J Glob Optim
DOI 10.1007/s10898-010-9638-1

Gauss–Newton method for convex composite optimizations on Riemannian manifolds

Jin-Hua Wang · **Jen-Chih Yao** · **Chong Li**

Received: 5 December 2010 / Accepted: 15 December 2010

Abstract A notion of quasi-regularity is extended for the inclusion problem $F(p) \in C$, where F is a differentiable mapping from a Riemannian manifold M to $\mathbb{R}^n$. When C is the set of minimum points of a convex real-valued function h on $\mathbb{R}^n$ and $\mathrm{D}F$ satisfies the L-average Lipschitz condition, we use the majorizing function technique to establish the semi-local convergence of sequences generated by the Gauss-Newton method (with quasi-regular initial points) for the convex composite function $h \circ F$ on Riemannian manifold. Two applications are provided: one is for the case of regularities on Riemannian manifolds and the other is for the case when C is a cone and $\mathrm{D}F(p_0)(\cdot) - C$ is surjective. In particular, the results obtained in this paper extend the corresponding one in Wang et al. (Taiwanese J Math 13:633–656, 2009).

Keywords Gauss–Newton method · Riemannian manifolds · L-average Lipschitz condition

---

Jin-Hua Wang was supported in part by the National Natural Science Foundation of China (grant 11001241).
Jen-Chih Yao was supported in part by the Grant NSC 99-2115-M-110-004-MY3.
Chong Li was supported in part by the National Natural Science Foundation of China (grant 10731060).

J.-H. Wang
Department of Mathematics, Zhejiang University of Technology,
Hangzhou 310032, People's Republic of China
e-mail: wjh@zjut.edu.cn

J.-C. Yao (✉)
Department of Mathematics, National Sun Yat-sen University, Kaohsiung, Taiwan
e-mail: yaojc@math.nsysu.edu.tw

C. Li
Department of Mathematics, Zhejiang University, Hangzhou 310027, People's Republic of China
e-mail: cli@zju.edu.cn

C. Li
Department of Mathematics, College of Sciences, King Saud University,
P. O. Box 2455, Riyadh 11451, Saudi Arabia

Mathematics Subject Classification (2000) 65H05 · 65J15

1 Introduction

Let F be a nonlinear Fréchet differentiable mapping from $\mathbb{R}^m$ to $\mathbb{R}^n$ and h be a real-valued convex function on $\mathbb{R}^n$. Consider the following convex composite optimization problem on $\mathbb{R}^m$:

$$\min_{x\in\mathbb{R}^m} f(x) := h(F(x)). \tag{1.1}$$

This problem has received a great deal of attention. A wide variety of its applications can be found throughout the mathematical programming literature especially in convex inclusion, minimax problems, penalization methods and goal programming [6,14,15,32,44]. As in [7,24,28], the study of (1.1) naturally relates to the convex inclusion problem

$$F(x) \in C, \tag{1.2}$$

where $C :=$ argmin h, the set of all minimum points of h. In [24], a notion of quasi-regularity for $x_0 \in \mathbb{R}^m$ with respect to the inclusion (1.2) was introduced. This new notion covers the case of regularity studied by Burke and Ferris [7] as well as the case when $F'(x_0)(\cdot) - C$ is surjective, presented by Robinson [41]. More importantly, in [24], the notions of the quasi-regular radius r_{x_0} and of the quasi-regular bound function β_{x_0} attached to each quasi-regular point x_0 were introduced, and it was established that if the initial point x_0 is a quasi-regular point with (r_{x_0}, β_{x_0}) and if F' satisfies a Lipschitz type condition, then the Gauss–Newton sequence $\{x_n\}$ provided by the following well-known Algorithm $\mathbf{GN}(\eta, \Delta, x_0)$ (cf. [7,21,28,55]) converges at a quadratic rate to some x^* with $F(x^*) \in C$ (in particular, x^* solves (1.1)).

Algorithm $\mathbf{GN}(\eta, \Delta, x_0)$. Let $\eta \geq 1$, $0 < \Delta \leq \infty$, and for each $x \in R^m$ define $\Psi_\Delta(x)$ by

$$\begin{aligned}\Psi_\Delta(x) := \{d \in \mathbb{R}^m \mid \|d\| \leq \Delta, h(F(x_k) + F'(x_k)d) \leq h(F(x_k)\\ +F'(x_k)d') \quad \forall d' \in \mathbb{R}^m \text{ with } \|d'\| \leq \Delta\}.\end{aligned}$$

Let $x_0 \in \mathbb{R}^m$ be given. For $k = 0, 1, \ldots$, having $x_0, x_1, \ldots, x_k$, determine x_{k+1} as follows.

If $0 \in \Psi_\Delta(x_k)$ then stop; if $0 \notin \Psi_\Delta(x_k)$, choose d_k such that $d_k \in \Psi_\Delta(x_k)$ and

$$\|d_k\| \leq \eta d(0, \Psi_\Delta(x_k)),$$

and set $x_{k+1} = x_k + d_k$, where $d(x, \Omega)$ denotes the distance from x to Ω in $\mathbb{R}^m$.

Note that $\Psi_\Delta(x)$ is nonempty and is the solution set of the following convex optimization problem

$$\min_{d\in\mathbb{R}^m, \|d\|\leq\Delta} h(F(x) + F'(x)d).$$

which can be solved by standard methods such as the subgradient method, the cutting plane method, the bundle method etc (cf. [20]).

Several problems related to numerical linear algebra can be expressed as optimizing a smooth function defined on a Riemannian manifold [3,36,37]. Applications appear in various areas, including computer vision [31], machine learning [34], maximum likelihood estimation [45,56], electronic structure computation [30], system balancing [19], model reduction [57], and robot manipulation [18]. A lot of algorithms such as steepest descent method, trust-region method, conjugate gradient method and so on have been extended to solve optimization problems on Riemannian manifolds (see, e.g., [1,23,46,47] and the references therein). Among these methods, Newton's method is one of the most powerful and simplest method. The convergence properties of Newton's method have been extensively explored in ([9,13,25–27,33,49] and the references therein). On the other hand, monotonicity notions in Banach spaces have been extended to Riemannian manifolds and a proximal-type method to find singular points has been developed for multivalued vector fields on Riemannian manifolds with nonpositive sectional curvatures, i.e., on Hadamard manifolds; see, e.g., [12,23] with other references. Furthermore, various derivative-like and subdifferential constructions for nondifferentiable functions on spaces with no linear structure are developed in ([2,11,22,29] and the references therein) and applied therein to the study of constrained optimization problems, nonclassical problems of the calculus of variations and optimal control, and generalized solutions to the first-order partial differential equations on Riemannian manifolds and other important classes of spaces with no linearity.

Motivated by what have been mentioned above, the purpose of the present paper is to extend the Gauss–Newton method to Riemannian manifold to solve the convex composite optimization on Riemannian manifold M which is formulated as follows:

$$\min_{p\in M} f(p) := h(F(p)), \tag{1.3}$$

where h is same as defined above and F is a differentiable mapping from M to $\mathbb{R}^n$. As mentioned before, the study of (1.3) naturally relates to the convex inclusion problem

$$F(p) \in C,$$

where $C = \operatorname{argmin} h$, the set of all minimum points of h. The extended Gauss–Newton method for convex composite optimization problem on Riemannian manifold (1.3) is defined as follows.

Algorithm R(η, Δ, p_0). Let $\eta \geq 1$, $0 < \Delta \leq \infty$. Let $p_0 \in M$ be given. For $k = 0, 1, \ldots$, having $p_0, p_1, \ldots, p_k$, determine p_{k+1} as follows.

If $0 \in \Lambda_\Delta(p_k)$ then stop; if $0 \notin \Lambda_\Delta(p_k)$, choose v_k such that $v_k \in \Lambda_\Delta(p_k)$ and

$$\|v_k\| \leq \eta d(0, \Lambda_\Delta(p_k)),$$

and set $p_{k+1} = \exp_{p_k} v_k$, where for each $p \in M$, $\Lambda_\Delta(p)$ is defined by

$$\Lambda_\Delta(p) := \{v \in T_pM \mid \|v\| \leq \Delta, h(F(p) + \mathrm{D}F(p)v) \leq h(F(p) \\ +\mathrm{D}F(p)v')\ \forall v' \in T_pM \text{ with } \|v'\| \leq \Delta\}.$$

In the present paper, the notions of the quasi-regularity for $p_0 \in M$, the quasi-regular radius r_{p_0} and the quasi-regular bound function β_{p_0} attached to quasi-regular point p_0 are extended to Riemannian manifolds. Our main results presented in section 3 show that if the initial point p_0 is a quasi-regular point with (r_{p_0}, β_{p_0}) and if $\mathrm{D}F$ satisfies a Lipschitz type condition, then the Gauss–Newton sequence $\{p_n\}$ generated by Algorithm $\mathbf{R}(\eta, \Delta, p_0)$ converges at a quadratic rate to some p^* with $F(p^*) \in C$ (in particular, p^* solves (1.3)). Furthermore, two applications to special cases are provided in section 4: one is for the case of regularities on Riemannian manifolds and the other is for the case when C is a cone and $\mathrm{D}F(p_0)(\cdot) - C$ is surjective. In particular, the results obtained in section 4 extend the corresponding one in [48].

2 Notions and preliminaries

The notations and notions about smooth manifolds used in the present paper are standard, see for example [10,17].

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 [10]), (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-functions 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 function such that

$$F = (F_1, F_2, \ldots, F_n)$$

with $F_i \in C^1(M)$ for each $i = 1, 2, \ldots, n$. Let ∇ be the Levi-Civita connection on M, and let $X \in C^1(TM)$. Following [9] (see also [27]), the derivative of F along the vector field X is defined by

$$\nabla_X F = (\nabla_X F_1, \nabla_X F_2, \ldots, \nabla_X F_n) = (X(F_1), X(F_2), \ldots, X(F_n)).$$

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

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

We use $\mathrm{D}F(p)$ 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

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

which only depends on the tangent vector v.

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.

3 Quasi-regularity and convergence criterion

The notion of the quasi-regular point was first introduced by Li and Ng in [24] for the study of convergence issue of convex composite optimization on $\mathbb{R}^n$. Below we extend this notion to Riemannian manifolds.

