©2026 Ada Academica https://adac.ee
Eur. J. Math. Anal. 6 (2026) 10
doi: 10.28924/ada/ma.6.10

Extended Gauss-Newton Method on Riemannian Manifolds for Convex Composite Optimization Problems

Ioannis K. Argyros[1,*], Nirjal Shrestha[2], Samundra Regmi[3]

[1] *Department of Computing and Mathematical Sciences, Cameron University, Lawton, OK, 73505, USA*
iargyros@cameron.edu
[2] *Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden, CO, 80401, USA*
nirjal.shrestha@mines.edu
[3] *Department of Mathematics, University of Houston, Houston, TX, 77021, USA*
sregmi5@uh.edu
[*] *Correspondence: iargyros@cameron.edu*

Abstract. The Gauss-Newton method has been used to generate a sequence to approximate a zero of a vector field defined on a Riemannian manifold. The sufficient convergence conditions are based on L-average continuity conditions on the covariant derivative. In this work, the convergence conditions are weakened with advantages: tighter error distances and more precise information about the zero. These improvements are realized since tighter majorizing sequences are generated than in earlier studies. The scalar functions controlling the derivative are at least as tight as the specializations of earlier ones. Therefore, the new results are obtained without additional computational effort. Numerical applications are utilized to verify the convergence conditions.

1. Introduction

Let h be a real-valued convex function on $\mathbb{R}^n$ and F stands for Fréchet differentiable mapping from $\mathbb{R}^j$ into $\mathbb{R}^i$. A plethora of mathematical programming applications, such as minimax problems, goal programming, and convex inclusion problems [5, 10, 11, 14–18, 26, 27, 29, 30, 34, 35, 37], are related to the composite optimization problem on $\mathbb{R}^j$:

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

Let $K := \arg\min h$ be the set of minimum points of the function h. Then, the problem (1) is connected to the convex inclusion

$$F(x) \in K. \tag{2}$$

Received: 25 Nov 2025.
Key words and phrases. Riemannian manifold on a vector field; L-average Lipschitz continuity; Zero of a vector field; Semi-local convergence.

A solution x_* of (2) is such that $F(x_*) \in K$ and also solves (1). However, the solution x_* can be found in analytical form only in special cases. That explains why most solution methods for (1) or (2) are iterative. The most popular iterative method is the Gauss-Newton method [7,20,21,24].

Algorithm 1 : Algorithm GN (Ω, δ, x_0)

1: Let $\Omega \geq 1, \delta \in (0, \infty)$.

2: Define for $x \in \mathbb{R}^j$ the set

$$\Phi(x) = \Phi_\delta(x) := \{d \in \mathbb{R}^j : \|d\| \leq \delta, h(F(x_n) + F'(x_n)d) \leq h(F(x_n) + F'(x_n)d_1$$
$$\text{for each } d_1 \in \mathbb{R}^j \text{ with } \|d_1\| \leq \delta\}.$$

3: Pick $x_0 \in \mathbb{R}^j$.

4: For each $n = 0, 1, \ldots,$ having $x_0, x_1, \ldots, x_n$, determine x_{n+1} as follows:

5: If $0 \in \Phi(x_n)$, then STOP;

6: If $0 \notin \Phi(x_n)$, choose d_k such that

$$d_n \in \Phi(x_n) \text{ and } \|d_n\| \leq \Omega d(0, \Phi(x_n)).$$

7: Set $x_{n+1} = x_n + d_n$, where d(x,S) stands for the distance from x to S in $\mathbb{R}^j$.

Notice that $\Phi(x) \neq \emptyset$, and solve the convex optimization problem

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

Such a problem can be solved by standard methods such as the bundle method, the cutting plane, and the subgradient method [17]. The sequence $\{x_n\}$ generated by the **Algorithm GN** (Ω, δ, x_0) converges at a quadratic rate to x_* [7, 18, 24, 37]. Numerous problems in robot manipulation [15], system balancing [16], computer vision [25], electronic structure computation [11], machine learning [27], and maximum likelihood estimation can be expressed using numerical linear algebra as optimizing a smooth function defined on a Riemannian manifold [1–3, 6, 8, 9, 12, 13, 20, 22, 36]. The trust-region method, the conjugate gradient method, the steepest descent method, and other methods have been extended on Riemannian manifolds [1–3, 6, 8, 9, 12, 13, 20, 22, 36].

