathcal{D}^k f(X_1,\cdots,X_k)$ at $p\in M$ only depends on the k-tuple of tangent vectors $(v_1,\cdots,v_k)=(X_1(p),\cdots,X_k(p))\in(T_pM)^k$. Consequently, for a given $p\in M$, the map $\mathcal{D}^k f(p):(T_pM)^k\to\mathbb{R}^n$, defined by

$$\mathcal{D}^k f(p)v_1\cdots v_k := \mathcal{D}^k f(X_1,\cdots,X_k)(p) \text{ for any } (v_1,\cdots,v_k)\in(T_pM)^k, \tag{4.5}$$

is well-defined, where $X_i\in C^\kappa(TM)$ satisfy $X_i(p)=v_i$ for each $i=1,\cdots,k$.

Let $r>0$ and $\gamma>0$ be such that $r\gamma<1$.

Definition 4.1. f is said to satisfy the γ-condition at p_0 in $\mathbf{B}(p_0,r)$, if for any two points $p,q\in\mathbf{B}(p_0,r)$, any geodesic c connecting p,q with $d(p_0,p)+l(c)<r$, one has

$$\|W_{p_0}^{-1}\|\|\mathcal{D}^2 f(q)\| \le \frac{2\gamma}{(1-\gamma(d(p_0,p)+l(c)))^3}. \tag{4.6}$$

Let L be the function defined by

$$L(u)=\frac{2\gamma}{(1-\gamma u)^3} \quad \text{for each } u \text{ with } 0\le u<\frac{1}{\gamma}. \tag{4.7}$$

The following proposition shows that the γ-condition implies the L-average Lipschitz condition.

Proposition 4.1. *Suppose that f satisfies the γ-condition at p_0 in $\mathbf{B}(p_0,r)$. Then $\mathcal{D}f$ satisfies the L-average Lipschitz condition in $\mathbf{B}(p_0,r)$ with L given by* (4.7).

Proof. Let $p,q\in\mathbf{B}(p_0,r)$, and let c be a geodesic connecting p,q such that $d(p_0,p)+l(c)<r$. It is sufficient to prove that

$$\|W_{p_0}^{-1}\|\|(\mathcal{D}f(q)P_{c,q,p}-\mathcal{D}f(p))\| \le \int_{d(p_0,p)}^{d(p_0,p)+l(c)} \frac{2\gamma}{(1-\gamma u)^3}\mathrm{d}u. \tag{4.8}$$

Let $v\in T_pM$ be arbitrary. Then there exists a unique vector field Y such that $Y(c(0))=v$ and $\nabla_{c'(t)}Y=0$. Then $Y(c(s))=P_{c,c(s),p}v$ for each $s\in[0,1]$. Thus by definition, one has that

$$\begin{aligned} \mathcal{D}f(q)P_{c,q,p}v - \mathcal{D}f(p)v &= \mathcal{D}f(Y)(q) - \mathcal{D}f(Y)(p) \\ &= \int_0^1 \mathcal{D}(\mathcal{D}f(Y)(c(s)))c'(s)\mathrm{d}s. \end{aligned} \tag{4.9}$$

Since $\nabla_{c'(s)}Y(c(s)) = 0$, it follows that

$$\begin{aligned} \mathcal{D}^2 f(c(s))Y(c(s))c'(s) &= c'(s)(\mathcal{D}f(Y)(c(s))) - \mathcal{D}f(\nabla_{c'(s)}Y(c(s))) \\ &= \mathcal{D}(\mathcal{D}f(Y)(c(s)))c'(s). \end{aligned}$$

This, together with (4.9), implies that

$$\mathcal{D}f(q)P_{c,q,p}v - \mathcal{D}f(p)v = \int_0^1 \mathcal{D}^2 f(c(s))Y(c(s))c'(s)\mathrm{d}s. \tag{4.10}$$

