J. Nonlinear Funct. Anal. 2016 (2016), Article ID 27

MATHRES
Journal of Nonlinear Functional Analysis
Available online at http://jnfa.mathres.org

EXTENDING THE APPLICABILITY OF GAUSS-NEWTON METHOD FOR CONVEX COMPOSITE OPTIMIZATION ON RIEMANNIAN MANIFOLDS USING RESTRICTED CONVERGENCE DOMAINS

IOANNIS K. ARGYROS[1,*], SANTHOSH GEORGE[2]

[1]Department of Mathematics Sciences, Cameron University, Lawton, OK 73505, USA

[2]Department of Mathematical and Computational Sciences,

National Institute of Technology, Karnataka, 575025, India

Abstract. In this study we present a semi-local convergence analysis of the Gauss-Newton method for solving convex composite optimization problems in Riemannian manifolds using the our idea of restricted convergence domains. Using this idea we introduce majorizing sequences for the Gauss-Newton method that are more precise than in earlier studies. Consequently, our semi-local convergence analysis for the Gauss-Newton method has the following advantages under the same computational cost: weaker sufficient convergence conditions; more precise estimates on the distances involved and an at least as precise information on the location of the solution.

Keywords. Gauss-Newton method; Riemannian manifold; Convex composite optimization; Semi-local convergence; Quasi-regularity.

1. Introduction

In this paper we are concerned with the convex composite optimization problem on a Riemannian manifold. We present a convergence analysis of Gauss–Newton method on Riemannian manifolds [11, 13, 24]. Using the restricted convergence domains, we obtained a finer convergence analysis, with the advantages $(\mathscr{A})$: tighter error estimates on the distances involved

*Corresponding author.

E-mail addresses: iargyros@cameron.edu (I.K. Argyros), sgeorge@nitk.ac.in (S. George)

Received March 24, 2016

and the information on the location of the solution is at least as precise. These advantages were obtained (under the same computational cost) using the same or weaker hypotheses as in [9, 24].

1.1. Gauss-Newton algorithms

The purpose of this paper is to study the convex composite optimization problem

$$\min_{x\in\mathbb{R}^l} f(x) := h(F(x)), \tag{1}$$

where $h:\mathbb{R}^m \longrightarrow \mathbb{R}$ is convex, $F:\mathbb{R}^l \longrightarrow \mathbb{R}^m$ is Fréchet-differentiable operator and $m,l\in\mathbb{N}^\star$ using the idea of restricted convergence domains. The important of the study of (1) can be found in (see, e.g., [3, 9, 7, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25]). We assume that the minimum h_{min} of the function h is attained. Problem (1) is related to

$$F(x)\in\mathscr{C}, \tag{2}$$

where

$$\mathscr{C} = \operatorname{argmin} h \tag{3}$$

is the set of all minimum points of h.

A semilocal convergence analysis for Gauss-Newton method (GNA) was presented using the algorithm (see, e.g., [5, 9, 14]):

Algorithm (GNA) : (ξ,Δ,x_0)

Let $\xi\in[1,\infty[$, $\Delta\in]0,\infty]$ and for each $x\in\mathbb{R}^l$, define $\mathscr{D}_\Delta(x)$ by

$$\begin{aligned}\mathscr{D}_\Delta(x) \;=\; \{&d\in\mathbb{R}^l : \| d \|\le\Delta,\ h(F(x)+DF(x)d)\le h(F(x)+DF(x)d')\\ &\text{for all } d'\in\mathbb{R}^l \text{ with } \| d' \|\le\Delta\}.\end{aligned} \tag{4}$$

Let also $x_0\in\mathbb{R}^l$ be given. Having $x_0, x_1, \cdots, x_k$ $(k\ge 0)$, determine x_{k+1} by:

If $0\in\mathscr{D}_\Delta(x_k)$, then STOP;

If $0\notin\mathscr{D}_\Delta(x_k)$, choose d_k such that $d_k\in\mathscr{D}_\Delta(x_k)$ and

$$\| d_k \|\le \xi\, d(0,\mathscr{D}_\Delta(x_k)). \tag{5}$$

Then, set $x_{k+1}=x_k+d_k$.

Here, $d(x,W)$ denotes the distance from x to W in the finite dimensional Banach space containing W. Note that the set $\mathscr{D}_\Delta(x)$ ($x \in \mathbb{R}^l$) is nonempty and is the solution of the following convex optimization problem

$$\min_{d\in\mathbb{R}^l,\,\|d\|\leq\Delta} h(F(x)+DF(x)\,d), \tag{6}$$

which can be solved by well known methods such as the subgradient or cutting plane or bundle methods (see, e.g., [19]). Many popular iterative methods such as trust region method, conjugate gradient method, steepest descent method, Newton-like methods, has been extended from a Banach space to a Riemannian manifold setting. Recently, Wang, Yao and Li [24] extended the Gauss-Newton method (GNA) to Riemmannian manifold to solve the convex composite optimization on Riemannian manifold $\mathscr{M}$ which is formulated as follows:

$$\min_{p\in\mathscr{M}} f(p) := h(F(p)), \tag{7}$$

where h is same as defined above and F is a differentiable mapping from $\mathscr{M}$ to $\mathbb{R}^l$. As mentioned before, the study of (7) naturally relates to the convex inclusion problem

$$F(p) \in C, \tag{8}$$

where $C = argminh$, the set of all minimum points of h. The extended Gauss-Newton method for convex composite optimization problem on Riemannian manifold (7) (GNAR) is defined as follows.