Let us extend the problem (1) on a Riemannian manifold M as:

$$\min_{q \in M} f(q) : h(F(q)), \tag{3}$$

where $F : M \to \mathbb{R}^i$ is a differentiable mapping. Then, the convex inclusion problem (2) reads as:

$$F(q) \in K. \tag{4}$$

Then, the extended Gauss-Newton algorithm for the convex composite optimization problem (3) [36] is defined as:

Algorithm 2 : Algorithm GNR (Ω, δ, q_0)

1: Let $\Omega \geq 1, \delta \in (0, \infty)$, and $q_0 \in M$.
2: For each $n = 0, 1, \ldots$, having $q_0, q_1, \ldots, q_n$, determine q_{n+1} as follows:
3: If $0 \in A(q_n)$, then STOP;
4: If $0 \notin A(q_n)$, choose u_n such that $u_n \in A(q_n)$ and $\|u_n\| \leq \Omega d(0, A(q_n))$ and take $q_{n+1} = \exp_{q_n}^{u_n}$ where for $q \in M$, $A_\delta(q) = A(q)$ is defined by:

$$A(q) = \{u \in T_qM : \|u\| \leq \delta, h(F(q) + F'(q)u) \leq h(F(q) + F'(q)u_1),$$
$$\text{for each } u_1 \in T_qM \text{ with } \|u_1\| \leq \delta\}$$

The local convergence analysis of the sequence $\{q_n\}$ generated by **Algorithm GNR** (Ω, δ, q_0) is our concern in this paper. In particular, we develop a technique that without additional convergence than in [36] we provide

(1) A larger radius of convergence.
(2) Tighter error estimates on $\|q_n - x_*\|$.

1.1. **Related Work.** Let us introduce $L-$average Lipschitz conditions required for the local convergence of **Algorithm GNR** (Ω, δ, q_0). From now on, we denote a positive value, increasing integrable function on $[0, \infty)$ by "L".

Definition 1.1: Let $R > 0$ and $q_0 \in M$. The operator F' is said to satisfy the center $L-$average Lipschitz condition on $B(q_0, R)$, if for each point $q \in B(q_0, R)$ and any geodesic γ connecting q_0, q with $\ell(\gamma) < R$, the following holds:

(C_1) $\|F'(q)P_{\gamma,q,q_0} - F'(q_0)\| \leq \int_0^{\ell(\gamma)} L_0(v)dv.$

(C_2) $a\int_0^s L_0(v)dv - 1 = 0, s \geq 0.$

Let $a > 0$ be such that, the equation has positive solutions. Denote the least such solution by R_a. Define the region $B_0 = B(q_0, R_0)$, where $R_0 = \min\{R, R_0\}$.

Definition 1.2: The operator F' is said to satisfy $L-$average Lipschitz condition on $B(q_0, R_a)$, if for any two points $q_1, q_2 \in B(q_0, R_0)$ and any geodesic connecting q_1, q_2 with $d(q_0, q_1) + \ell(\gamma) < R_0$, the following holds:

(C_3) $\|F'(q_2)P_{\gamma,q_1,q_2} - F'(q_1)\| \leq \int_{d(q_0,q_1)}^{d(q_0,q_1)+\ell(\gamma)} L(v)dv.$

Defintion 1.3: The operator F' is said to satisfy the L_1-average Lipschitz conditions on $B(q_0, R)$, if for any two points $q_1, q_2 \in B(q_0, R)$ and any geodesic connecting q_1, q_2 with $d(q_0, q_1) + \ell(\gamma) < R$, the following holds:

(C_4) $\|F'(q_2)P_{\gamma,q_1,q_2} - F'(q_1)\| \leq \int_{d(q_0,q_1)}^{d(q_0,q_1)+\ell(\gamma)} L_1(v)dv.$

Remark 1.4: It follows by these definitions that since $R_0 \leq R$, the following estimates hold:

$$L_0(v) \leq L_1(v), \tag{5}$$

and

$$L(v) \leq L_1(v), \tag{6}$$

for each $v \in [0, R_0]$. It is worth noting that $L_0 = L_0(q_0, R)$ i.e., depends on q_0 and R,$L_1 = L_1(q_0, R)$ but $L = L(q_0, R_0)$. In the local convergence analysis of the **Algorithm GNR**(Ω, δ, q_0) [36], the condition (C_4) is used.