Noting that c is a geodesic connecting p and q, there exists $\overline{v} \in T_pM$ such that $q = \exp_p(\overline{v})$ and $l(c) = \|\overline{v}\|$. It follows from (4.10) and (4.6) that

$$\begin{aligned} &\|W_{p_0}^{-1}\|\|\mathcal{D}f(q)P_{c,q,p}v - \mathcal{D}f(p)v\| \\ \le &\int_0^1 \frac{2\gamma}{(1-\gamma(d(p_0,p)+s\|\overline{v}\|))^3}\|\overline{v}\|\|v\|\mathrm{d}s \\ = &\int_{d(p_0,p)}^{d(p_0,p)+l(c)} \frac{2\gamma}{(1-\gamma u)^3}\mathrm{d}u\|v\|. \end{aligned}$$

As $v \in T_pM$ is arbitrary, (4.8) is seen to hold. ∎

Corresponding to the function L given by (4.7), one has from (3.9) and elementary calculation (cf. [37]) that

$$r_0 = \frac{2-\sqrt{2}}{2\gamma}, \quad b = \frac{3-2\sqrt{2}}{\gamma} \tag{4.11}$$

and

$$\phi(t) = \xi - t + \frac{\gamma t^2}{1-\gamma t} \quad \text{for each } t \text{ with } 0 \le t < \frac{1}{\gamma}. \tag{4.12}$$

Thus, from [37], we have the following lemma.

Lemma 4.1. *Assume that $\xi \le b$, namely,*

$$\gamma\xi \le 3 - 2\sqrt{2}. \tag{4.13}$$

Then the following assertions hold:

(i) ϕ *has two zeros given by*

$$\left.\begin{array}{l} r_1 \\ r_2 \end{array}\right\} = \frac{1+\gamma\xi \mp \sqrt{(1+\gamma\xi)^2 - 8\gamma\xi}}{4\gamma}; \tag{4.14}$$

(ii) *the sequence* $\{t_n\}$ *generated by Newton′s method for* ϕ *with the initial point* $t_0 = 0$ *satisfies*

$$t_n = \frac{1-\eta^{2^n-1}}{1-\eta^{2^n-1}\zeta} r_1 \quad \textit{for each } n = 0, 1, \cdots, \tag{4.15}$$

where

$$\eta := \frac{1-\gamma\xi - \sqrt{(1+\gamma\xi)^2 - 8\gamma\xi}}{1-\gamma\xi + \sqrt{(1+\gamma\xi)^2 - 8\gamma\xi}} \quad \textit{and} \quad \zeta := \frac{1+\gamma\xi - \sqrt{(1+\gamma\xi)^2 - 8\gamma\xi}}{1+\gamma\xi + \sqrt{(1+\gamma\xi)^2 - 8\gamma\xi}}; \tag{4.16}$$

(iii) *we have*

$$\frac{t_{n+1}-t_n}{t_n - t_{n-1}} = \frac{1-\eta^{2^n}}{1-\eta^{2^{n-1}}} \cdot \frac{1-\eta^{2^{n-1}-1}\zeta}{1-\eta^{2^{n+1}-1}\zeta} \eta^{2^{n-1}} \le \eta^{2^{n-1}} \quad \textit{for each } n = 1, 2, \cdots. \tag{4.17}$$

In view of Proposition 4.1 and Lemma 4.1, the following corollary follows directly from Theorem 3.1.

Corollary 4.2. *Suppose that*

$$\xi = \|W_{p_0}^{-1}\| d(f(p_0), K) \le \frac{3-2\sqrt{2}}{\gamma}$$

and that f *satisfies the* γ*-condition at* p_0 *in* $\mathbf{B}(p_0, r_1)$*. Then the sequence* $\{p_k\}$ *generated by Algorithm 1.1 with initial point* p_0 *is well-defined and converges to a solution* $p^* \in \overline{\mathbf{B}(p_0, r_1)}$ *of the inclusion problem* (1.2)*. Moreover, if* $\{v_k\}$ *and* $\{p_k\}$ *are sequences generated by* (1.3)*, then the following assertions hold:*