Algorithm (GNAR) : (ξ,Δ,x_0)

Let $\xi \geq 1$, $0 < \Delta \geq \infty$. Let $p_0 \in \mathscr{M}$ 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 $v_k \in \Lambda_\Delta(p_k)$ and

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

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

$$\Lambda_\Delta(p) := \big\{ v \in T_p\mathscr{M} \|v\| \leq \Delta, h(F(p)+DF(p)v) \leq h(F(p)) + DF(p)v')$$

$$\forall v' \in T_p\mathscr{M} \text{ with } \|v'\| \le \Delta\big\}.$$

1.2. Gauss-Newton algorithms

We used the same notations and notions about smooth manifolds used in the papers [3, 9, 7, 8, 24]. “Let $\mathscr{M}$ be a complete connected m-dimensional Riemannian manifold with the Levi-Civita connection ∇ on $\mathscr{M}$. Let $p \in \mathscr{M}$, and let $T_p\mathscr{M}$ denote the tangent space at p to $\mathscr{M}$. Let $< \cdot, \cdot >$ be the scalar product on $T_p\mathscr{M}$ with the associated norm $\|\cdot\|_p$, where the subscript p is sometimes omitted. For any two distinct elements $p, q \in \mathscr{M}$, let $c : [0,1] \to \mathscr{M}$ be a piecewise smooth curve connecting p and q. Then the arc-length of c is defined by $l(c) := \int_0^1 \|c'(t)\|dt$, 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 $c : [0,1] \to \mathscr{M}$ connecting p and q. Thus, by the HopfRinow Theorem (see [19]), (M,d) is a complete metric space and the exponential map at p, $\exp_p : T_p\mathscr{M} \to \mathscr{M}$ is well-defined on $T_p\mathscr{M}$.

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

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

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

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 $T_{c(a)}\mathscr{M}$ to $T_{c(b)}\mathscr{M}$. Note that, for any $a, b, b_1, b_2 \in \mathbb{R}$,

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

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 $C^1(T\mathscr{M})$ denote the set of all the C^1-vector fields on $\mathscr{M}$ and $C^i(\mathscr{M})$ the set of all C^i-functions from $\mathscr{M}$ to $\mathbb{R}$ ($i = 0, 1$, where C^0-mappings mean continuous mappings), respectively. Let $F : \mathscr{M} \to \mathbb{R}^l$ 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 $\mathscr{M}$, and let $X \in C^1(T\mathscr{M})$. Then, 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 $DF : (C^1(T\mathscr{M})) \to (C^0(\mathscr{M}))^n$ defined by

$$DF(X) = \nabla_X F \text{ for each } X \in C^1(T\mathscr{M}). \tag{9}$$

We use $DF(p)$ to denote the derivative of F at p. Let $v \in T_p\mathscr{M}$. Taking $X \in C^1(T\mathscr{M})$ such that $X(p) = v$, and any nontrivial smooth curve $c : (-\varepsilon, \varepsilon) \to \mathscr{M}$ with $c(0) = p$ and $c'(0) = v$, one has that

$$DF(p)v := DF(X)(p) = \nabla_X F(p) = \left(\frac{d}{dt}(F \circ c)(t)\right)_{t=0}, \tag{10}$$

which only depends on the tangent vector v.

Let $U(x,r)$ denote the open ball in $\mathbb{R}^l$ (or $\mathbb{R}^m$) centered at x and of radius $r > 0$. By $\overline{U}(x,r)$ we denote its closure. Let W be a closed convex subset of $\mathbb{R}^l$ (or $\mathbb{R}^m$). The negative polar of W denoted by $W^{\ominus}$ is defined as

$$W^{\ominus} = \{z :< z, w >\leq 0 \quad \text{for each} \quad w \in W\}''. \tag{11}$$

1.3. Quasi-regularity and convergence criterion

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

$$F(p) \in C. \tag{12}$$

Let $p \in \mathscr{M}$ and

$$\Lambda(p) := \{v \in T_p\mathscr{M} | F(p) + DF(p)v \in C\}. \tag{13}$$

Remark 1.1. In the case when C is the set of all minimum points of h and if there exists $v_0 \in T_p\mathscr{M}$ with $v_0 \leq \Delta$ such that $v_0 \in \Lambda(p)$, then $v_0 \in \Lambda_\Delta(p)$ and for each $v \in T_p\mathscr{M}$ with $\|v\| = \Delta$ one has

$$v \in \Lambda_\Delta(p) \Leftrightarrow v \in \Lambda(p) \Leftrightarrow v \in \Lambda_\infty(p). \tag{14}$$

Definition 1.2. A point $p_0 \in \mathscr{M}$ is called a quasi-regular point of the inclusion (12) if there exist $r > 0$ and an increasing positive-valued function β on $[0,r)$ such that

$$\Lambda(p) \neq \emptyset \text{ and } d(0,\Lambda(p)) \leq \beta(d(p_0,p))d(F(p),C) \text{ for all } p \in B(p_0,r). \tag{15}$$

Following [24], let r_{p_0} denote the supremum of r such that (15) holds for some increasing positive-valued function β on $[0,r)$. Let $r \in [0,r_{p_0}]$ and let $B_r(p_0)$ denote the set of all increasing positive-valued function β on $[0,r)$ such that (15) holds. Define

$$\beta_{p_0}(t) = \inf\{\beta(t) : \beta \in B_{r_{p_0}}(p_0)\} \text{ for each } t \in [0,r_{p_0}). \tag{16}$$

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

$$\beta_{p_0}(t) = inf\{\beta(t) : \beta \in B_r(p_0)\} \text{ for each } t \in [0,r). \tag{17}$$

We call 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 Banach spaces was first introduced in [24] by Wang for the study of Smales point estimate theory, where the terminology "the center Lipschitz condition in the inscribed sphere with L average" was used (see [23]). Recently, this notion has been extended and applied to sections on Riemannian manifolds in [1, 3, 4, 7, 8, 9, 13, 14, 24]. Note that a mapping$F : \mathscr{M} \to \mathbb{R}^m$ is a special example of sections.