Let C be a closed convex set in $\mathbb{R}^n$. Consider the inclusion

$$F(p) \in C. \tag{3.1}$$

Let $p \in M$ and

$$\Lambda(p) := \{v \in T_pM |\ F(p) + \mathrm{D}F(p)v \in C\}. \tag{3.2}$$

Remark 3.1 In the case when C is the set of all minimum points of h and if there exists $v_0 \in T_pM$ with $\|v_0\| \leq \Delta$ such that $v_0 \in \Lambda(p)$, then $v_0 \in \Lambda_\Delta(p)$ and for each $v \in T_pM$ with $\|v\| \leq \Delta$ one has

$$v \in \Lambda_\Delta(p) \Longleftrightarrow v \in \Lambda(p) \Longleftrightarrow v \in \Lambda_\infty(p). \tag{3.3}$$

Definition 3.1 A point $p_0 \in M$ is called a quasi-regular point of the inclusion (3.1) if there exist $r > 0$ and an increasing positive-valued function β on $[0,r)$ such that

$$\Lambda(p) \neq \emptyset \quad \text{and} \quad d(0,\Lambda(p)) \leq \beta(d(p_0,p))\, d(F(p),C) \quad \text{for all } p \in \mathbf{B}(p_0,r). \tag{3.4}$$

Following [24], let $\mathbf{r}_{p_0}$ denote the supremum of r such that (3.4) holds for some increasing positive-valued function β on $[0,r)$. Let $r \in [0,\mathbf{r}_{p_0}]$ and let $\mathcal{B}_r(p_0)$ denote the set of all increasing positive-valued function β on $[0,r)$ such that (3.4) holds. Define

$$\beta_{p_0}(t) = \inf\{\beta(t) : \beta \in \mathcal{B}_{\mathbf{r}_{p_0}}(p_0)\} \quad \text{for each } t \in [0,\mathbf{r}_{p_0}). \tag{3.5}$$

Note that each $\beta \in \mathcal{B}_r(p_0)$ with $\lim_{t\to r^-}\beta(t) < +\infty$ can be extended to an element of $\mathcal{B}_{\mathbf{r}_{p_0}}(p_0)$. From this we can verify that

$$\beta_{p_0}(t) = \inf\{\beta(t) : \beta \in \mathcal{B}_r(p_0)\} \quad \text{for each } t \in [0,r). \tag{3.6}$$

We call $\mathbf{r}_{p_0}$ and β_{p_0} respectively the quasi-regular radius and the quasi-regular bound function of the quasi-regular point p_0.

Let L be a positive-valued increasing 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 [51] 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 [51]). Recently, this notion has been extended and applied to sections on Riemannian manifolds in [27]. Note that a mapping $F : M \to \mathbb{R}^n$ is a special example of sections. Definition 3.2 below is a slight modification of the corresponding one in [27] for mappings on Riemannian manifolds.

Definition 3.2 Let $r > 0$ and $p_0 \in M$. DF 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

$$\|\mathrm{D}F(p)P_{c,p,p_0} - \mathrm{D}F(p_0)\| \le \int_0^{l(c)} L(u)\mathrm{d}u; \tag{3.7}$$

(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

$$\|\mathrm{D}F(q)P_{c,q,p} - \mathrm{D}F(p)\| \le \int_{d(p_0,p)}^{d(p_0,p)+l(c)} L(u)\mathrm{d}u. \tag{3.8}$$

Obviously, the L-average Lipschitz condition implies the center L-average Lipschitz condition.

Let $\alpha > 0$, and let $r_\alpha > 0$ and $b_\alpha > 0$ be such that

$$\alpha \int_0^{r_\alpha} L(u)\ \mathrm{d}u = 1 \quad \text{and} \quad b_\alpha = \alpha \int_0^{r_\alpha} L(u)u\ \mathrm{d}u \tag{3.9}$$

(thus $b_\alpha < r_\alpha$). Let $\xi \ge 0$ and define

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

Thus

$$\phi'_\alpha(t) = -1 + \alpha \int_0^t L(u)\ \mathrm{d}u \text{for each } t \ge 0. \tag{3.11}$$

Let $t_{\alpha,n}$ denote the sequence generated by Newton's method for ϕ_α with the initial point $t_{\alpha,0} = 0$:

$$t_{\alpha,n+1} = t_{\alpha,n} - \phi'_\alpha(t_{\alpha,n})^{-1}\phi_\alpha(t_{\alpha,n}) \quad n = 0, 1, \ldots. \tag{3.12}$$

In particular, by (3.10) and (3.11),

$$t_{\alpha,1} = \xi. \tag{3.13}$$

Below we list a series of useful lemmas for our propose, where Lemma 3.1 is taken from [51] (see also [24]), while Lemmas 3.2 and 3.3 from [24].

Lemma 3.1 *Suppose that* $0 < \xi \le b_\alpha$. *Then* $b_\alpha < r_\alpha$ *and the following assertions hold:*

(i) ϕ_α *is strictly decreasing in* $[0, r_\alpha]$ *and strictly increasing in* $[r_\alpha, \infty)$ *with*

$$\phi_\alpha(\xi) > 0, \quad \phi_\alpha(r_\alpha) = \xi - b_\alpha \le 0, \quad \phi_\alpha(\infty) \ge \xi > 0. \tag{3.14}$$

Moreover, if $\xi < b_\alpha$, ϕ_α *has two zeros, denoted respectively by* r_α^* *and* r_α^{**}, *such that*

$$\xi < r_\alpha^* < \frac{r_\alpha}{b_\alpha}\xi < r_\alpha < r_\alpha^{**}, \tag{3.15}$$

and, if $\xi = b_\alpha$, ϕ_α *has a unique zero* r_α^* *in* (ξ, ∞) *(in fact* $r_\alpha^* = r_\alpha$*).*

(ii) $\{t_{\alpha,n}\}$ *is strictly monotonically increasing and converges to* r_α^*.

(iii) *The convergence of* $\{t_{\alpha,n}\}$ *is of quadratic rate if* $\xi < b_\alpha$, *and linear if* $\xi = b_\alpha$.

Lemma 3.2 *Let* r_α, b_α *and* ϕ_α *be defined by* (3.9) *and* (3.10). *Let* $\alpha' > \alpha$ *with the corresponding* $\phi_{\alpha'}$. *Then the following assertions hold.*

(i) *The functions* $\alpha \mapsto r_\alpha$ *and* $\alpha \mapsto b_\alpha$ *are strictly decreasing on* $(0, \infty)$.

(ii) $\phi_\alpha < \phi_{\alpha'}$ *on* $(0, \infty)$.

(iii) *The function* $\alpha \mapsto r_\alpha^*$ *is strictly increasing on the interval* $I(\xi)$, *where* $I(\xi)$ *denotes the set of all* $\alpha > 0$ *such that* $\xi \le b_\alpha$.

Lemma 3.3 *Define*

$$\omega_\alpha(t) = \phi_\alpha'(t)^{-1}\phi_\alpha(t) \qquad t \in [0, r_\alpha^*).$$

Suppose that $0 < \xi \le b_\alpha$. *Then* ω_α *is increasing on* $[0, r_\alpha^*)$.

We assume throughout the remainder of this paper that C is the set of all minimum points of h. Let $p_0 \in M$ be a quasi-regular point of the inclusion (3.1) with the quasi-regular radius $\mathbf{r}_{p_0}$ and the quasi-regular bound function β_{p_0}. Let $\eta \in [1, \infty)$ and let

$$\xi := \eta\beta_{p_0}(0)\, d(F(p_0), C). \tag{3.16}$$

For all $\mathbf{r} \in (0, \mathbf{r}_{p_0}]$, we define

$$\alpha_0(\mathbf{r}) := \sup\left\{\frac{\eta\beta_{p_0}(t)}{\eta\beta_{p_0}(t)\int_0^t L(u)\ \mathrm{d}u + 1} : \xi \le t < \mathbf{r}\right\} \tag{3.17}$$

with the usual convention that $\sup\emptyset = -\infty$.

Now, we are ready to give the main theorem of this section whose proof is a modification of the one for [24, Theorem 4.1] to suit the Riemannian manifold setting.

Theorem 3.1 *Let* $\eta \ge 1$ *and* $0 < \Delta \le \infty$. *Let* $p_0 \in M$ *be a quasi-regular point of the inclusion* (3.1) *with the quasi-regular radius* $\mathbf{r}_{p_0}$ *and the quasi-regular bound function* β_{p_0}. *Let* $\xi > 0$, $0 < \mathbf{r} \le \mathbf{r}_{p_0}$ *and* $\alpha_0(\mathbf{r})$ *be defined by* (3.17). *Let* $\alpha \ge \alpha_0(\mathbf{r})$ *be a positive constant and let* b_α, r_α *be defined by* (3.9). *Let* r_α^* *denote the smaller zero of the function* ϕ_α *defined by* (3.10). *Suppose that* $\mathrm{D}F$ *satisfies the* L*-average Lipschitz condition on* $\mathbf{B}(p_0, r_\alpha^*)$, *and that*

$$\xi \le \min\{b_\alpha, \Delta\} \quad \textit{and} \quad r_\alpha^* \le \mathbf{r}. \tag{3.18}$$

Let $\{p_n\}$ *denote the sequence generated by Algorithm* $\mathbf{R}(\eta, \Delta, p_0)$. *Then,* $\{p_n\}$ *converges to some* p^* *such that* $F(p^*) \in C$, *and the following assertions hold for each* $n = 0, 1, \ldots$*:*

$$d(p_n, p^*) \le r_\alpha^* - t_{\alpha,n} \tag{3.19}$$

and

$$F(p_n) + \mathrm{D}F(p_n)v_n \in C. \tag{3.20}$$

Proof To complete the proof, it's sufficient to prove that for each $n = 0, 1, \ldots$:

$$d(p_n, p_{n+1}) \le \|v_n\| \le t_{\alpha,n+1} - t_{\alpha,n} \tag{3.21}$$

and (3.20) holds. Granting this, $\{p_n\}$ is a Cauchy sequence by the monotonicity of $\{t_n\}$ and hence converges to some p^* such that $F(p^*) \in C$, and (3.19) holds.