$$\|v_k\| \le \eta^{2^{k-1}} \|v_{k-1}\| \quad \textit{for each } k \ge 1, \tag{4.18}$$

and

$$d(p_k, p^*) \le \eta^{2^k-1} r_1 \quad \textit{for each } k \ge 0, \tag{4.19}$$

where η *is given by* (4.16)*.*

Proof. By Theorem 3.1, the sequence $\{p_k\}$ with initial point p_0 converges to a solution $p^* \in \overline{\mathbf{B}(p_0, r_1)}$ of the inclusion problem (1.2), and the following estimates hold for each k:

$$\begin{array}{rcl} \|v_k\| & \le & (t_{k+1}-t_k)\left(\frac{\|v_{k-1}\|}{t_k - t_{k-1}}\right)^2 \\ & \le & \left(\frac{t_{k+1}-t_k}{t_k-t_{k-1}}\right)\|v_{k-1}\| \end{array}$$

and

$$d(p_k, p^*) \le r_1 - t_k.$$

Hence (4.18) and (4.19) are true because, by (4.15) and (4.17), one has

$$r_1 - t_k = \frac{\eta^{2^k-1}(1-\zeta)}{1-\eta^{2^k-1}\zeta} r_1 \le \eta^{2^k-1} r_1$$

and

$$\frac{t_{k+1} - t_k}{t_k - t_{k-1}} \le \eta^{2^{k-1}}.$$

■

4.3. Smale's type theorem

Throughout the remainder of the present paper, we always assume that f is an analytic mapping. We define, for a point $p_0 \in M$,

$$\gamma(f, p_0) := \sup_{k\ge 2} \left(\frac{\|W_{p_0}^{-1}\| \|\mathcal{D}^k f(p_0)\|}{k!} \right)^{\frac{1}{k-1}}. \tag{4.20}$$

The following proposition is taken from [23].

Proposition 4.2. *Let* $r = \dfrac{1}{\gamma(f, p_0)}$. *Let* $p \in M$, *and let* $v \in T_{p_0}M$ *be such that* $\|v\| < r$ *and* $p = \exp_{p_0}(v)$. *Then*

$$\mathcal{D}^j f(p) = \left(\sum_{k=0}^{\infty} \frac{1}{k!} \mathcal{D}^{k+j} f(p_0) v^k \right) P^j_{c,p_0,p} \quad \textit{for each } j = 0, 1, 2, \cdots, \tag{4.21}$$

where c *is a geodesic connecting* p_0, p *such that* $c(t) = \exp_{p_0} tv$ *for each* $t \in [0, 1]$, *and* $P^j_{c,p_0,p}$ *stands for the map from* $(T_pM)^j$ *to* $(T_{p_0}M)^j$ *defined by*

$$P^j_{c,p_0,p}(v_1, \cdots, v_j) = (P_{c,p_0,p}v_1, \cdots, P_{c,p_0,p}v_j) \quad \textit{for each } (v_1, \cdots, v_j) \in (T_pM)^j.$$

We will show that any analytic mapping satisfies the γ-condition. For this purpose, we need a simple known fact (cf. [2, P.150]):

$$\sum_{j=0}^{\infty} \frac{(k+j)!}{k!\, j!} t^j = \frac{1}{(1-t)^{k+1}} \quad \text{for each } t \in [-1, 1] \text{ and } k = 0, 1, \cdots. \tag{4.22}$$

For simplicity, we use the function ψ defined by

$$\psi(u) := 1 - 4u + 2u^2 \quad \text{for each } u \in [0, 1 - \frac{\sqrt{2}}{2}). \tag{4.23}$$

Note that ψ is strictly decreasing on $[0, 1 - \frac{\sqrt{2}}{2})$.