1.2. **Contributions.** In view of the estimates (5) and (6), the benefits 1. and 2. stated in the introduction can be obtained if the function L_1 is replaced by at least as tight L_0 and L in the proof of the main Theorem 3.1 in [36]. Indeed, let us look at the functions g_0, g and g_1 defined on the interval $[0, \infty)$ for some $\mu \geq 0$:

$$g_0(t) = \mu - t + a \int_0^t L_0(v)(t - v)dv, \tag{7}$$

$$g(t) = \mu - t + a \int_0^t L(v)(t - v)dv, \tag{8}$$

and

$$g_1(t) = \mu - t + a \int_0^t L_1(v)(t - v)dv. \tag{9}$$

Then, in view of (5) - (9), we see that for each $t \in [0, R_0]$

$$g_0(t) \leq g_1(t), \tag{10}$$

and

$$g(t) \leq g_1(t). \tag{11}$$

Moreover, define the scalar sequences $\{s_i\}, \{t_i\}$ for $s_0 = t_0 = 0$, $s_1 = t_1 = \mu$ by

$$\begin{aligned} s_2 &= s_1 - \frac{a \int_0^{s_1 - s_0} L_0(s_0 + v)(s_1 - s_0 - v)dv}{g_0'(s_1)}, \\ s_{i+1} &= s_i - \frac{a \int_0^{s_i - s_{i-1}} L_0(s_{i-1} + v)(s_i - s_{i-1} - v)dv}{g_0'(s_i)}, \end{aligned} \tag{12}$$

$$\begin{aligned} t_2 &= t_1 - \frac{a \int_0^{t_1 - t_0} L_1(t_0 + v)(t_1 - t_0 - v)dv}{g_1'(t_1)}, \\ t_{i+1} &= t_i - \frac{a \int_0^{t_i - t_{i-1}} L_1(t_{i-1} + v)(t_i - t_{i-1} - v)dv}{g_1'(t_i)}. \end{aligned} \tag{13}$$

The sequence $\{t_i\}$ can also be given in a more condensed form than (13):

$$t_{i+1} = t_i - \frac{g_1(t_i)}{g_1'(t_i)}. \tag{14}$$

Then, in view of (10 - (13), a simple inductive argument shows that

$$0 \leq s_i \leq t_i, \tag{15}$$

$$0 \leq s_{i+1} - s_i \leq t_{i+1} - t_i, \tag{16}$$

and

$$s_* = \lim_{i\to\infty} s_i \leq t_* = \lim_{i\to\infty} t_i. \tag{17}$$

Provided that s_* and t_* exist (see also Section 4). It follows by (15) - (17) that the advantages (1) and (2) hold. Note that the new functions L_0 and L replacing L_1 used in [36] (see Section 3) are specializations (of L_1). Hence, in practice, the computational cost is the same, i.e., no additional conditions are required.

1.3. **Limitations.** The convergence conditions for GNR (Ω, δ, q_0) are sufficient conditions. It is interesting to see if necessary conditions can also be found.

The rest of the work is structured as follows: Section 2 contains the mathematical background. The semi-local analysis of convergence can be found in Section 3 and the numerical examples in Section 4. The conclusions appear in Section 5. Finally, we have included the code used in the numerical section after the references in the Appendix.

2. Mathematical Background

The definitions and notations on manifolds are standard [10, 14]. But we reproduce some of them in this section to make the chapter as self-contained as possible. More details can be found in [10, 14]. Let us consider M to stand for a completely connected $m-$dimensional Riemannian manifold equipped with the Levi-Civita relation δ on M. Take $q \in M$, and let T_qM be the tangent space at q to M. The scalar product $< \cdot, \cdot >$ on T_qM with related norm $\|\cdot\|_q = \|\cdot\|$. Let $q_1, q_2 \in M$ with $q_1 \neq q_2$ and $\gamma : [0,1] \to M$ be a piecewise smooth curve connecting q_1 to q_2. Then, the arc-length of γ is defined by $\ell(\gamma) = \int_0^1 \|\gamma'(t)\| dt$. Moreover, the Riemannian distance from q_1 to q_2 is defined by $d(q_1, q_2) = \inf_{\gamma} \ell(\gamma)$. The infimum is taken over all piecewise smooth curves $\gamma : [0,1] \to M$ connecting the exponential map at q. That is $\exp : T_qM \to M$ exists on T_qM.

A geodesic γ in M connecting q_1 and q_2 is called minimizing, provided its arc length equals the Riemannian distance between q_1 and q_2. Then, clearly the curve $\gamma : [0,1] \to M$ is a minimizing geodesic connecting q_1 and q_2 if and only if there exists a vector $u \in T_qM$ such that $\|u\| = d(q_1, q_2)$ and $\gamma(t) = \exp_q(tu)$ for each $t \in [0,1]$.