The rest of the paper is organized as follows: Section 2 contains the L-average condition and majorizing sequences. In Section 5, a numerical example is given to illustrate our theoretical results.

2. L-average condition and majorizing sequences

We used the following notion of generalized Lipschitz condition due to Wang in [23] for our convergence analysis. From now on $L : [0,\infty[\longrightarrow]0,\infty[$ (or K) denotes a nondecreasing and absolutely continuous function. Moreover, ξ and α denote given positive numbers.

Definition 2.1. Let $r > 0$ and $p_0 \in \mathscr{M}$. Then, G is said to satisfy: Let $G : \mathbb{R}^l \longrightarrow \mathscr{Y}$. Then, G is said to satisfy:

(a) The center L_0-average condition on $U(p_0, r)$, if for any point $p \in U(p_0, r)$ and any geodesic c connecting p_0, p with $l(c) < r$, we have

$$\| DF_{P_{c,p,p_0}}(p) - DF(p_0) \| \leq \int_0^{l(c)} L_0(u)\, du \tag{18}$$

(b) Let $r_0 = \sup\{t \in [0, r) : L_0(t)t < 1\}$. The K-average Lipschitz condition on $U(x_0, r_0)$, if for any two points $p, q \in U(p_0, r_0)$ and any geodesic c connecting p, q with $d(p_0, p) + l(c) < r_0$, we have

$$\| DF_{P_{c,p,q}}(q) - DF(p) \| \leq \int_{d(p_0,p)}^{d(p_0,p)+l(c)} K(u)\, du \tag{19}$$

(c) The $L-$ average Lipschitz condition on $U(x_0, r)$, if for any two points $p, q \in U(p_0, r)$ and any geodesic c connectiong p, q with $d(p_0, p) + l(c) < r_0$ we have

$$\|DF_{p_0,p,q}(q) - DF(p)\| \leq \int_{d(p_0,p)}^{d(p_0,p)+l(c)} L(u) du. \tag{20}$$

Notice that in view of (18)-(20)

$$L_0(u) \leq L(u) \text{ for each } u \in [0, r) \tag{21}$$

and

$$K(u) \leq L(u) \text{ for each } u \in [0, r). \tag{22}$$

Definition 2.2. Define majorizing function φ_α on $[0, +\infty)$ by

$$\varphi_\alpha(t) = \xi - t + \alpha \int_0^t K(u)\,(t - u)\, du \quad for\ each \quad t \geq 0 \tag{23}$$

and majorizing sequence $\{s_{\alpha,n}\}$ by

$$s_{\alpha,0} = 0, \quad s_{\alpha,n+1} = s_{\alpha,n} - \frac{\varphi_\alpha(s_{\alpha,n})}{\varphi'_\alpha(s_{\alpha,n})} \quad for\ each \quad n = 0, 1, \cdots. \tag{24}$$

From now on we show how our convergence analysis for (GNAR) is finer than the one in [9, 24]. Define supplementary majorizing functions $\varphi_{\alpha,0}, \varphi_\alpha$ on $[0,+\infty)$ by

$$\varphi_{\alpha,0}(t) = \xi - t + \alpha \int_0^t L_0(u)\,(t-u)\,du \quad \text{for each} \quad t \geq 0,$$

$$\varphi_\alpha(t) = \xi - t + \alpha \int_0^t L(u)\,(t-u)\,du \quad \text{for each} \quad t \geq 0$$

and corresponding majorizing sequences $\{r_{\alpha,n}\}, \{t_{\alpha,n}\}$ by

$$r_{\alpha,0} = 0, \quad r_{\alpha,1} = \xi, \quad r_{\alpha,n+1} = r_{\alpha,n} - \frac{\psi_\alpha(s_{\alpha,n})}{\psi'_{\alpha,0}(s_{\alpha,n})} \quad \text{for each} \quad n = 1,2,\cdots, \tag{25}$$

$$t_{\alpha,0} = 0, \quad t_{\alpha,1} = \xi, \quad t_{\alpha,n+1} = t_{\alpha,n} - \frac{\psi_\alpha(t_{\alpha,n})}{\varphi'_{\alpha,0}(t_{\alpha,n})} \quad \text{for each} \quad n = 1,2,\cdots. \tag{26}$$

The results concerning $\{t_{\alpha,n}\}$ are already in the literature (see, e.g., [9]), whereas the corresponding ones for sequence $\{s_{\alpha,n}\}$ can be derivated in an analogous way by simple using $\varphi'_{\alpha,0}$ instead of φ'_α. First, we need some auxiliary results for the properties of functions φ_α, $\varphi_{\alpha,0}$ and the relationship between sequences $\{s_{\alpha,n}\}$ and $\{t_{\alpha,n}\}$. The proofs of the next four lemmas involving the φ_α function can be found in [9, 13, 23, 24], whereas the proofs for function $\varphi_{\alpha,0}$ are analogously obtained by simply replacing L by L_0.