Lemma 4.2. *Let $p \in M$, and let c be a geodesic connecting p_0 and p such that*

$$u := \gamma(f, p_0)\, l(c) < 1 - \frac{\sqrt{2}}{2}. \tag{4.24}$$

Then W_p carries T_pM onto $\mathbb{R}^n$ and

$$\|W_p^{-1}\| \le \frac{\|W_{p_0}^{-1}\|(1-u)^2}{\psi(u)} \quad \textit{and} \quad \gamma(f, p) \le \frac{\gamma(f, p_0)}{(1-u)\psi(u)}. \tag{4.25}$$

Proof. Assume that c is defined by

$$c(t) = \exp_{p_0}(tv) \quad \text{for each } t \in [0, 1], \tag{4.26}$$

where $v \in T_{p_0}M$. Then $p = \exp_{p_0}(v)$ and $l(c) = \|v\|$; hence, by (4.24),

$$\gamma\, \|v\| = u < 1 - \frac{\sqrt{2}}{2} < 1. \tag{4.27}$$

Thus, Proposition 4.2 (with $j = 1$) is applicable to concluding that

$$\|W_{p_0}^{-1}\|\|\mathcal{D}f(p)P_{c,p,p_0} - \mathcal{D}f(p_0)\| \le \sum_{k=1}^{\infty} \frac{\|W_{p_0}^{-1}\|\|\mathcal{D}^{k+1}f(p_0)\|}{k!}\|v\|^k.$$

It follows from (4.20), (4.22) and (4.27) that

$$\|W_{p_0}^{-1}\|\|\mathcal{D}f(p)P_{c,p,p_0} - \mathcal{D}f(p_0)\| \le \sum_{k=1}^{\infty}(k+1)\gamma(f, p_0)^k\|v\|^k = \frac{1}{(1-u)^2} - 1 < 1. \tag{4.28}$$

Note that the convex process

$$W_pv = \mathcal{D}f(p)v - K = [\mathcal{D}f(p_0)P_{c,p_0,p} + (\mathcal{D}f(p) - \mathcal{D}f(p_0)P_{c,p_0,p})]v - K. \tag{4.29}$$

Since $P_{c,p_0,p}$ is an isomorphism from T_pM to $T_{p_0}M$ (cf. [16, p.30]), $W_{p_0} \circ P_{c,p_0,p}$ is a convex process from T_pM into $\mathbb{R}^n$ and $\mathcal{D}f(p) - \mathcal{D}f(p_0)P_{c,p_0,p}$ is a linear transformation from T_pM to $\mathbb{R}^n$. Moreover, $\|W_{p_0} \circ P_{c,p_0,p}\| = \|W_{p_0}\|$ and $\|\mathcal{D}f(p) - \mathcal{D}f(p_0)P_{c,p_0,p}\| = \|\mathcal{D}f(p)P_{c,p,p_0} - \mathcal{D}f(p_0)\|$ thanks to the fact that $P_{c,p_0,p}$ is an isometry (cf. [9, p.56]). Thus, by (4.28) and (4.29), Lemma 2.1 is applicable to concluding that W_p carries T_pM onto $\mathbb{R}^n$ and

$$\|W_p^{-1}\| \le \frac{\|W_{p_0}^{-1}\|}{1 - \|W_{p_0}^{-1}\|\|\mathcal{D}f(p)P_{c,p,p_0} - \mathcal{D}f(p_0)\|} \le \frac{\|W_{p_0}^{-1}\|(1-u)^2}{\psi(u)}.$$

To establish the second inequality of (4.25), let $k=2,3,\cdots$, and observe that

$$\|W_p^{-1}\|\frac{\|\mathcal{D}^k f(p)\|}{k!} \le \frac{(1-u)^2}{\psi(u)}\|W_{p_0}^{-1}\|\frac{\|\mathcal{D}^k f(p)\|}{k!}. \tag{4.30}$$