We proceed by mathematical induction. By (3.18), (3.13) and Lemma 3.1, one has that, for each n,

$$\xi \leq t_{\alpha,n} < r_\alpha^* \leq \mathbf{r} \leq \mathbf{r}_{x_0}. \tag{3.22}$$

By the quasi-regularity assumption, it follows that

$$\Lambda(p) \neq \emptyset \quad \text{and} \quad d(0, \Lambda(p)) \leq \beta_{p_0}(d(p_0, p))\, d(F(p), C) \quad \text{for each } p \in \mathbf{B}(p_0, \mathbf{r}). \tag{3.23}$$

In particular, $\Lambda(p_0) \neq \emptyset$ and

$$\eta d(0, \Lambda(p_0)) \leq \eta\beta_{p_0}(d(p_0, p_0))\, d(F(p_0), C) = \eta\beta_{p_0}(0) d(F(p_0), C) = \xi \leq \Delta. \tag{3.24}$$

Since $\eta \geq 1$, it follows that $d(0, \Lambda(p_0)) \leq \Delta$ and so there exists $v \in T_{p_0}M$ with $\|v\| \leq \Delta$ such that $F(p_0) + \mathrm{D}F(p_0)v \in C$. Consequently, by Remark 3.1,

$$\Lambda_\Delta(p_0) = \{v \in T_{p_0}M \| \|v\| \leq \Delta,\ \ F(p_0) + \mathrm{D}F(p_0)v \in C\}$$

and

$$d(0, \Lambda_\Delta(p_0)) = d(0, \Lambda(p_0)).$$

Then, by the definition of Algorithm $\mathbf{R}(\eta, \Delta, p_0)$ and (3.24), we can choose $v_0 \in \Lambda_\Delta(p_0)$ such that

$$\|v_0\| \leq \eta d(0, \Lambda_\Delta(p_0)) = \eta d(0, \Lambda(p_0)) \leq \eta\beta_{p_0}(0) d(F(p_0), C) = \xi = t_{\alpha,1} - t_{\alpha,0},$$

which implies that (3.21) and (3.20) hold for $n = 0$. Assume that (3.21) and (3.20) hold for each $n \in \{0, 1, \ldots, k-1\}$ and define the geodesic $c : [0, 1] \to M$ by

$$c(\tau) = \exp_{p_{k-1}} \tau v_{k-1} \quad \tau \in [0, 1]. \tag{3.25}$$

Note that

$$d(p_0, p_k) \leq \sum_{i=1}^{k} d(p_i, p_{i-1}) \leq \sum_{i=1}^{k} (t_{\alpha,i} - t_{\alpha,i-1}) = t_{\alpha,k} \tag{3.26}$$

and

$$d(p_{k-1}, p_0) \leq t_{\alpha,k-1} \leq t_{\alpha,k}. \tag{3.27}$$

It follows from (3.25) and (3.22) that $c(\tau) \in \mathbf{B}(p_0, r_\alpha^*) \subseteq \mathbf{B}(p_0, \mathbf{r})$ for each $\tau \in [0, 1]$. Hence (3.23) holds for $p = p_k$, namely,

$$\Lambda(p_k) \neq \emptyset \quad \text{and} \quad d(0, \Lambda(p_k)) \leq \beta_{p_0}(d(p_k, p_0))\, d(F(p_k), C). \tag{3.28}$$

We claim that

$$\eta d(0, \Lambda(p_k)) \leq t_{\alpha,k+1} - t_{\alpha,k}. \tag{3.29}$$

To do this, using (3.20) for $n = k - 1$ and the fact that $\mathrm{D}F$ satisfies L-average Lipschitz condition on $\mathbf{B}(p_0, r_\alpha^*)$, we have by (3.28) that

$$\begin{aligned}\eta d(0, \Lambda(p_k)) &\le \eta\beta_{p_0}(d(p_0, p_k))d(F(p_k), C)\\ &\le \eta\beta_{p_0}(d(p_0, p_k))\|F(p_k) - F(p_{k-1}) - \mathrm{D}F(p_{k-1})v_{k-1}\|\\ &\le \eta\beta_{p_0}(d(p_0, p_k))\left\|\int_0^1 (\mathrm{D}F(c(\tau))P_{c,c(\tau),p_{k-1}} - \mathrm{D}F(p_{k-1}))v_{k-1}\ \mathrm{d}\tau\right\|\\ &\le \eta\beta_{p_0}(d(p_0, p_k))\int_0^1\left(\int_{d(p_{k-1},p_0)}^{\tau\|v_{k-1}\|+d(p_{k-1},p_0)} L(u)\ \mathrm{d}u\right)\|v_{k-1}\|\ \mathrm{d}\tau\\ &= \eta\beta_{p_0}(d(p_0, p_k))\left(\int_0^{\|v_{k-1}\|} L(d(p_{k-1}, p_0) + u)(\|v_{k-1}\| - u)\ \mathrm{d}u\right)\\ &\le \eta\beta_{p_0}(t_{\alpha,k})\left(\int_0^{\|v_{k-1}\|} L(t_{\alpha,k-1} + u)(\|v_{k-1}\| - u)\ \mathrm{d}u\right),\end{aligned}$$

where the last inequality is valid because L and β_{x_0} are increasing and thanks to (3.26) and (3.27). Hence, it follows that

$$\begin{aligned}\eta d(0, \Lambda(p_k)) &\le \eta\beta_{p_0}(t_{\alpha,k})\int_0^{t_{\alpha,k}-t_{\alpha,k-1}} L(t_{\alpha,k-1} + u)(t_{\alpha,k} - t_{\alpha,k-1} - u)\mathrm{d}u\\ &= \frac{\eta\beta_{p_0}(t_{\alpha,k})\phi_\alpha(t_{\alpha,k})}{\alpha}.\end{aligned} \tag{3.30}$$

On the other hand, by (3.22) and (3.17),

$$\frac{\eta\beta_{p_0}(t_{\alpha,k})}{\alpha_0(\mathbf{r})} \le \left(1 - \alpha_0(\mathbf{r})\int_0^{t_{\alpha,k}} L(u)\,\mathrm{d}u\right)^{-1}.$$

Since $\alpha \ge \alpha_0(\mathbf{r})$ and by (3.11), it follows that

$$\frac{\eta\beta_{p_0}(t_{\alpha,k})}{\alpha} \le \left(1 - \alpha\int_0^{t_{\alpha,k}} L(u)\ \mathrm{d}u\right)^{-1} = -(\phi_\alpha'(t_{\alpha,k}))^{-1}. \tag{3.31}$$

Combining (3.30) and (3.31) together with (3.12), (3.29) is seen to hold. Moreover by Lemma 3.3 and (3.18), we have

$$t_{\alpha,k+1} - t_{\alpha,k} = -\phi_\alpha'(t_{\alpha,k})^{-1}\phi_\alpha(t_{\alpha,k}) \le -\phi_\alpha'(t_{\alpha,0})^{-1}\phi_\alpha(t_{\alpha,0}) = \xi \le \Delta,$$

so (3.29) implies that $d(0, \Lambda(p_k)) \le \Delta$. Hence there exists $v \in T_{p_k}M$ with $\|v\| \le \Delta$ such that $F(p_k) + \mathrm{D}F(p_k)v \in C$. Consequently, by Remark 3.1,

$$\Lambda_\Delta(p_k) = \{v \in T_{p_k}M \mid \|v\| \le \Delta,\ F(p_k) + \mathrm{D}F(p_k)v \in C\}$$

and

$$d(0, \Lambda_\Delta(p_k)) = d(0, \Lambda(p_k)).$$

Choosing $v_k \in \Lambda_\Delta(p_k)$ according to Algorithm $\mathbf{R}(\eta, \Delta, p_0)$, it follows that (3.20) holds for $n = k$. Furthermore, one has that

$$\|v_k\| \leq \eta d(0, \Lambda_\Delta(p_k)) = \eta d(0, \Lambda(p_k)).$$

This with (3.29) yields that (3.21) holds for $n = k$. The proof of the theorem is complete. □

Remark 3.2 A more convenient condition than (3.18) is

$$\xi \leq \min\{b_\alpha, \ \frac{b_\alpha}{r_\alpha}\mathbf{r}, \ \Delta\}. \tag{3.32}$$

In fact, by (3.15) and (3.32), we have

$$r_\alpha^* \leq \frac{r_\alpha}{b_\alpha}\xi \leq \mathbf{r}. \tag{3.33}$$

Hence (3.32) $\Longrightarrow$ (3.18).

Remark 3.3 Let $0 < \mathbf{r} \leq \mathbf{r}_{p_0}$, $\beta_0 > 0$ and $0 \leq \beta \leq \beta_0$. Below we consider one important class of quasi-regular bound function β_{p_0} satisfying

$$\beta_{p_0}(t) \leq \frac{\beta_0}{1 - \beta \int\limits_0^t L(u)\mathrm{d}u} \quad \text{for each } t \in [0, \mathbf{r}). \tag{3.34}$$

Then, we have

$$\alpha_0(\mathbf{r}) \leq \frac{\eta\beta_0}{1 + (\eta\beta_0 - \beta)\int\limits_0^\xi L(u)\mathrm{d}u}. \tag{3.35}$$

In fact by (3.34), for each $t \in [\xi, \mathbf{r})$, we have

$$\eta\int\limits_0^t L(u)\ \mathrm{d}u + \frac{1}{\beta_{x_0}(t)} \geq \frac{1}{\beta_0} + \left(\eta - \frac{\beta}{\beta_0}\right)\int\limits_0^t L(u)\ \mathrm{d}u \geq \frac{1}{\beta_0} + \left(\eta - \frac{\beta}{\beta_0}\right)\int\limits_0^\xi L(u)\ \mathrm{d}u,$$

that is,

$$\frac{\beta_{x_0}(t)}{1 + \eta\beta_{x_0}(t)\int\limits_0^t L(u)\mathrm{d}u} \leq \frac{\beta_0}{1 + (\eta\beta_0 - \beta)\int\limits_0^\xi L(u)\mathrm{d}u}.$$

Hence, (3.35) follows by the definition of $\alpha_0(\mathbf{r})$ in (3.17).

We specialize our results in two important cases: the classical Lipschitz condition and the γ-condition, which have been used extensively in the study of convergence of Newton's method both in Banach spaces and Riemannian manifolds, see for example [25–27,48,49,51–54].