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

$$\alpha \int_0^{r_\alpha} L(u)\,du = 1, \quad b_\alpha = \alpha \int_0^{r_\alpha} L(u)\,u\,du \tag{27}$$

and

$$\alpha \int_0^{r_{\alpha,0}} K(u)\,du = 1, \quad b_{\alpha,0} = \alpha \int_0^{r_{\alpha,0}} K(u)\,u\,du. \tag{28}$$

Clearly, we have that

$$b_\alpha < r_\alpha \tag{29}$$

and

$$b_{\alpha,0} < r_{\alpha,0}. \tag{30}$$

In view of (21), (22), (27) and (28), we get that

$$r_\alpha \leq r_{\alpha,0} \tag{31}$$

and

$$b_\alpha \leq b_{\alpha,0}. \tag{32}$$

Lemma 2.3. *Suppose that $0 < \xi \leq b_\alpha$. Then, $b_\alpha < r_\alpha$ and the following assertions hold:*

(i) *φ_α is strictly decreasing on $[0, r_\alpha]$ and strictly increasing on $[r_\alpha, \infty)$ with $\varphi_\alpha(\xi) > 0$, $\varphi_\alpha(r_\alpha) = \xi - b_\alpha \leq 0$, $\varphi_\alpha(+\infty) \geq \xi > 0$.*

(ii) *$\varphi_{\alpha,0}$ is strictly decreasing on $[0, r_{\alpha,0}]$ and strictly increasing on $[r_{\alpha,0}, \infty)$ with $\varphi_{\alpha,0}(\xi) > 0$, $\varphi_{\alpha,0}(r_{\alpha,0}) = \xi - b_{\alpha,0} \leq 0$, $\varphi_{\alpha,0}(+\infty) \geq \xi > 0$. Moreover, if $\xi < b_\alpha$, then ψ_α has two zeros, denoted by $t^\star_\alpha$ and $t^{\star\star}_\alpha$, such that*

$$\xi < t^\star_\alpha < \frac{r_\alpha}{b_\alpha}\xi < r_\alpha < t^{\star\star}_\alpha \tag{33}$$

and if $\xi = b_\alpha$, ψ_α has an unique zero $t^\star_\alpha = r_\alpha$ in (ξ, ∞);

$\varphi_{\alpha,0}$ has two zeros, denoted by $t^\star_{\alpha,0}$ and $t^{\star\star}_{\alpha,0}$, such that

$$\xi < t^\star_{\alpha,0} < \frac{r_{\alpha,0}}{b_{\alpha,0}}\xi < r_{\alpha,0} < t^{\star\star}_{\alpha,0},$$

$$t^\star_{\alpha,0} \leq t^\star_\alpha, \tag{34}$$

$$t^{\star\star}_{\alpha,0} \leq t^{\star\star}_\alpha \tag{35}$$

and if $\xi = b_{\alpha,0}$, $\varphi_{\alpha,0}$ has an unique zero $t^\star_{\alpha,0} = r_{\alpha,0}$ in (ξ, ∞).

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

(iv) *$\{s_{\alpha,n}\}$ is strictly monotonically increasing and converges to its unique least upper bound $t^\star_\alpha \leq t^\star_{\alpha,0}$.*

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

Lemma 2.4. *Let r_α, $r_{\alpha,0}$, b_α, $b_{\alpha,0}$, φ_α, $\psi_{\alpha,0}$ be as defined above. Let $\overline{\alpha} > \alpha$. Then, the following assertions hold:*

(i) *Functions $\alpha \longrightarrow r_\alpha$, $\alpha \longrightarrow r_{\alpha,0}$, $\alpha \longrightarrow b_\alpha$, $\alpha \longrightarrow b_{\alpha,0}$ are strictly decreasing on $[0, \infty)$.*

(ii) $\varphi_\alpha < \varphi_{\overline{\alpha}}$ *and* $\varphi_{\alpha,0} < \varphi_{\overline{\alpha},0}$ *on* $[0,\infty)$.

(iii) *Function* $\alpha \longrightarrow t^\star_\alpha$ *is strictly increasing on* $I(\xi)$, *where* $I(\xi) = \{\alpha > 0 : \xi \leq b_\alpha\}$.

(iv) *Function* $\alpha \longrightarrow t^\star_{\alpha,0}$ *is strictly increasing on* $I(\xi)$.

Lemma 2.5. *Let* $0 \leq \lambda < \infty$. *Define functions*

$$\chi(t) = \frac{1}{t^2}\int_0^t L(\lambda+u)\,(t-u)\,du \quad for\ all \quad t \geq 0 \tag{36}$$

and

$$\chi_0(t) = \frac{1}{t^2}\int_0^t L_0(\lambda+u)\,(t-u)\,du \quad for\ all \quad t \geq 0. \tag{37}$$

Then, functions χ *and* χ_0 *are increasing on* $[0,\infty)$.

Lemma 2.6. *Define function*

$$g_\alpha(t) = \frac{\varphi_\alpha(t)}{\varphi'_\alpha(t)} \quad for\ all \quad t \in [0, r^\star_{\alpha,0}).$$

Suppose $0 < \xi \leq b_\alpha$. *Then, function* g_α *is increasing on* $[0, r^\star_{\alpha,0})$.