In view of (4.26) and (4.27), one can apply Proposition 4.2 (with $j=k$) to conclude that

$$\begin{aligned} \|W_{p_0}^{-1}\|\frac{\|\mathcal{D}^k f(p)\|}{k!} &= \|W_{p_0}^{-1}\|\left\|\frac{1}{k!}\sum_{j=0}^{\infty}\frac{1}{j!}\mathcal{D}^{k+j}f(p_0)v^jP^k_{c,p_0,p}\right\| \\ &\le \sum_{j=0}^{\infty}\frac{(k+j)!}{k!j!}\|W_{p_0}^{-1}\|\left\|\frac{\mathcal{D}^{k+j}f(p_0)}{(k+j)!}\right\|\|v\|^j \\ &\le \sum_{j=0}^{\infty}\frac{(k+j)!}{k!j!}\gamma(f,p_0)^{k+j-1}\|v\|^j. \end{aligned} \tag{4.31}$$

This, together with (4.22), implies that

$$\|W_{p_0}^{-1}\|\frac{\|\mathcal{D}^k f(p)\|}{k!} \le \frac{\gamma(f,p_0)^{k-1}}{(1-\gamma(f,p_0)\|v\|)^{k+1}}. \tag{4.32}$$

Combining this with the first inequality of (4.25) and (4.30) gives that

$$\|W_p^{-1}\|\frac{\|\mathcal{D}^k f(p)\|}{k!} \le \frac{1}{\psi(u)}\left(\frac{\gamma}{1-u}\right)^{k-1}, \tag{4.33}$$

thanks to (4.27). Consequently,

$$\gamma(f,p)=\sup_{k\ge 2}\left(\frac{\|W_{p_0}^{-1}\|\cdot\|\mathcal{D}^k f(p)\|}{k!}\right)^{\frac{1}{k-1}} \le \frac{\gamma(f,p_0)}{1-u}\sup_{k\ge 2}\frac{1}{\psi(u)^{\frac{1}{k-1}}} = \frac{\gamma(f,p_0)}{(1-u)\psi(u)},$$

where the last equality holds because the supremum attains at $k=2$ as $0<\psi(u)\le 1$ by (4.24). ■

Proposition 4.3. *Let $0<r\le\frac{2-\sqrt{2}}{2\gamma(f,p_0)}$. Then f satisfies the γ-condition at p_0 in $\mathbf{B}(p_0,r)$ with $\gamma=\gamma(f,p_0)$.*

Proof. Let $p,q\in\mathbf{B}(p_0,r)$. Let c_1 be a minimizing geodesic connecting p_0,p and c_2 a geodesic connecting p,q such that, for some $v_1\in T_{p_0}M$ and $v_2\in T_pM$, $p=\exp_{p_0}v_1$, $q=\exp_p v_2$, $l(c_1)=\|v_1\|=d(p_0,p)$, $l(c_2)=\|v_2\|$ with $l(c_1)+l(c_2)<r$. Write $u=l(c_1)\gamma(f,p_0)$. Then $u<1-\frac{\sqrt{2}}{2}$, and so Lemma 4.2 is applicable to concluding that

$$\gamma(f,p)\le\frac{\gamma(f,p_0)}{(1-u)\psi(u)}.$$

Hence,

$$\frac{1}{\gamma(f,p)} \geq \frac{(1-u)\psi(u)}{\gamma(f,p_0)}. \tag{4.34}$$

Since

$$\|v_1\| + \|v_2\| < r \leq \frac{1-\frac{\sqrt{2}}{2}}{\gamma(f,p_0)},$$

it follows from (4.34) that

$$\|v_2\| < \frac{1-\frac{\sqrt{2}}{2}}{\gamma(f,p_0)} - \|v_1\| = \frac{1-\frac{\sqrt{2}}{2} - \gamma(f,p_0)\|v_1\|}{\gamma(f,p_0)} \leq \frac{(1-u)\psi(u)}{\gamma(f,p_0)} \leq \frac{1}{\gamma(f,p)}.$$