First, we consider the classical Lipschitz condition.

Corollary 3.1 *Let* $\eta \geq 1$, $\beta_0 > 0$ *and* $0 < \Delta \leq \infty$. *Let* $0 \leq \beta \leq \beta_0$.

Let $p_0 \in M$ *be a quasi-regular point of the inclusion* (3.1) *with the quasi-regular radius* $\mathbf{r}_{p_0}$ *and the quasi-regular bound function* β_{p_0}. *Let* $0 < \mathbf{r} \leq \mathbf{r}_{p_0}$. *Suppose that*

$$\beta_{p_0}(t) \leq \frac{\beta_0}{1-\beta L t} \quad \text{for each } t \in [0, \mathbf{r}).$$

Let $\xi = \eta\beta d(F(p_0), C)$,

$$R^* = \frac{1+(\eta\beta_0-\beta)L\xi - \sqrt{[1+(\eta\beta_0-\beta)L\xi][1-(\eta\beta_0+\beta)L\xi]}}{L\eta\beta_0} \tag{3.36}$$

and

$$Q = \frac{1-\beta L\xi - \sqrt{[1+(\eta\beta_0-\beta)L\xi][1-(\eta\beta_0+\beta)L\xi]}}{L\eta\beta_0\xi}.$$

Suppose that $\mathrm{D}F$ *is Lipschitz continuous on* $\mathbf{B}(p_0, R^*)$ *with modulus L, that is,*

$$\|\mathrm{D}F(p_2)P_{c,p_2,p_1} - \mathrm{D}F(p_1)\| \leq L d(p_1, p_2) \tag{3.37}$$

holds for each $p_1, p_2 \in \mathbf{B}(p_0, R^*)$ *and any geodesic c connecting* p_1, p_2. *Suppose that*

$$\xi \leq \min\left\{\frac{1}{L\beta_0\eta + L\beta}, \; \Delta\right\} \quad \text{and} \quad \mathbf{r} \geq R^*. \tag{3.38}$$

Let $\{p_n\}$ *denote the sequence generated by Algorithm* $\mathbf{R}(\eta, \Delta, p_0)$. *Then* $\{p_n\}$ *converges to some* p^* *with* $F(p^*) \in C$ *and the following assertion holds:*

$$d(p_n, p^*) \leq \frac{Q^{2^n-1}}{\sum_{i=0}^{2^n-1} Q^i} R^* \quad \text{for each } n = 0, 1, \ldots. \tag{3.39}$$

Proof Let $\alpha = \frac{\eta\beta_0}{1+(\eta\beta_0-\beta)L\xi}$. Let $\alpha_0(\mathbf{r})$ be defined by (3.17) with $L(u) \equiv L$. Then, we have from (3.35) that

$$\alpha \geq \frac{\eta\beta_{p_0}(t)}{1+\eta\beta_{p_0}(t)Lt} \quad \text{for each } t \in [\xi, r).$$

Hence $\alpha \geq \alpha_0(\mathbf{r})$ by (3.17). Noting that $L(u) \equiv L$, and by (3.9), (3.10), we have that,

$$r_\alpha = \frac{1}{\alpha L}, \quad b_\alpha = \frac{1}{2\alpha L} \tag{3.40}$$

and

$$\phi_\alpha(t) = \xi - t + \frac{\alpha L}{2} t^2 \quad \text{for each } t \geq 0.$$

Moreover, if $\xi \leq \frac{1}{2\alpha L}$, then the zeros of ϕ_α are given by

$$\left.\begin{matrix} r_\alpha^* \\ r_\alpha^{**} \end{matrix}\right\} = \frac{1 \mp \sqrt{1-2\alpha L\xi}}{\alpha L}. \tag{3.41}$$

It is also known (see for example [16,35,50]) that $\{t_{\alpha,n}\}$ has the closed form

$$t_{\alpha,n} = \frac{1-q_\alpha^{2^n-1}}{1-q_\alpha^{2^n}} r_\alpha^* \quad \text{for each } n = 0, 1, \ldots, \tag{3.42}$$

where

$$q_\alpha := \frac{r_\alpha^*}{r_\alpha^{**}} = \frac{1-\sqrt{1-2\alpha L\xi}}{1+\sqrt{1-2\alpha L\xi}}. \tag{3.43}$$

Moreover, by (3.40)–(3.43), one has that

$$r_\alpha^* = R^*, \quad q_\alpha = Q, \quad r_\alpha = \frac{1+(\eta\beta_0-\beta)L\xi}{L\eta\beta_0}, \quad b_\alpha = \frac{1+(\eta\beta_0-\beta)L\xi}{2L\eta\beta_0} \tag{3.44}$$

and

$$t_{\alpha,n} = \frac{1-Q^{2^n-1}}{1-Q^{2^n}} R^*. \tag{3.45}$$

Hence condition (3.18) is equivalent to the three inequalities together:

$$\xi \le \frac{1+(\eta\beta_0-\beta)L\xi}{2L\eta\beta_0}, \quad \xi \le \Delta \quad \text{and} \quad r_\alpha^* \le \mathbf{r}$$

and is hence, by (3.44), also equivalent to condition (3.38). Thus we apply Theorem 3.1 to conclude that the sequence $\{p_n\}$ converges to some p^* with $F(p^*) \in C$ and, for each $n = 1, 2, \ldots,$

$$d(p_n, p^*) \le r_\alpha^* - t_{\alpha,n}.$$

Noting, by (3.44) and (3.45), that

$$r_\alpha^* - t_{\alpha,n} = \left(1 - \frac{1-Q^{2^n-1}}{1-Q^{2^n}}\right) R^* = \frac{Q^{2^n-1}}{\sum_{i=0}^{2^n-1} Q^i} R^*,$$

it follows that (3.39) holds and the proof is complete. □

Let k, κ be positive integers such that $k \le \kappa$. Let $F : M \to \mathbb{R}^n$ be a C^κ-mapping. Following [27] (see also [9,48]), we 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 mapping $\mathrm{D}^1 F = \mathrm{D}F : (C^\kappa(TM))^1 \to C^{\kappa-1}(M)$ is defined by

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

Define the mapping $\mathrm{D}^k F : (C^\kappa(TM))^k \to C^{\kappa-k}(M)$ by

$$\begin{aligned} &\mathrm{D}^k F(X_1, \ldots, X_{k-1}, X_k) \\ &\quad = \nabla_{X_k}(\mathrm{D}^{k-1} F(X_1, \ldots, X_{k-1})) - \sum_{i=1}^{k-1} \mathrm{D}^{k-1} F(X_1, \ldots, \nabla_{X_k} X_i, \ldots, X_{k-1}) \end{aligned}$$

for each $X_1, \ldots, X_{k-1}, X_k \in C^\kappa(TM)$.

Then, one can use mathematical induction to prove easily that $\mathrm{D}^k F(X_1, \ldots, X_k)$ 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 $\mathrm{D}^k F(X_1, \ldots, X_k)$ at $p \in M$ only depends on the k-tuple of tangent vectors $(v_1, \ldots, v_k) = (X_1(p), \ldots, X_k(p)) \in (T_p M)^k$. Consequently, for a given $p \in M$, the map $\mathrm{D}^k F(p) : (T_p M)^k \to \mathbb{R}^n$, defined by

$$\mathrm{D}^k F(p) v_1 \ldots v_k := \mathrm{D}^k F(X_1, \ldots, X_k)(p) \quad \text{for any } (v_1, \ldots, v_k) \in (T_p M)^k,$$

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

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

Definition 3.3 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

$$\|\mathrm{D}^2 F(q)\| \le \frac{2\gamma}{(1 - \gamma(d(p_0, p) + l(c)))^3}. \tag{3.46}$$

Let L be the function defined by

$$L(u) = \frac{2\gamma}{(1-\gamma u)^3} \quad \text{for each } u \text{ with } 0 \le u < \frac{1}{\gamma}. \tag{3.47}$$

The following proposition shows that the γ-condition implies the L-average Lipschitz condition. Noting that the mapping F is a special example of the section, so Proposition 3.1 below is a direct consequence of [27, Proposition 5.1] (see also [48] for a direct proof).

Proposition 3.1 *Suppose that F satisfies the γ-condition at p_0 in $\mathbf{B}(p_0, r)$. Then $\mathrm{D}F$ satisfies the L-average Lipschitz condition in $\mathbf{B}(p_0, r)$ with L given by* (3.47).

According to the L defined by (3.47), one has from (3.9), (3.10) and elementary calculation (cf. [54]) that for all $\alpha > 0$,

$$r_\alpha = \left(1 - \sqrt{\frac{\alpha}{1+\alpha}}\right)\frac{1}{\gamma}, \quad b_\alpha = \left(1 + 2\alpha - 2\sqrt{\alpha(1+\alpha)}\right)\frac{1}{\gamma} \tag{3.48}$$

and

$$\phi_\alpha(t) = \xi - t + \frac{\alpha\gamma t^2}{1-\gamma t} \quad \text{for each } t \text{ with } 0 \le t < \frac{1}{\gamma}. \tag{3.49}$$

Thus, from [51] (see also [24]), we have the following lemma.