Remark 2.7. The majorizing function φ_α plays a crucial role in the study of the convergence of (GNAR). According to the definition of function φ_α in (2.21), we need function L in order to construct φ_α. Then, given F and DF the general formula for computing function L is given by

$$\sup \|DF_{P_{c,p,q}}(q) - DF(p)\| = \int_{d(p_0,p)}^{d(p_0,p)+l(c)} L(u)du, \tag{38}$$

where the supremum is taken for any two points $p,q \in \bar{U}(p_0,r_0)$ for any geodesic c connecting p,q with $d(p_0,p)+l(c) \leq r_0$ (see e.g. Remark 4.1 and Example 4.2 for the construction of function L, i.e. the construction of majorizing function ψ_α). Similarly, the construction of function $\varphi_{\alpha,0}$ defined in (2.6) requires finding function L_0. According to (2.18) function L_0 must satisfy

$$\sup \|DF_{P_{c,p,q}}(q) - DF(p_0)\| = \int_0^{l(c)} L_0(u)du, \tag{39}$$

where the supremum is taken for any point $p \in \bar{U}(p_0,r)$ and any geodesic c connecting p_0,p with $l(c) \leq r$.

3. Semilocal convergence analysis for (GNAR)

Assume that the set $\mathscr{C}$ satisfies (3). Let $p_0 \in \mathscr{M}$ be a quasi-regular point of (3) with the quasi-regular radius R_{p_0} and the quasi-regular bound function β_{p_0} (i.e., see (2.16) and (2.17)). Let $\delta \in [1, +\infty)$ and let

$$\xi = \delta\, \beta_{p_0}(0)\, d(F(p_0), \mathscr{C}). \tag{40}$$

For all $R \in (0, R_{p_0}]$, we define

$$\alpha_0(R) = \sup\left\{ \frac{\delta\, \beta_{p_0}(t)}{\delta\, \beta_{p_0}(t) \displaystyle\int_0^t L_0(s)\, ds + 1} : \xi \le t < R \right\}. \tag{41}$$

Theorem 3.1. *Let $\delta \in [1, +\infty)$ and $\Delta \in (0, +\infty]$. Let $p_0 \in \mathscr{M}$ be a quasi-regular point of (3) with the quasi-regular radius R_{p_0} and the quasi-regular bound function β_{p_0}. Let $\xi > 0$ and $\alpha_0(R)$ be given in (40) and (41), respectively. Let $0 < R < R_{p_0}$, $\alpha \ge \alpha_0(R)$ be a positive constant and let b_α, r_α be as defined in (27). Let $\{s_{\alpha,n}\}$ ($n \ge 0$) be given by (24). Suppose that DF satisfies the K-average Lipschitz on $U(p_0, r_0)$ and the center L_0-average Lipschitz conditions on $U(p_0, s^\star_\alpha)$. Suppose that*

$$L_0(t) \le K(t), t \in [0, r_0),\ \ \xi \le \min\{b_\alpha, \Delta\} \quad \textit{and} \quad s^\star_\alpha := \lim_{n \longrightarrow \infty} s_{\alpha,n} \le R. \tag{42}$$

Then, sequence $\{p_n\}$ generated by (GNAR) is well defined, remains in $\overline{U}(p_0, s^\star_\alpha)$ for all $n \ge 0$ and converges to some $p^\star$ such that $F(p^\star) \in \mathscr{C}$. Moreover, the following estimates hold for each $n = 1, 2, \cdots$

$$d(p_n, p^\star) \le s^\star_\alpha - s_{\alpha,n}, \tag{43}$$

and

$$F(p_n) + DF(p_n) v_n \in C. \tag{44}$$

Proof. Simply replace L by K in the proof of Theorem 3.1 in [9] (or see[11, 13, 24]). Notice also that the iterates p_k remain in $U(p_0, r_0)$ which is more accurate location than $U(x_0, r)$ used earlier, since $U(x_0, r_0) \subset U(x_0, r)$.

We proceed by mathematical induction. By (42) and Lemma 2.6, we have that, for each n,

$$\xi \le s_{\alpha,n} < s^*_\alpha \le s \le s_{x_0}. \tag{45}$$

Using the quasi-regularity assumption, we get that

$$\Lambda(p) \neq \emptyset \text{ and } d(0,\Lambda(p)) \leq \beta_{p_0}(d(p_0,p))d(F(p),C) \text{ for each } p \in B(p_0,r_0). \tag{46}$$

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

$$\delta d(0,\Lambda(p_0)) \leq \delta\beta_{p_0}(d(p_0,p_0))d(F(p_0),C) = \delta\beta_{p_0}(0)d(F(p_0),C) = \xi \leq \Delta. \tag{47}$$

Since $\delta \geq 1$, it follows that $d(0,\Lambda(p_0)) \leq \Delta$ and so there exists $v \in T_{p_0}\mathscr{M}$ with $\|\| \leq \Delta$ such that $F(p_0)+DF(p_0)v \in C$. Then, by Remark 2.7,

$$\Lambda_\Delta(p_0) = \left\{v \in T_{p_0}\mathscr{M} | \|v\| \leq \Delta, F(p_0)+DF(p_0)v \in C\right\}$$

and

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

Hence, by the definition of (GNAR) and (47), we can choose $v_0 \in \Lambda_\Delta(p_0)$ such that

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

which implies that (43) and (45) hold for $n = 0$. Assume that (43) and (44) hold for each $n \in 0,1,...,k-1$ and define the geodesic $c : [0,1] \to \mathscr{M}$ by

$$c(\tau) = \exp_{p_{k-1}} \tau v_{k-1},\ \tau \in [0,1] \tag{48}$$

It follows from the induction hypotheses that

$$d(p_0,p_k) \leq \sum_{i=1}^{k} d(p_i,p_{i-1}) \leq \sum_{i=1}^{k}(s_{\alpha,i} - s_{\alpha,i-1}) = s_{\alpha,k} \tag{49}$$

and

$$d(p_{k-1},p_0) \leq s_{\alpha,k-1} \leq s_{\alpha,k}. \tag{50}$$

It follows from (48) and (45) that $c(\tau) \in U(p_0,s^*_\alpha) \subseteq B(p_0,r_0)$ for each $\tau \in [0,1]$. Hence, (46) holds for $p = p_k$. That is

$$\Lambda(p_k) \neq \emptyset \text{ and } d(0,\Lambda(p_k)) \leq \beta_{p_0}(d(p_k,p_0))d(F(p_k),C). \tag{51}$$