Therefore, $q \in \mathbf{B}(p, \frac{1}{\gamma(\xi,p)})$ and it follows from Proposition 4.2 that

$$\begin{aligned} \mathcal{D}^2 f(q) &= \sum_{i=0}^{\infty} \frac{1}{i!} \mathcal{D}^{i+2} f(p) v_2^i P^2_{c_2,p,q} \\ &= \sum_{i=0}^{\infty} \frac{1}{i!} \sum_{j=0}^{\infty} \frac{1}{j!} \mathcal{D}^{j+i+2} f(p_0) v_1^j P^{i+2}_{c_1,p_0,p} v_2^i P^2_{c_2,p,q}. \end{aligned} \tag{4.35}$$

Since $\dfrac{\|W_{p_0}^{-1}\| \cdot \|\mathcal{D}^{j+i+2} f(p_0)\|}{(j+i+2)!} \leq \gamma(f,p_0)^{j+i+1}$, one has from (4.35) that

$$\|W_{p_0}^{-1}\| \cdot \|\mathcal{D}^2 f(q)\| \leq \sum_{i=0}^{\infty} \frac{1}{i!} \sum_{j=0}^{\infty} \frac{(j+i+2)!}{j!} \gamma(f,p_0)^{j+i+1} \|v_1\|^j \|v_2\|^i. \tag{4.36}$$

Using (4.22) to calculate the quantity on the right-hand side of the inequality (4.36), we obtain that

$$\|W_{p_0}^{-1}\| \cdot \|\mathcal{D}^2 f(q)\| \leq \frac{2\gamma}{(1-\gamma(\|v_1\| + \|v_2\|)^3} = \frac{2\gamma}{(1-\gamma(d(p_0,p) + l(c_2))^3}$$

and the proof is complete. ∎

The following corollary results from Proposition 4.3 and Corollary 4.2.

Corollary 4.3. *Let* $\gamma = \gamma(f,p_0)$. *Suppose that*

$$\xi = \|W_{p_0}^{-1}\| d(f(p_0), K) \leq \frac{3-2\sqrt{2}}{\gamma}.$$

Then the sequence $\{p_k\}$ *generated by Algorithm 1.1 with initial point* p_0 *is well-defined and converges to a solution* p^* *of the inclusion problem* (1.2). *Moreover, if* $\{v_k\}$ *and* $\{p_k\}$ *are sequences generated by* (1.3), *then* (4.18) *and* (4.19) *hold.*

4.4. Extension of Smale's approximate zeros

The following notion of approximate singular point was introduced in [23] for Newton's method on Riemannian manifold. Let ν be a section from a Riemannian manifold M to another one E. Recall that Newton's method for ν is defined as follows:

$$p_{k+1} = \exp_{p_k}(-\mathrm{D}\nu(p_k)^{-1}\nu(p_k)) \quad \text{for each } k = 0, 1, \cdots, \tag{4.37}$$

where D is a connection of ν.

Definition 4.2. Suppose that $p_0 \in M$ is such that the sequence $\{p_k\}$ generated by Newton's method (4.37) with initial point p_0 is well-defined for ν and satisfies

$$\Theta(p_k) \le \left(\frac{1}{2}\right)^{2^{k-1}} \Theta(p_{k-1}) \quad \text{for all } k = 1, 2, \cdots, \tag{4.38}$$

where $\Theta(p_k)$ denotes some measurement of the approximation degree between p_k and the singular point p^*. Then p_0 is said to be an approximate singular point of ν in the sense of $\Theta(p_k)$.

We now extend the notion of approximate singular point to the extended Newton's method for mappings on Riemannian manifolds with values in a cone.

Definition 4.3. Suppose that $p_0 \in M$ is such that the sequence $\{p_k\}$ generated by Algorithm 1.1 with initial point p_0 is well-defined, converges to a solution of the inclusion problem (1.2) and satisfies (4.38). Then p_0 is said to be an approximate solution of (1.2) in the sense of $\Theta(p_k)$.

Then we have the following corollary about approximate solution of (1.2).