Lemma 3.4 *Let $\alpha > 0$. Assume that $\xi \le b_\alpha$, namely,*

$$\gamma\xi \le 1 + 2\alpha - 2\sqrt{\alpha(1+\alpha)}. \tag{3.50}$$

Then the following assertions hold:

(i) *ϕ_α has two zeros given by*

$$\left.\begin{matrix} r_\alpha^* \\ r_\alpha^{**} \end{matrix}\right\} = \frac{1 + \gamma\xi \mp \sqrt{(1+\gamma\xi)^2 - 4(1+\alpha)\gamma\xi}}{2(1+\alpha)\gamma}; \tag{3.51}$$

(ii) *the sequence $\{t_{\alpha,n}\}$ generated by Newton's method for ϕ_α with the initial point $t_{\alpha,0} = 0$ has the closed form:*

$$t_{\alpha,n} = \frac{1 - q_\alpha^{2^n-1}}{1 - q_\alpha^{2^n-1} p_\alpha} r_\alpha^* \quad \textit{for each } n = 0, 1, \dots, \tag{3.52}$$

where

$$q_\alpha := \frac{1 - \gamma\xi - \sqrt{(1+\gamma\xi)^2 - 4(1+\alpha)\gamma\xi}}{1 - \gamma\xi + \sqrt{(1+\gamma\xi)^2 - 4(1+\alpha)\gamma\xi}} \quad \textit{and}$$

$$p_\alpha := \frac{1 + \gamma\xi - \sqrt{(1+\gamma\xi)^2 - 4(1+\alpha)\gamma\xi}}{1 + \gamma\xi + \sqrt{(1+\gamma\xi)^2 - 4(1+\alpha)\gamma\xi}}. \tag{3.53}$$

Corollary 3.2 *Let $\eta \geq 1$, $\beta_0 > 0$ and $0 < \Delta \leq \infty$. Let $0 \leq \beta \leq \beta_0$. Let $p_0 \in M$ be a quasi-regular point of the inclusion (3.1) with the quasi-regular radius $\mathbf{r}_{p_0}$ and the quasi-regular bound function β_{p_0}. Let $0 < \mathbf{r} \leq \mathbf{r}_{p_0}$. Suppose that*

$$\beta_{p_0}(t) \leq \frac{\beta_0(1-\gamma t)^2}{(1+\beta)(1-\gamma t)^2 - \beta} \quad \textit{for each } t \in [0, \mathbf{r}).$$

Let $\xi = \eta\beta_0 d(F(p_0), C)$ and

$$\alpha = \frac{\eta\beta_0(1-\gamma\xi)^2}{[1-(\eta\beta_0-\beta)](1-\gamma\xi)^2 + (\eta\beta_0-\beta)}. \tag{3.54}$$

Set

$$r_\alpha^* = \frac{1+\gamma\xi-\sqrt{(1+\gamma\xi)^2-4(1+\alpha)\gamma\xi}}{2(1+\alpha)\gamma} \quad \textit{and}$$
$$q_\alpha = \frac{1-\gamma\xi-\sqrt{(1+\gamma\xi)^2-4(1+\alpha)\gamma\xi}}{1-\gamma\xi+\sqrt{(1+\gamma\xi)^2-4(1+\alpha)\gamma\xi}}.$$

Suppose that F satisfies the γ-condition at p_0 in $\mathbf{B}(p_0, r_\alpha^*)$, and that

$$\xi \leq \min\left\{\frac{1+2\alpha-2\sqrt{\alpha(1+\alpha)}}{\gamma}, \ \Delta\right\} \quad \textit{and } r_\alpha^* \leq \mathbf{r}. \tag{3.55}$$

Let $\{p_n\}$ denote the sequence generated by Algorithm $\mathbf{R}(\eta, \Delta, p_0)$. Then, $\{p_n\}$ converges to some p^* such that $F(p^*) \in C$, and the following assertion holds for each $n = 0, 1, \ldots$:

$$d(p_n, p^*) \leq {q_\alpha}^{2^n-1} r_\alpha^*. \tag{3.56}$$

Proof By assumption and Proposition 3.1, we have that $\mathrm{D}F$ satisfies the L-average Lipschitz condition in $\mathbf{B}(p_0, r_\alpha^*)$ with L given by (3.47). According to the L given by (3.47), we have $\int_0^\xi L(u)\,\mathrm{d}u = (1-\gamma\xi)^{-2} - 1$ and so that the α given by (3.54) satisfies $\alpha \geq \alpha_0(\mathbf{r})$ thanks to (3.35). Furthermore, corresponding to the L given by (3.47), it follows from (3.48) that (3.55) is equivalent to (3.18). Thus, Theorem 3.1 is applicable to concluding that $\{p_n\}$ converges to some p^* such that $F(p^*) \in C$, and $d(p_n, p^*) \leq r_\alpha^* - t_{\alpha,n}$ holds for each $n = 0, 1, \ldots$. Noting that by (3.52), one has

$$r_\alpha^* - t_{\alpha,n} = \frac{q_\alpha^{2^n-1}(1-p_\alpha)}{1-q_\alpha^{2^n-1}p_\alpha} r_\alpha^* \leq {q_\alpha}^{2^n-1} r_\alpha^*.$$

Hence (3.56) follows from (3.19). The proof of the corollary is complete. □

We end this section with a conclusion remark.

Remark 3.4 It is known in [27] that if F is analytic, then F satisfies the γ-condition with $\gamma = \sup_{k\geq 2}\left\|\frac{\mathrm{D}^k F(p)}{k!}\right\|_p^{\frac{1}{k-1}}$. Hence, we can also use the standard technique to obtain the Smale's type theory and establish the theory about the Smale's approximate zeros as that in [27]. However, the technique is very similar and so is omitted here.

4 Applications

This section is devoted to two applications: one is for the case of regularities on Riemannian manifolds and the other is for the case when C is a cone and $\mathrm{D}F(p_0)(\cdot) - C$ is surjective. In particular, the results obtained in this section extends the corresponding one in [48].

Let $\Omega \subset \mathbb{R}^n$ and write

$$\Omega^{\ominus} = \{z \in \mathbb{R}^n \,|\, \langle w, z \rangle \le 0 \quad \text{for each } w \in \Omega\}.$$

Let $A_0 \in \mathbb{R}^{n\times m}$. We use $\ker(A_0)$ to denote the kernel of A_0. Recall that $F : M \to \mathbb{R}^n$ is a C^1 mapping. The following definition extends the notion of regular point in [4,5,38,39,44] to the Riemannian manifold setting.

Definition 4.1 A point $p_0 \in M$ is a regular point of the inclusion (3.1) if

$$\ker(\mathrm{D}F(p_0)^*) \cap (C - F(p_0))^{\ominus} = \{0\}, \tag{4.1}$$

where $\mathrm{D}F(p_0)^*$ is the conjugate operator of $\mathrm{D}F(p_0)$.

Proposition 4.1 below extends the corresponding one in [8] to Riemannian manifolds. First, we need the following lemma which is a consequence of [8, Lemma 3.2 and Proposition 3.3].

Lemma 4.1 *Let $a_0 \in \mathbb{R}^n$ and $A_0 \in \mathbb{R}^{n\times m}$. If $\ker(A_0^T) \cap (C - a_0)^{\ominus} = \{0\}$. Then there exist $\beta_0 > 0$ and open neighborhoods $U_1 \subset \mathbb{R}^n$, $U_2 \subset \mathbb{R}^{n\times m}$ with $a_0 \in U_1$ and $A_0 \in U_2$ such that for each $(a, A) \in U_1 \times U_2$, we have*

$$\bar{\Lambda}(a, A) := \{u \in \mathbb{R}^m \,|\, a + Au \in C\} \ne \emptyset \text{ and } d(0, \bar{\Lambda}(a, A)) \le \beta_0 d(a, C). \tag{4.2}$$

Proposition 4.1 *Let p_0 be a regular point of (3.1). Then there are constants $r > 0$ and $\beta_0 > 0$ such that (3.4) holds for r and $\beta(\cdot) = \beta_0$; consequently, p_0 is a quasi-regular point with the quasi-regular radius $\mathbf{r}_{p_0} \ge r$ and the quasi-regular bound function $\beta_{p_0}(\cdot) \le \beta_0$ on $[0, r]$.*

Proof Let p_0 be a regular point of (3.1). Let $\{e_1, \dots, e_m\}$ be an orthonormal basis of $T_{p_0}M$. Let $p \in M$. Set

$$e_i^p = P_{p,p_0} e_i \quad \text{for each } i = 1, \dots, m.$$

Then it is easy to prove (see [10]) that $\{e_1^p, \dots, e_m^p\}$ is an orthonormal basis of T_pM. The mapping $A : M \to \mathbb{R}^{n\times m}$ is defined for each $p \in M$ by

$$A(p)u := \mathrm{D}F(p)\left(\sum_{i=1}^{m} u_i e_i^p\right) \quad \forall u = (u_1, \dots, u_m)^T \in \mathbb{R}^m. \tag{4.3}$$

Then, by definition, it follows that

$$\ker(\mathrm{D}F(p_0)^*) = \ker(A(p_0)^T). \tag{4.4}$$

Since p_0 is a regular point of (3.1), we have from (4.1) and (4.4) that

$$\ker(A(p_0)^T) \cap (C - F(p_0))^{\ominus} = \{0\}.$$

Thus, Lemma 4.1 is applicable to concluding that there exist $\beta_0 > 0$ and open neighborhoods $U_1 \subset \mathbb{R}^n$, $U_2 \subset \mathbb{R}^{n\times m}$ with $F(p_0) \in U_1$ and $A(p_0) \in U_2$ such that for each $(a, A) \in U_1 \times U_2$, (4.2) holds. Since F is a C^1 mapping, and for each $i = 1, \dots, m$, the mapping $p \to e_i^p$ is continuous (cf. [23, Lemma 2.4]), one has that $A(\cdot)$ is continuous. Hence, there exists $r > 0$ such that $A(\mathbf{B}(p_0, r)) \subset U_2$ and $F(\mathbf{B}(p_0, r)) \subset U_1$. Then for each $p \in \mathbf{B}(p_0, r)$, we have $(F(p), A(p)) \in U_1 \times U_2$ and so (4.2) holds for $a = F(p)$ and $A = A(p)$, that is,

$$\begin{aligned}\bar{\Lambda}(F(p), A(p)) = \{u \in \mathbb{R}^m \mid F(p) + A(p)u \in C\} \neq \emptyset \text{ and } d(0, \bar{\Lambda}(F(p), A(p)))\\ \leq \beta_0 d(F(p), C).\end{aligned} \tag{4.5}$$

Take $p \in \mathbf{B}(p_0, r)$. Note that $\Lambda(p) = \{v \in T_pM \mid F(p) + \mathrm{D}F(p)v \in C\}$. Therefore, by (4.3), the following implication holds:

$$v = \sum_{i=1}^{m} u_i e_i^p \in \Lambda(p) \Longleftrightarrow u = (u_1, \ldots, u_m)^T \in \bar{\Lambda}(F(p), A(p)). \tag{4.6}$$

Hence, $\Lambda(p) \neq \emptyset$ thanks to (4.5). Since $\{e_1^p, \ldots, e_m^p\}$ is an orthonormal basis of T_pM, we have

$$\left\| \sum_{i=1}^{m} u_i e_i^p \right\|_p = \sqrt{\sum_{i=1}^{m} u_i^2}.$$

This together with (4.27) implies that $d(0, \bar{\Lambda}(F(p), A(p))) = d(0, \Lambda(p))$ and so

$$d(0, \Lambda(p)) \leq \beta_0\, d(F(p), C).$$

Since $p \in \mathbf{B}(p_0, r)$ is arbitrary, we get

$$\Lambda(p) \neq \emptyset \quad \text{and} \quad d(0, \Lambda(p)) \leq \beta_0\, d(F(p), C) \quad \text{for all } p \in \mathbf{B}(p_0, r). \tag{4.7}$$

This completes the proof of the proposition. □

In the case when $p_0 \in M$ is a regular point of the inclusion (3.1), by Proposition 4.1, there exist $r > 0$ and $\beta_0 > 0$ such that p_0 is a quasi-regular point with the quasi-regular radius $\mathbf{r}_{p_0} \geq r$ and the quasi-regular bound function $\beta_{p_0}(\cdot) \leq \beta_0$. Thus, the following two corollaries follows from Corollaries 3.1 and 3.2 (with β replaced by 0), respectively.