We claim that

$$\delta d(0,\Lambda(p_k)) \leq s_{\alpha,k+1} - s_{\alpha,k}. \tag{52}$$

Using (43) for $n = k-1$ and the fact that DF satisfies $K-$average Lipschitz condition on $U(p_0, s^*_\alpha)$, we have by (52) in turn that

$$\begin{aligned}
\delta d(0,\Lambda(p_k)) &\leq \delta\beta_{p_0}(d(p_0,p_k))d(F(p_k),C)\\
&\leq \delta\beta_{p_0}(d(p_0,p_k))\|F(p_k)-F(p_{k-1})-DF(p_{k-1})v_{k-1}\|\\
&\leq \delta\beta_{p_0}(d(p_0,p_k))\|\textstyle\int_0^1(DF(c(\tau))P_{c,c(\tau),p_{k-1}}-DF(p_{k-1}))v_{k-1}\tau\|\\
&\leq \delta\beta_{p_0}(d(p_0,p_k))\textstyle\int_0^1\left(\int_{d(p_{k-1},p_0)}^{\tau\|v_{k-1}\|+d(p_{k-1},p_0)}K(u)du\right)\|v_{k-1}\|d\tau\\
&= \delta\beta_{p_0}(d(p_0,p_k))\left(\textstyle\int_0^{\|v_{k-1}\|}K(d(p_{k-1},p_0)+u)(\|v_{k-1}\|-u)du\right)\\
&\leq \delta\beta_{p_0}(s_{\alpha,k})\left(\textstyle\int_0^{\|v_{k-1}\|}K(s_{\alpha,k-1}+u)(\|v_{k-1}\|-u)u\right)
\end{aligned}$$

where the last inequality is valid because K and β_{x_0} are increasing and thanks to (49) and (50). Hence, it follows that

$$\begin{aligned}
\delta d(0,\Lambda(p_k)) &\leq \delta\beta_{p_0}(s_{\alpha,k})\textstyle\int_0^{s_{\alpha,k}-s_{\alpha,k-1}}K(s_{\alpha,k-1}+u)(s_{\alpha,k}-s_{\alpha,k-1}-u)du\\
&= \frac{\delta\beta_{p_0}(s_{\alpha,k})\psi_\alpha(s_{\alpha,k})}{\alpha}.
\end{aligned}\tag{53}$$

On the other hand, by (45) and (41),

$$\frac{\delta\beta_{p_0}(s_{\alpha,k})}{\alpha_0(r)} \leq \left(1-\alpha_0(r)\int_0^{s_{\alpha,k}}L_0(u)du\right)^{-1}.$$

Since $\alpha \geq \alpha_0(r)$ and by (41), it follows that

$$\frac{\delta\beta_{p_0}(s_{\alpha,k})}{\alpha} \leq \left(1-\alpha_0(s)\int_0^{s_{\alpha,k}}L_0(u)du\right)^{-1} = -\varphi'_{\alpha,0}(s_{\alpha,k})^{-1} \leq -\varphi'_\alpha(s_{\alpha,k})^{-1}.\tag{54}$$

In view of (53), (54) and (39) we deduce that (52) hold. Moreover by Lemma 2.9 and (42), we have

$$s_{\alpha,k+1}-s_{\alpha,k} = -\varphi'_{\alpha,0}(s_{\alpha,k})^{-1}\varphi_\alpha(s_{\alpha,k}) \leq -\varphi'_\alpha(s_{\alpha,0})^{-1}\varphi_\alpha(s_{\alpha,0}) = \xi \leq \Delta,$$

it follows from (52) that $d(0,\Lambda(p_k)) \leq \Delta$. Hence there exists $v \in T_{p_k}\mathscr{M}$ with $v \leq \Delta$ such that $F(p_k)+DF(p_k)v \in C$. Then, by Remark 2.7,

$$\Lambda_\Delta(p_k) = \left\{v \in T_{p_k}\mathscr{M} \,|\, \|v\| \leq \Delta,\, F(p_k)+DF(p_k)v \in C\right\}$$

and

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

Choosing $v_k \in \Lambda_\Delta(p_k)$ according to (GNAR), we deduce that (44) holds for $n = k$. Finally, we have that

$$\|v_k\| \le \delta d(0,\Lambda\Delta(p_k)) = \delta d(0,\Lambda(p_k)).$$

The last estimate with together with (52) complete the induction.

Remark 3.2. Let $0 < r_0 \le r_{p_0}, \beta_0 > 0$ and $0 \le \beta \le \beta_0$. Let us consider one important class of quasi-regular bound function β_{p_0} satisfying

$$\beta_{p_0}(t) \le \frac{\beta_0}{1-\beta\int_0^t L_0(u)du} \text{ for each } t \in [0,r_0). \tag{55}$$

Then, we have

$$\alpha_0(r) \le \frac{\delta\beta_0}{1+(\delta\beta_0-\beta)\int_0^\delta L_0(u)du}. \tag{56}$$

In fact by (55), for each $t \in [\xi, r)$, we have

$$\delta\int_0^t L_0(u)du + \frac{1}{\beta_{x_0}(t)} \ge \frac{1}{\beta_0}\left(\delta - \frac{\beta}{\beta_0}\right)\int_0^t L_0(u)du \ge \frac{1}{\beta_0}\left(\delta - \frac{\beta}{\beta_0}\right)\int_0^\xi L_0(u)du,$$

that is,

$$\frac{\beta_{x_0}(t)}{1+\delta\beta_{x_0}(t)\int_0^t L_0(u)du} \le \frac{\beta_0}{1+(\delta\beta_0-\beta)\int_0^\xi L_0(u)du}.$$

Hence, (56) follows by the definition of $\alpha_0(r)$ in (41).