Corollary 4.4. *Let $\gamma = \gamma(f, p_0)$. Suppose that*

$$\xi = \|W_{p_0}^{-1}\| d(f(p_0), K) \le \frac{13 - 3\sqrt{17}}{4\gamma}.$$

Then the sequence $\{p_k\}$ generated by Algorithm 1.1 with initial point p_0 is well-defined and converges to a solution p^* of the inclusion problem (1.2)*. Furthermore, if $\{v_k\}$ and $\{p_k\}$ are sequences generated by* (1.3)*, then p_0 is an approximate solution of* (1.2) *in the sense of $\|v_k\|$.*

Proof. Since $\frac{13-3\sqrt{17}}{4\gamma} < \frac{3-2\sqrt{2}}{\gamma}$, Corollary 4.3 is applicable to concluding that the sequence $\{p_k\}$ generated by Algorithm 1.1 with initial point p_0 is well-defined and converges to a solution p^* of the inclusion problem (1.2). Moreover, if $\{v_k\}$

and $\{p_k\}$ are sequences generated by (1.3), then (4.18) holds for η given by (4.16). Noting that η increases as ξ dose on $(0, \frac{13-3\sqrt{17}}{4\gamma}]$, and the value of η at $\xi = \frac{13-3\sqrt{17}}{4\gamma}$ is $\frac{1}{2}$. Hence, (4.18) holds for $\eta = \frac{1}{2}$. Therefore, p_0 is an approximate solution of (1.2) in the sense of $\|v_k\|$. ■

REFERENCES

1. R. Adler, J. P. Dedieu, J. Margulies, M. Martens and M. Shub, Newton's method on Riemannian manifolds and a geometric model for human spine, *IMA J. Numer. Anal.*, **22** (2002), 359-390.
2. L. Blum, F. Cucker, M. Shub and S. Smale, *Complexity and Real Computation*, New York: Springer-Verlag, 1997.
3. J. V. Burke, Second order necessary and sufficient conditions for convex composite NDO,*Math. Program.*, **38** (1987), 287-302.
4. J. V. Burke and M. C. Ferris, A Gauss-Newton method for convex composite optimization, *Math. Program. (Ser. A)*, **71** (1995), 179-194.
5. J. W. Daniel, Newton's method for nonlinear inequalities, *Numer. Math.*, **40** (1973), 381-387.
6. J. P. Dedieu, P. Priouret and G. Malajovich, Newton's method on Riemannian manifolds: covariant alpha theory , *IMA J. Numer. Anal.*, **23** (2003), 395-419.
7. J. E. Dennis, Jr. and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliks, NJ, 1983.
8. J. E. Dennis, Jr., E.A. Mahmoud and W. Karen, A trust-region approach to nonlinear systems of equalities and inequalities. *SIAM J. Optim.*, **9** (1999), 291-315.
9. M. P. DoCarmo, *Riemannian Geometry*, Birkhauser, Boston, 1992.
10. J. A. Ezquerro and M. A. Hern¨¢ndez, Generalized differentiability conditions for Newton′s method, *IMA J. Numer. Anal.*, **22** (2002), 187-205.
11. J. A. Ezquerro and M. A. Hern¨¢ndez, On an application of Newton′s method to nonlinear operators with w-conditioned second derivative, *BIT* **42** (2002), 519-530.
12. O. P. Ferreira and B. F. Svaiter, Kantorovich's Theorem on Newton's method in Riemannian manifolds, *J. Complexity*, **18** (2002), 304-329.
13. W. B. Gragg and R. A. Tapia, Optimal error bounds for the Newton-Kantorovich theorem, *SIAM J. Numer. Anal.*, **11** (1974), 10-13.
14. J. M. Gutiérrez, A new semilocal convergence theorem for Newton′s method, *J. Comput. Appl. Math.*, **79** (1997), 131-145.
15. J. M. Gutiérrez and M. A. Hernández, Newton′s method under weak Kantorovich conditions, *IMA J. Numer. Anal.*, **20** (2000), 521-532.