Corollary 4.1 *Let $p_0 \in M$ be a regular point of the inclusion* (3.1) *with $r > 0$ and $\beta_0 > 0$ such that* (3.4) *holds. Let $\eta \geq 1$, $0 < \Delta \leq \infty$, $\xi = \eta\beta_0 d(F(p_0), C)$,*

$$R^* = \frac{1 + L\eta\beta_0\xi - \sqrt{1 - (L\eta\beta_0\xi)^2}}{L\eta\beta_0} \quad \textit{and} \quad Q = \frac{1 - \sqrt{1 - (L\eta\beta_0\xi)^2}}{L\eta\beta_0\xi}.$$

Assume that $\mathrm{D}F$ *is Lipschitz continuous on* $\mathbf{B}(p_0, R^*)$ *with modulus L, and that*

$$\xi \leq \min\left\{\frac{1}{L\beta_0\eta},\ \Delta\right\} \quad \textit{and} \quad r \geq R^*$$

Then the same conclusions hold as in Corollary 3.1.

Corollary 4.2 *Let $p_0 \in M$ be a regular point of the inclusion* (3.1) *with $r > 0$ and $\beta_0 > 0$ such that* (3.4) *holds.*

Let $\eta \geq 1$, $0 < \Delta \leq \infty$, $\xi = \eta\beta_0 d(F(p_0), C)$ and $\alpha = \frac{\eta\beta_0(1-\gamma\xi)^2}{\eta\beta_0+(1-\eta\beta_0)(1-\gamma\xi)^2}$.

Set

$$r_\alpha^* = \frac{1+\gamma\xi - \sqrt{(1+\gamma\xi)^2 - 4(1+\alpha)\gamma\xi}}{2(1+\alpha)\gamma} \quad \textit{and}$$
$$q_\alpha = \frac{1-\gamma\xi - \sqrt{(1+\gamma\xi)^2 - 4(1+\alpha)\gamma\xi}}{1-\gamma\xi + \sqrt{(1+\gamma\xi)^2 - 4(1+\alpha)\gamma\xi}}. \tag{4.8}$$

Assume that F satisfies the γ-condition at p_0 in $\mathbf{B}(p_0, r_\alpha^*)$, and that

$$\xi \le \min\left\{\frac{1+2\alpha-2\sqrt{\alpha(1+\alpha)}}{\gamma}, \quad \Delta\right\} \quad \textit{and } r_\alpha^* \le r.$$

Then the same conclusions hold as in Corollary 3.2.

Recall that the concept of convex process was introduced by Rockafellar [42,43] for convexity problems (see also Robinson [40]).

Definition 4.2 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 $\mathcal{D}(W)$, $\mathrm{R}(W)$, W^{-1}; i.e.,

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

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

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

where, following [24,40] , for a set G in a Banach space, $\|G\|$ denotes its distance to the origin, that is,

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

The convex process W is said to be normed if $\|W\| < +\infty$.

Recall that F is a C^1 mapping. Let $p \in M$ and let C be a closed convex cone in $\mathbb{R}^n$. We define a set-valued mapping W_p from T_pM into $\mathbb{R}^n$ by

$$W_p v := \mathrm{D}F(p)v - C \quad \text{for each } v \in T_pM. \tag{4.9}$$

Then, W_p is a convex process. Obviously, for each $p \in M$, $\mathcal{D}(W_p) = T_pM$. The inverse of W_p is

$$W_p^{-1}y := \{v \in T_pM \mid \ \mathrm{D}F(p)v \in y + K\} \quad \text{for each } y \in \mathbb{R}^n. \tag{4.10}$$

W_p is said to carry T_pM onto $\mathbb{R}^n$ if $\mathcal{D}(W_p^{-1}) = \mathbb{R}^n$.

The following lemma is taken from [41] and will be useful.

Lemma 4.2 *Let C be a closed convex cone in $\mathbb{R}^n$. 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

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

Proposition 4.2 *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) *p_0 is a regular point of (3.1).*
- (ii) *Suppose further that C is a closed convex cone in $\mathbb{R}^n$ and $\mathrm{D}F$ satisfies the center L-average Lipschitz condition on $\mathbf{B}(p_0, r)$ for some $r > 0$. Let $\beta_0 = \|W_{p_0}^{-1}\|$ and let r_{β_0} be defined by*

$$\beta_0 \int_0^{r_{\beta_0}} L(u)\ \mathrm{d}u = 1. \tag{4.11}$$

Then the quasi-regular radius $\mathbf{r}_{p_0}$ and the quasi-regular bound function β_{p_0} satisfy $\mathbf{r}_{p_0} \ge \min\{r, r_{\beta_0}\}$ and

$$\beta_{p_0}(t) \le \frac{\beta_0}{1 - \beta_0 \int_0^t L(u)\ \mathrm{d}u} \quad \textit{for each } t \textit{ with } 0 \le t < \min\{r, r_{\beta_0}\}. \tag{4.12}$$

Proof The proof of (i) is similar to that of [24] and so is omitted here.

Below, we give the proof of (ii), which is a modification to suit the Riemannian manifold setting of the one for [24, Proposition 3.7(ii)]. Now let $r > 0$ and suppose that $\mathrm{D}F$ satisfies the center L-average Lipschitz condition on $\mathbf{B}(p_0, r)$. Let $p \in M$ be such that $d(p_0, p) < \min\{r, r_{\beta_0}\}$, and let c be a minimizing geodesic connecting p_0, p. Then

$$\|\mathrm{D}F(p)P_{c,p,p_0} - \mathrm{D}F(p_0)\| \le \int_0^{l(c)} L(u)\ \mathrm{d}u = \int_0^{d(p_0,p)} L(u)\ \mathrm{d}u < \int_0^{r_{\beta_0}} L(u)\ \mathrm{d}u;$$

hence, by (4.11),

$$\|W_{p_0}^{-1}\|\|\mathrm{D}F(p)P_{c,p,p_0} - \mathrm{D}F(p_0)\| < \|W_{p_0}^{-1}\| \int_0^{r_{\beta_0}} L(u)\ \mathrm{d}u = 1. \tag{4.13}$$

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

$$W_p v = \mathrm{D}F(p)v - C = [\mathrm{D}F(p_0)P_{c,p_0,p} + (\mathrm{D}F(p) - \mathrm{D}F(p_0)P_{c,p_0,p})]v - C. \tag{4.14}$$

Since $P_{c,p_0,p}$ is an isomorphism from T_pM to $T_{p_0}M$ (see [17, p.30]), $W_{p_0} \circ P_{c,p_0,p}$ is a convex process from T_pM into $\mathbb{R}^n$ and $\mathrm{D}f(p) - \mathrm{D}f(p_0)P_{c,p_0,p}$ 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 $\|\mathrm{D}f(p) - \mathrm{D}f(p_0)P_{c,p_0,p}\| = \|\mathrm{D}f(p)P_{c,p,p_0} - \mathrm{D}f(p_0)\|$ thanks to the fact that $P_{c,p_0,p}$ is an isometry (see [10, p. 56]). Thus, by (4.13) and (4.14), Lemma 4.2 is applicable to concluding that W_p carries T_pM onto $\mathbb{R}^n$ and

$$\|W_p^{-1}\| \le \frac{\|W_{p_0}^{-1}\|}{1 - \|W_{p_0}^{-1}\|\|\mathrm{D}F(p)P_{c,p,p_0} - \mathrm{D}F(p_0)\|} \le \frac{\|W_{p_0}^{-1}\|}{1 - \|W_{p_0}^{-1}\| \int\limits_0^{d(p_0,p)} L(u)\mathrm{d}u}. \tag{4.15}$$

Since W_p is surjective, we have that $\Lambda(p)$ is nonempty; in particular, for each $c \in C$,

$$W_p^{-1}(c - F(p)) \subseteq \Lambda(p) \tag{4.16}$$

To see this, let $v \in W_p^{-1}(c - F(p))$. Then, by (4.10), one has that $\mathrm{D}F(p)\,v \in c - F(p) + C \subseteq C - F(p)$ and so $F(p) + \mathrm{D}F(p)\,v \in C$, that is, $v \in \Lambda(p)$. Hence (4.16) is true. Consequently,

$$d(0, \Lambda(p)) \le \|W_p^{-1}(c - F(p))\| \le \|W_p^{-1}\|\|c - F(p)\|.$$

Since this is valid for each $c \in C$, it is seen that

$$d(0, \Lambda(p)) \le \|W_p^{-1}\| d(F(p), C).$$

Combining this with (4.15) and (4.13) gives the desired result (4.12), and the proof is complete. □

To obtain one of the main results in this section, we still need the following lemma.

Lemma 4.3 *Let $\alpha > 0, \beta_0 > 0$, and let r_α^* be given by Lemma 3.1. Let r_{β_0} be defined by (4.11). Then $r_{\beta_0} \ge r_\alpha^*$.*

Proof We consider the two cases namely: (i) $\alpha \ge \beta_0$ and (ii) $\alpha < \beta_0$. In (i), since by Lemma 3.2 r_α is decreasing with respect to α, we have that $r_\alpha^* \le r_\alpha \le r_{\beta_0}$. In (ii), since r_α^* is increasing with respect to α by Lemma 3.2, we have that $r_\alpha^* \le r_{\beta_0}^* \le r_{\beta_0}$. The proof is complete. □

Then we have the following corollaries.