Notice that the results in [9] use L instead of L_0. Hence, if $L_0 < K$, then our results improve the older ones.

4. Special cases and applications

Remark 4.1.

(a) If $L_0(u) = K(u) = L(u)$ for each $u \in (0,r_0]$ and $r_0 = r$ then Theorem 3.1 reduces to Theorem 3.1 in [9] (which in turn improve the corresponding result in [24]). However, if $L_0(u) \le K(u) < L(u)$ for each $u \in [0,r_0]$, we have improvements. Indeed, we have that, if function ψ_α has a solution t_α^*, then, since $\varphi_\alpha(t_\alpha^*) \leq \psi_\alpha(t_\alpha^*) = 0$ and $\varphi_\alpha(0) = \psi_\alpha(0) = \xi > 0$, we get that function ψ_α has a solution r_α^* such that

$$r_\alpha^* \leq t_\alpha^* \tag{57}$$

but not necessarily vice versa. It also follows from (57) that the new information about the location of the p^* is at least as precise as the one given in [9, 24].

Let us specialize conditions (18)-(20) even further in the case when L_0, K and L are constant functions. Then, we have:

$$\psi_\alpha(t) = \frac{L}{2}t^2 - t + \xi \tag{58}$$

and

$$\varphi_\alpha(t) = \frac{K}{2}t^2 - t + \xi, \tag{59}$$

respectively. In this case the convergence criteria become, respectively

$$h = L\eta \leq \frac{1}{2} \tag{60}$$

and

$$h_1 = K\eta \leq \frac{1}{2}. \tag{61}$$

Notice that

$$h \leq \frac{1}{2} \Longrightarrow h_1 \leq \frac{1}{2} \tag{62}$$

but not necessaruly vice versa unless, if $K = L$. Criterion (60) is the famous for its simplicity and clarity Kantorovich hypothesis for the semilocal convergence of Newton's method to a solution x^* of nonlinear equation $F(x) = 0$ [3, 8]. In the case of Wang's conditions [24] we have

$$\psi(t) == \frac{\gamma t^2}{1-\gamma t} - t + \eta,$$

$$\varphi(t) = \frac{\beta t^2}{1-\beta t} - t + \eta,$$

$$L(u) = \frac{2\gamma}{(1-\gamma u)^3}, \ \gamma > 0, 0 \leq t < \frac{1}{\gamma}$$

and

$$K(u) = \frac{2\beta}{(1-\beta u)^3},\ \beta > 0, 0 \leq t < \frac{1}{\beta}$$

with convergence criteria, given respectively by

$$H = \gamma\eta \leq 3 - 2\sqrt{2} \tag{63}$$

$$H_1 = \beta\eta \leq 3 - 2\sqrt{2}. \tag{64}$$

Then, again we have that

$$H \leq 3 - 2\sqrt{2} \Longrightarrow H_1 \leq 3 - 2\sqrt{2} \tag{65}$$

but not necessarily vice versa, unless, if $\beta = \gamma$. Concerning the error bounds and the limit of majorizing sequence, suppose that

$$-\frac{\varphi_\alpha(s)}{\varphi'_\alpha(s)} \leq -\frac{\psi_\alpha(t)}{\psi'_\alpha(t)} \tag{66}$$

for each $s,t \in [0,\rho_0]$ with $s \leq t$. Then, we have that

$$s_{\alpha,n} \leq t_{\alpha,n},$$

$$s_{\alpha,n+1} - s_{\alpha,n} \leq t_{\alpha,n+1} - t_{\alpha,n},$$

and

$$s^*_\alpha \leq t^*_\alpha.$$

The first two preceding inequalities are also strict for $n \geq 2$, if strict inequality holds in (66).

The aformensioned results can be improved even further, if (19) holds on $U_0 := U(p_0, r_0 - d(p_0,p_1))$ instead of $U(p_0,r_0)$. In particular, G is said to satisfy the K_0- average Lipschitz condition on U_0, if for every two points $p,q \in U_0$ and any geodesic c connecting p,q with $d(p_0,p) + l(c) < r_0$, we have

$$\|DF_{c,p,q}(q) - DF(p)\| \leq \int_{d(p_0,p)}^{d(p_0,p)+l(c)} K_0(u)du. \tag{67}$$

Then, we have that

$$\bar{K}(u) \leq K(u) \text{ for each } u \in [0, r_0),$$

since $U_0 \subseteq U(p_0, r_0)$. Therefore, $\bar{K}$ can replace K in the preceeding results leading to an even finer convergence analysis. It is worth noticing that these advantages are obtained under the same computational cost, since in practice the computation of function L requires the computation of functions $L_0, \bar{K}$ and K as special cases. Moreover, upon using U_0 instead of $U(p_0, r)$ or $U(p_0, r_0)$, we still use the initial data, since $d(p_0, p_1)$ involves only $p_0, F(p_0)$ and $DF(p_0)$.

The rest of the results in [9, 24] can be improved along the same lines by also using K instead of L. We leave the details to the motivated reader.

5. Numerical examples

We present an example to show that $L_0 < K < L$.