16. S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*, Academic Press, Inc, New York, 1978.

17. L. V. Kantorovich, On Newton method for functional equations, *Dokl. Acad. Nauk.* **59** (1948), 1237-1240.

18. L. V. Kantorovich and G. P. Akilov, *Functional Analysis*, Oxford: Pergamon, 1982.

19. C. Li, G. López and V. Mart´n-Márquez, Monotone vector fields and the proximal point algorithm on Hadamard manifolds, J. London Math. Soc., to appear

20. C.Li and K. F. Ng, Majorizing functions and convergence of the Gauss-Newton method for convex composite cptimization, *SIAM J. Optimization,* **18** (2007), 613-642.

21. C. Li and J. H. Wang, Convergence of the Newton method and uniqueness of zeros of vector fields on Riemannian manifolds, *Sci. China Ser. A.*, **48** (2005), 1465-1478.

22. C. Li and J. H. Wang, Newton′s method on Riemannian manifolds: Smale′s point estimate theory under the γ-condition, *IMA J. Numer. Anal*., **26** (2006), 228-251.

23. C. Li and J. H. Wang, Newton's Method for Sections on Riemannian Manifolds: Generalized Covariant α-Theory, *J of Complexity,* **24** (2008), 423-451.

24. C. Li and X. H. Wang, On convergence of the Gauss-Newton methods for convex composite optimization, *Math. Program. (Ser.A),* **91** (2002), 349-356.

25. S. L. Li, C. Li, Y. C. Liou and J. C. Yao, Existence of solutions for variational inequalities on Riemannian manifolds, submitted.

26. A. M. Ostrowski, *Solutions of Equations in Euclidean and Banach Spaces*, Academic Press, New York, 1973.

27. B. N. Pshenichnyi, A Newton method for the solution of systems of equalities and inequalities. (Russian) *Mat. Zametki*., **8** (1970), 635-640

28. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

29. R. T. Rockafellar, Monotone processes of convex and concave type, *Mem. Amer. Math. Soc. 77, AMS.*, Providence, R.I., 1967.

30. S. M. Robinson, Normed convex process, *Trans. Amer. Math. Soc.*, **174** (1972), 127-140.

31. S. M. Robinson, Extension of Newton's method to nonlinear functions with values in a cone, *Numer. Math.*, **19** (1972), 341-347.

32. S. Smale, Newton's method estimates from data at one point, The Merging of Disciplines: New Directions in Pure, Applied and Computational Mathematics (R.Ewing, K.Gross and C. Martin, eds), New York: Springer, (1986), 185-196.

33. S. T. Smith, Optimization techniques on Riemannian manifolds, in "Fields Institute Communications," Vol. 3, pp. 113-146, American Mathematical Society, Providence, RI, 1994.

34. C. Udriste, *Convex Functions and Optimization Methods on Riemannian Manifolds,* Mathematics and Its Applications, Vol. 297, Kluwer Academic, Dordrecht, 1994.

35. J. H. Wang and C. Li, Uniqueness of the singular points of vector fields on Riemannian manifolds under the γ-condition, *J. Complexity*, **22** (2006), 533-548.

36. X. H. Wang, Convergence of an iterative procedure, *KeXue TongBao*, **20** (1975), 558-559.

37. X. H. Wang, Convergence of Newton's method and inverse function theorem in Banach space, *Math. of Comput.*, **225** (1999), 169-186.

Jin-Hua Wang
Department of Mathematics,
Zhejiang University of Technology,
Hangzhou 310032,
P. R. China
E-mail: wjh@zjut.edu.cn

Shuechin Huang
Department of Applied Mathematics,
Dong Hwa University,
Hualien 97401, Taiwan
E-mail: shuang@mail.ndhu.edu.tw

Chong Li
Department of Mathematics,
Zhejiang University, Hangzhou 310027,
P. R. China
E-mail: cli@zju.edu.cn