Corollary 4.3 *Let $\eta \ge 1$, $0 < \Delta \le \infty$ and let C be a cone. Let $p_0 \in M$ be such that W_{p_0} carries $T_{p_0}M$ onto $\mathbb{R}^n$. Let*

$$\xi = \eta\|W_{p_0}^{-1}\|d(F(p_0), C) \tag{4.17}$$

and

$$\alpha = \frac{\eta\|W_{p_0}^{-1}\|}{1 + (\eta - 1)\|W_{p_0}^{-1}\| \int\limits_0^{\xi} L(u)\mathrm{d}u}. \tag{4.18}$$

let b_α, r_α be defined by (3.9). Let r_α^* denote the smaller zero of the function ϕ_α defined by (3.10).

Suppose that $\mathrm{D}F$ *satisfies the* L*-average Lipschitz condition on* $\mathbf{B}(p_0, r_\alpha^*)$*, and that*

$$\xi \le \min\{b_\alpha, \Delta\}. \tag{4.19}$$

Then the same conclusions hold as in Theorem 3.1.

Proof Let $\beta_0 = \|W_{p_0}^{-1}\|$ and let r_{β_0} be defined by (4.11). Then, by Proposition 4.2(ii), we know that p_0 is a quasi-regular point with the quasi-regular radius

$$\mathbf{r}_{p_0} \ge \min\{r_\alpha^*, r_{\beta_0}\} \tag{4.20}$$

and the quasi-regular bound function

$$\beta_{p_0}(t) \le \frac{\beta_0}{1 - \beta_0 \int\limits_0^t L(u)\mathrm{d}u} \quad \text{for each } t \text{ with } 0 \le t < \min\{r_\alpha^*, r_{\beta_0}\}. \tag{4.21}$$

Let $\mathbf{r} := \min\{r_\alpha^*, r_{\beta_0}\}$, and let $\alpha_0(\mathbf{r})$ be defined by (3.17). Then we have from (3.35) (with β replaced by β_0) that $\alpha \ge \alpha_0(\mathbf{r})$. Furthermore, it follows from Lemma 4.3 that $\mathbf{r} = r_\alpha^* \le \mathbf{r}_{p_0}$. On the other hand, we note that $\beta_{p_0}(0) \le \beta_0$ by (4.21) and so the ξ defined by (3.16) is majorized by that defined by (4.17); thus (4.19) entails that (3.18) holds and Theorem 3.1 is applicable. The proof is complete. □

Taking $L(u) \equiv L$ in Corollary 4.3, we get Corollary 4.4 below.

Corollary 4.4 *Let* $\eta \ge 1$, $0 < \Delta \le \infty$*, and let* C *be a cone. Let* $p_0 \in M$ *be such that* W_{p_0} *carries* $T_{p_0}M$ *onto* $\mathbb{R}^n$*. Let* $\xi = \eta\|W_{p_0}^{-1}\|d(F(p_0), C)$*. Write*

$$R^* = \frac{1 + (\eta - 1)L\|W_{p_0}^{-1}\|\xi - \sqrt{1 - 2L\|W_{p_0}^{-1}\|\xi - (\eta^2 - 1)(L\|W_{p_0}^{-1}\|\xi)^2}}{L\|W_{p_0}^{-1}\|\eta} \tag{4.22}$$

and

$$Q = \frac{1 - L\|W_{p_0}^{-1}\|\xi - \sqrt{1 - 2L\|W_{p_0}^{-1}\|\xi - (\eta^2 - 1)(L\|W_{p_0}^{-1}\|\xi)^2}}{L\|W_{p_0}^{-1}\|\eta\xi}. \tag{4.23}$$

Suppose that $\mathrm{D}F$ *is Lipschitz continuous on* $\mathbf{B}(x_0, R^*)$ *with modulus L, and that*

$$\xi \le \min\left\{\frac{1}{L\|W_{p_0}^{-1}\|(\eta + 1)}, \Delta\right\}.$$

Then the same conclusions hold as in Corollary 3.1.

Corollary 4.5 follows from Proposition 3.1 and Corollary 4.3.

Corollary 4.5 *Let* $\eta \ge 1$, $0 < \Delta \le \infty$*, and let* C *be a cone. Let* $p_0 \in M$ *be such that* W_{p_0} *carries* $T_{p_0}M$ *onto* $\mathbb{R}^n$*. Let* $\xi = \eta\|W_{p_0}^{-1}\|d(F(p_0), C)$ *and*

$$\alpha = \frac{\eta\|W_{p_0}^{-1}\|(1 - \gamma\xi)^2}{(\eta - 1)\|W_{p_0}^{-1}\| + (1 - (\eta - 1)\|W_{p_0}^{-1}\|)(1 - \gamma\xi)^2}.$$

Set, as in (4.8)*,*

$$r_\alpha^* = \frac{1+\gamma\xi-\sqrt{(1+\gamma\xi)^2-4(1+\alpha)\gamma\xi}}{2(1+\alpha)\gamma} \quad \text{and}$$
$$q_\alpha = \frac{1-\gamma\xi-\sqrt{(1+\gamma\xi)^2-4(1+\alpha)\gamma\xi}}{1-\gamma\xi+\sqrt{(1+\gamma\xi)^2-4(1+\alpha)\gamma\xi}}. \tag{4.24}$$

Suppose that F *satisfies the* γ*-condition at* p_0 *in* $\mathbf{B}(p_0, r_\alpha^*)$*, and that*

$$\xi \le \min\left\{\frac{1+2\alpha-2\sqrt{\alpha(1+\alpha)}}{\gamma},\ \Delta\right\}. \tag{4.25}$$

Then the same conclusions hold as in Corollary 3.2*.*

Recall that in [48], Robinson's generalized Newton's method is extended to solve the inclusion problem on Riemannian manifolds, i.e., finding a point $p^* \in M$ such that

$$F(p^*) \in C, \tag{4.26}$$

where C is a nonempty closed convex cone in $\mathbb{R}^n$ and F is a C^1 mapping from a manifold M onto $\mathbb{R}^n$. The extended Newton's method for the inclusion problem (4.26) is defined as follows.

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

$$\|v_k\| := \min\{\|v\| \mid v \in \theta(p_k)\} \text{ and set } p_{k+1} = \exp_{p_k} v_k, \tag{4.27}$$

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

$$\theta(p) := \{v \in T_pM \mid F(p) + \mathcal{D}F(p)v \in C\}.$$

Hence, by Corollary 4.3 (with η and Δ replaced by 1 and ∞, respectively), we obtain Corollary 4.6 below, which has been presented in [48, Theorem 3.1].

Let r_0 and $b > 0$ be such that

$$\int_0^{r_0} L(u)\mathrm{d}u = 1 \quad \text{and} \quad b = \int_0^{r_0} L(u)u\mathrm{d}u.$$

Let $\xi > 0$ and define the majorizing function ϕ by

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

Thus

$$\phi'(t) = -1 + \int_0^t L(u)\mathrm{d}u.$$

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

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

Recall from [51] (see also Lemma 3.1), if $\xi < b$, ϕ has two zeros, denoted respectively by r_1 and r_2, such that

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

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

Corollary 4.6 *Let C be a cone, and let $p_0 \in M$ be such that W_{p_0} carries $T_{p_0}M$ onto $\mathbb{R}^n$. Suppose that*

$$\xi = \|W_{p_0}^{-1}\| d(F(p_0), C) \le b$$

and that

$$\|W_{p_0}^{-1}\| \cdot \|\mathrm{D}F(q)P_{c,q,p} - \mathrm{D}F(p)\| \le \int_{d(p_0,p)}^{d(p_0,p)+l(c)} L(u)\mathrm{d}u \tag{4.28}$$

holds for any $p, q \in \mathbf{B}(p_0, r_1)$ and any geodesic c connecting p, q with $d(p_0, p)+l(c) < r_1$. Then, Algorithm $\mathbf{N}(p_0)$ is well-defined and any sequence $\{p_n\}$ so generated converges to some p^* such that $F(p^*) \in C$, and the following assertion holds for each $n = 0, 1, \ldots$:

$$d(p_n, p^*) \le r_1 - t_n.$$

Proof Define

$$\overline{L(u)} = \frac{1}{\|W_{p_0}^{-1}\|} L(u) \quad \text{for each } u \ge 0.$$

Let $\alpha = \|W_{p_0}^{-1}\|$. Replacing $L(u)$ and α in (3.9) and Lemma 3.1 by $\overline{L(u)}$ and $\|W_{p_0}^{-1}\|$, respectively, we get

$$b_\alpha = b, \quad t_{\alpha,n} = t_n \quad \text{and} \quad r_\alpha^* = r_1.$$

Furthermore, by (4.28), DF satisfies the $\overline{L}$-average Lipschitz condition on $\mathbf{B}(p_0, r_\alpha^*)$. Thus, Corollary 4.3 (with η and Δ replaced by 1 and ∞, respectively) is applicable to concluding that any sequence $\{p_n\}$ generated by Algorithm $\mathbf{R}(1, \infty, p_0)$ converges to some p^* such that $F(p^*) \in C$, and the following assertions hold for each $n = 0, 1, \ldots$:

$$d(p_n, p^*) \le r_1 - t_{\alpha,n} = r_1 - t_n$$

and

$$F(p_n) + \mathrm{D}F(p_n)v_n \in C. \tag{4.29}$$

To complete the proof, it's sufficient to prove that Algorithm $\mathbf{N}(p_0)$ is well-defined and any sequence so generated is also a sequence generated by Algorithm $\mathbf{R}(1, +\infty, p_0)$. Since W_{p_0} carries $T_{p_0}M$ onto $\mathbb{R}^n$, we have $\theta(p_0) \ne \emptyset$. Hence, there exists $v_0 \in \theta(p_0)$ such that $\|v_0\| = d(0, \theta(p_0))$ and $p_1 = \exp_{p_0} v_0$. Noting that $\theta(p_0) = \Lambda_\infty(p_0)$, by Remark 3.1, p_1 can be regarded as a point obtained by Algorithm $\mathbf{R}(1, +\infty, x_0)$ at its first iteration. Therefore, (4.29) holds for $n = 1$. Then, it follows from Remark 3.1 that there exists $v_1 \in \theta(p_1) = \Lambda_\infty(p_1)$ such that $\|v_1\| = d(0, \theta(p_1))$ and $p_2 = \exp_{p_1} v_1$. Hence, p_2 is also a point obtained by Algorithm $\mathbf{R}(1, +\infty, x_0)$ at its second iteration. Inductively, we see that, for each n, $\emptyset \ne \theta(p_n) = \Lambda_\infty(p_n)$, and this means that Algorithm $\mathbf{N}(p_0)$ is well-defined and any sequence $\{p_n\}$ so generated is also a sequence generated by Algorithm $\mathbf{R}(1, \infty, p_0)$. The proof is complete. □