Example 5.1. *Let* $X = Y = \mathbb{R}, D = \bar{U}(x_0, 1-\lambda), x_0 = 1, \lambda \in [0, \frac{1}{2})$. *Define function* F *on* D *by*

$$F(x) = x^3 - \lambda.$$

Then, we have $\xi = 1-\lambda, L_0 = 3(1-\lambda), K = 6(1+\frac{1}{L_0}), \bar{K} = 6(1+\frac{1}{L_0} - \xi)$ *and* $L = 6(2-\lambda)$. *Define*

$$\varphi_\alpha(t) = \frac{K}{2}t^2 - t + \xi,\ \varphi_{\alpha,0}(t) = \frac{L_0}{2}t^2 - t + \xi$$

and

$$\psi_\alpha(t) = \frac{L}{2}t^2 - t + \xi.$$

Then, since $L_0 < K < L, \bar{K} < K$, *we get* $\varphi_{\alpha,0}(t) < \varphi_\alpha(t) < \psi_\alpha(t)$ *and* $\bar{\varphi}_\alpha(t) < \varphi_\alpha(t)$. *Therefore, the new results have the advantages* ($\mathscr{A}$) *over the corresponding ones in* [9, 23, 24].

References

[1] S. Amat, I.K. Argyros, S. Busquier, R. Castro, S. Hilout, S. Plaza, On a bilinear operator free third order method on Riemannian manifolds, Appl. Math. Comput. 14 (2013), 7429–7444.

[2] I.K. Argyros, Approximating solutions of equations using Newton's method with a modified Newton's method iterate as a starting point, Rev. Anal. Numér. Théor. Approx. 36 (2007), 123–138.

[3] I.K. Argyros, Convergence and Applications of Newton-type Iterations, Springer, New York, (2008).

[4] I.K. Argyros, S. Hilout, Newton's method for approximating zeros of vector fields on Riemannian manifolds, J. Appl. Math. Comput. 29 (2009), 417–427.

[5] I.K. Argyros, S. Hilout, Extending the applicability of the Gauss-Newton method under average Lipschitz-type conditions, Numer. Algorithms 58 (2011), 23–52.

[6] I.K. Argyros, S. Hilout, Weaker conditions for the convergence of Newton's method, J. Complexity 28 (2012), 364–387.

[7] I.K. Argyros, S. Hilout, Extending the applicability of Newtons method on Lie groups, Appl. Math. Comput. 219 (2013), 10355–10365.

[8] I.K. Argyros, S. Hilout, Computational methods in nonlinear analysis. Efficient algorithms, fixed point theory and applications, World Scientific (2013).

[9] I.K. Argyros, A.A. Magreñán, Extending the applicabitity of Gauss-Newton method for convex composite optimization on Riemannian manifolds, Appl. Math. Comput. 249 (2014), 453–467.

[10] J.V. Burke, M.C. Ferris, A Gauss-Newton method for convex composite optimization, Math. Programming Ser A. 71 (1995), 179–194.

[11] O.P. Ferreira, O.P. Oliveira, Proximal point algorithm or Riemannian manifolds, Optimization, 51 (2002), 257–270.

[12] J.B. Hiriart-Urruty, C. Lemaréchal, Convex analysis and minimization algorithms (two volumes). I. Fundamentals, II. Advanced theory and bundle methods, 305 and 306, Springer–Verlag, Berlin (1993).

[13] C. Li, K.F. Ng, Majorizing functions and convergence of the Gauss-Newton method for convex composite optimization, SIAM J. Optim. 18 (2007), 613–642.

[14] C. Li, N. Hu, J. Wang, Convergence bahavior of Gauss-Newton's method and extensions to the Smale point estimate theory, J. Complexity 26 (2010), 268–295.

[15] Y. Ma, J. Koseck, S.S. Sastry, Optimization Criteria and Geometric Algorithms for Motion and Structure Estimation, Int. J. Comput. Vision 44 (1999), 219-249.

[16] Y. Nishimori, S. Akaho, Learning algorithms utilizing quasi-geodesic flows on the Stiefel manifold, Neurocomputing, 67 (2005), 106–135.

[17] T. Rapcsák, Smooth Nonlinear Optimization of $\mathbb{R}^n$, Nonconvex Optimization and Its Applications, (1997).

[18] S.M. Robinson, Extension of Newton's method to nonlinear functions with values in a cone, Numer. Math. 19 (1972), 341–347.

[19] R.T. Rockafellar, Convex analysis, Princeton Mathematical Series, 28, Princeton University Press, Princeton, N.J. (1970).

[20] S. Smale, Algorithms for solving equations. Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Berkeley, Calif., 1986), 172–195, Amer. Math. Soc., Providence, RI, 1987.

[21] S.T. Smith, Optimizationn techniques on Riemannian manifolds, IN: Fields Institute Communications, 3, 113–145, American Mathematical Society, Providence, RI (1994).

[22] C. Udriste, Convex Functions and Optimization Methods on Riemannian Manifolds, Mathematics and Its Applications, 297, Kluwer, Dordrecht (1994).

[23] X.H. Wang, Convergence of Newton's method and inverse function theorem in Banach space, Math. Comput. 68 (1999), 169–186.

[24] J.H. Wang, J.C. Yao, C. Li, Gauss-Newton mehotd for convex composite optimizations on Riemannian manifolds, J. Glob. Optim. 53 (2012), 5–28.

[25] W.Y. Yan, J. Lam, An approximate approach $\mathscr{H}_2$ optimal model reduction, IEEE Trans. Automat. Control 44 (1999), 1341–1358.