References

1. Adler, R., Dedieu, J.P., Margulies, J., Martens, M., Shub, M.: Newton's method on Riemannian manifolds and a geometric model for human spine. IMA J. Numer. Anal. **22**, 359–390 (2002)
2. Azagra, D., Ferrera, J., López-Mesas, F.: Nonsmooth analysis and Hamilton–Jacobi equations on Riemannian manifolds. J. Funct. Anal. **220**, 304–361 (2005)
3. Balogh, J., Csendes, T., Rapcsák, T.: Some global optimization problems on Stiefel manifolds. J. Global Optim. **30**, 91C101 (2004)
4. Borwein, J.M.: Stability and regular points of inequality systems. J. Optim. Theory Appl. **48**, 9–52 (1986)
5. Burke, J.V.: An exact penalization viewpoint of constrained optimization. SIAM J. Control Optim. **29**, 968–998 (1991)
6. Burke, J.V.: Descent methods for composite nondifferentiable optimization problems. Math. Program. **33**(3), 260–279 (1985)
7. Burke, J.V., Ferris, M.C.: A Gauss–Newton method for convex composite optimization. Math. Program. **71**, 179–194 (1995)
8. Burke, J.V., Ferris, M.C.: A Gauss–Newton method for convex composite optimization. Math. Program. (Ser. A) **71**, 179–194 (1995)
9. Dedieu, J.P., Priouret, P., Malajovich, G.: Newton's method on Riemannian manifolds: covariant alpha theory. IMA J. Numer. Anal. **23**, 395–419 (2003)
10. DoCarmo, M.P.: Riemannian Geometry. Birkhauser, Boston (1992)
11. Ferreira, O.P., Oliveira, P.R.: Proximal point algorithm on Riemannian manifolds. Optimization **51**, 257–270 (2002)
12. Ferreira, O.P., Pérez, L.R., Lucambio, Németh, S.Z.: Singularities of monotone vector fields and an extragradient-type algorithm. J. Global Optim. **31**(1), 133–151 (2005)
13. Ferreira, O.P., Svaiter, B.F.: Kantorovich's theorem on Newton's method in Riemannian manifolds. J. Complex. **18**, 304–329 (2002)
14. Flecher, R.: Second order correction for nondifferentiable optimization. In: Watson, G.A. (ed.) Numerical Analysis, Lecture Notes in Mathematics, vol. 912, pp. 85–114. Spring, Berlin (1982)
15. Flecher, R.: Practical Methods of Optimization, 2nd edn. Wiley, New York (1987)
16. Gragg, W.B., Tapai, R.A.: Optimal error bounds for the Newton–Kantorovich theorems. SIAM J. Numer. Anal. **11**, 10–13 (1974)
17. Helgason, S.: Differential Geometry, Lie Groups, and Symmetric Spaces. Academic Press, Inc, New York (1978)
18. Helmke, U., Huper, K., Moore, J.B.: Quadratically convergent algorithms for optimal dexterous hand grasping. IEEE Trans. Robot. Automat. **18**(2), 138–146 (2002)
19. Helmke, U., Moore, J.B.: Optimization and Dynamical Systems. Springer, New York (1994)
20. Hiriart-Urruty, J., Lemarechal, C.: Convex Analysis and Minimization Algorithms II, Vol. 305 of Grundlehren der Mathematschen Wissenschaften. Springer, New York (1993)
21. Jittorntrum, K., Osborne, M.R.: Strong uniqueness and second order convergence in nonlinear discrete approximation. Numer. Math. **34**, 439–455 (1980)
22. Ledyaev, Y.S., Zhu, Q.J.: Nonsmooth analysis on smooth manifolds. Trans. Am. Math. Soc. **359**, 3687–3732 (2007)
23. Li, C., López, G., Martín-Márquez, V.: Monotone vector fields and the proximal point algorithm on Hadamard manifolds. J. Lond. Math. Soc. **79**(2), 663–683 (2009)
24. Li, C., Ng, K.F.: Majorizing functions and convergence of the Gauss–Newton method for convex composite optimization. SIAM J. Optim. **18**, 613–642 (2007)
25. Li, C., Wang, J.H.: Convergence of the Newton method and uniqueness of zeros of vector fields on Riemannian manifolds. Sci. China Ser. A **48**, 1465–1478 (2005)
26. Li, C., Wang, J.H.: Newton's method on Riemannian manifolds: Smale's point estimate theory under the γ-condition. IMA J. Numer. Anal. **26**, 228–251 (2006)
27. Li, C., Wang, J.H.: Newton's method for sections on Riemannian manifolds: generalized covariant α-theory. J. Complex. **24**, 423–451 (2008)
28. Li, C., Wang, X.H.: On convergence of the Gauss–Newton method for convex composite optimization. Math. Program. **91**, 349–356 (2002)
29. Li, S.L., Li, C., Liou, Y.C., Yao, J.C.: Existence of solutions for variational inequalities on Riemannian manifolds. Nonlinear Anal. **71**, 5695–5706 (2009)
30. Lippert, R., Edelman, A.: Nonlinear eigenvalue problems with orthogonality constraints (Section 9.4). In: Bai Z, Demmel J, Dongarra J, Ruhe A, van der Vorst H (eds) Templates for the Solution of Algebraic Eigenvalue Problems, pp. 290–314. SIAM, Philadelphia (2000)

31. Ma, Y., Kosecka, J., Sastry, S.S.: Optimization criteria and geometric algorithms for motion and structure estimation. Int. J. Comput. Vis. **44**, 219–249 (2001)
32. Madsen, K.: Minimization of nonlinear approximation function. Ph.D.Thesis, Institute of Numerical Anakysis. Technical University of Denmark, Lyngby (1985)
33. Mahony, R., Manton, J.H.: The geometry of the Newton method on non-compact Lie groups. J. Global Optim. **23**(3–4), 309–327 (2002)
34. Nishimori, Y., Akaho, S.: Learning algorithms utilizing quasi-geodesic flows on the Stiefel manifold. Neurocomputing **67**, 106–135 (2005)
35. Ostrowski, A.M.: Solutions of Equations in Euclidean and Banach Spaces. Academic Press, New York (1973)
36. Rapcsák, T.: Smooth Nonlinear Optimization in $\mathbb{R}^n$. Nonconvex Optimization and Its Applications, vol. 19. Kluwer, Dordrecht (1997)
37. Rapcsák, T.: Sectional curvatures in nonlinear optimization. J. Global Optim. **40**, 375–388 (2008)
38. Robinson, S.M.: Stability theory for systems of inequalities, Part I: Linear systems. SIAM J. Numer. Anal. **12**, 754–769 (1975)
39. Robinson, S.M.: Stability theory for systems of inequalities, Part II: Differentiable nonlinear systems. SIAM J. Numer. Anal. **13**, 479–513 (1976)
40. Robinson, S.M.: Normed convex process. Trans. Am. Math. Soc. **174**, 127–140 (1972)
41. Robinson, S.M.: Extension of Newton's method to nonlinear functions with values in a cone. Numer. Math. **19**, 341–347 (1972)
42. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton, NJ (1970)
43. Rockafellar, R.T.: Monotone Processes of Convex and Concave Type. Mem. American Mathematical Society 77. AMS, Providence, RI (1967)
44. Rockafellar, R.T.: First and second order epi-differentiability in nonlinear programming. Trans. Am. Math. Soc. **307**, 75–108 (1988)
45. Smith, S.T.: Covariance, subspace, and intrinsic Cramér-Rao bounds. IEEE Trans. Signal Process. **53**, 1610–1630 (2005)
46. Smith, S.T.: Optimization techniques on Riemannian manifolds. In: Fields Institute Communications, vol. 3, pp. 113–146. American Mathematical Society, Providence, RI (1994)
47. Udriste, C.: Convex Functions and Optimization Methods on Riemannian Manifolds. Mathematics and Its Applications, vol. 297. Kluwer, Dordrecht (1994)
48. Wang, J.H., Huang, S., Li, C.: Extended Newton's method for mappings on Riemannian manifolds with values in a cone. Taiwanese J. Math. 13:633–656 (2009)
49. Wang, J.H., Li, C.: Uniqueness of the singular points of vector fields on Riemannian manifolds under the γ-condition. J. Complex. **22**, 533–548 (2006)
50. Wang, X.H.: Convergence of an iteration process. Kexue Tongbao **20**, 558–559 (1975)
51. Wang, X.H.: Convergence of Newton's method and inverse function theorem in Banach space. Math. Comput. **225**, 169–186 (1999)
52. Wang, X.H.: Convergence on the iteration of Halley family in weak conditions. Chin. Sci. Bull. **42**(7), 552–555 (1997)
53. Wang, X.H., Han, D.F.: Criterion α and Newton's method. Chin. J. Numer. Appl. Math. **19**(2), 96–105 (1997)
54. Wang, X.H., Li, C.: Local and global behaviors for algorithms of solving equations. Chin. Sci. Bull. **46**, 441–447 (2001)
55. Womersley, R.S.: Local properties of algorithms for minimizing nonsmooth composite function. Math. Program. **32**, 69–89 (1985)
56. Xavier, J., Barroso, V.: Intrinsic variance lower bound (IVLB): An extension of the Cramer-Rao bound to Riemannian manifolds. In: Proceedings of 2005 IEEE international conference on acoustics, speech, and signal processing (ICASSP 2005), Philadelphia, PA (2005)
57. Yan, W.Y., Lam, J.: An approximate approach to $\mathcal{H}_2$ optimal model reduction. IEEE Trans. Automat. Control. **44**(7), 1341–1358 (1999)