Let $\gamma : \mathbb{R} \to M$ be a C^∞ curve. Then, the parallel transport P is defined by

$$P_{\gamma,\gamma(\beta),\gamma(\alpha)}(u) = W(\gamma(\beta))$$

for each $\alpha, \beta \in \mathbb{R}$ and $u \in T_\gamma(\alpha)M$, where W is the unique C^∞ vector such that $\delta_{\gamma'(t)=0}$ and $W(\gamma(\alpha) = u$. It follows that $P_{\gamma,\gamma(\beta),\gamma(\alpha)}$ is an isometry mapping $T_\gamma(\alpha)M$ into $T_\gamma(\beta)M$ and for $\alpha, \beta, \beta_1, \beta_2 \in \mathbb{R}$,

$$P_{\gamma,\gamma(\beta_2),\gamma(\beta_1)} \circ P_{\gamma,\gamma(\beta_1),\gamma(\alpha)} = P_{\gamma,\gamma(\beta_2),\gamma(\alpha)},$$

and

$$P^{-1}_{\gamma,\gamma(\beta),\gamma(\alpha)} = P_{\gamma,\gamma(\alpha),\gamma(\beta)}.$$

If γ is a minimizing geodesic relating q_1 to q_2, we abbreviate P_{γ,q_1,q_2} as P_{q_1,q_2}. By $C^1(TM)$ we denote the set of all C^1-vector fields on M and $C^p(M)(p=0,1)$ the set of all C^p-functions from M into $\mathbb{R}$, with C^0 standing for continuous mappings.

Let $F_p \in C^1(M), p=1,2,\ldots,i$ and $F : M \to \mathbb{R}^i$ be a C^1 function so that $F=(F_1,F_2,\ldots,F_i)$. Then, the derivative of the mapping F along the vector field V is defined by [9, 22]

$$\delta_V F = (\delta_V F_1, \delta_V F_2, \ldots, \delta_V F_i) = (V(F_1), V(F_2), \ldots, V(F_i)).$$

Consequently, $F' : C^1(TM) \to (C^0(M))^i$ is given by

$$F'(V) = \delta_V F \text{ for each } V \in C^1(TM). \tag{18}$$

Let $q \in M$. Then, the notation $F'(q)$ stands for the derivative of F at q. Let $u \in C^1(TM)$ be such that $V(q)=u$. Moreover, consider any non-trivial smooth curve $\gamma : (-\lambda,\lambda) \to M(\lambda > 0)$ with $(\gamma(0)=q)$ and $\gamma'(0)=u$. Then, it follows that

$$\begin{aligned} F'(q)u &= F'(V)(q) \\ &= \delta_V F(q) = \Big(\frac{d}{dt}(F\circ\gamma)(t)\Big)_{t=0}. \end{aligned} \tag{19}$$

The definition of quasi-regular point was initiated in [21] for the study of (1). Then, in [36], this notion is extended to Riemannian manifolds. Let K be a closed convex set in $\mathbb{R}^i$, and consider the inclusion

$$F(q) \in K. \tag{20}$$

Then, for $q \in M$

$$A(q) := \{u \in T_qM : F(q) + F'(q)u \in K\}. \tag{21}$$

If K is the set of all minimum points of h and there exists $u_0 \in T_qM$ with $\|u_0\| \le \delta$ such that $u_0 \in A(q)$. Then, for each $u \in T_qM$ with $\|u\| \in \delta$, we have

$$u \in A_\delta(q) \iff u \in A(q) \iff u \in A_\infty(q). \tag{22}$$

Definition 2.1: A point $q_0 \in M$ is called quasi-regular for the inclusion (18) if there exists $R > 0$ and an increasing positive-valued function b on $[0,R)$ such that

$$A(q) \neq \emptyset \text{ and } d(0, A(q)) \le b(d(q_0,q))d(F(q),K) \text{ for each } q \in B(q_0,R). \tag{23}$$

As in [21], let R_{q_0} be the supremum of R such that (23) holds for some increasing positive-valued function b on the interval $[0,R)$. Let $R \in [0,R_{q_0}]$ and let $S(q_o,R)$ stand for the set of all increasing positive-valued function b on $[0,R)$ such that (23) holds. Define

$$b_{q_0}(t) = \text{infimum}\{b(t) : b \in S_{R_{q_0}}(q_0) \text{ for each } t \in [0,R_{q_0})\}. \tag{24}$$

Note that, each $b \in S_R(q_0)$ can be extended to an element of $S(q_0,R_{q_0})$ provided that $\lim_{t\to R^-} b(t) < \infty$. So,

$$b_{q_0}(t) = \text{infimum}\{b(t) : b \in S(q_0,R)\} \text{ for each } t \in [0,R). \tag{25}$$

From now on, we denote R_{q_0} as the quasi-regular radius and b_{q_0} as the quasi-regular bound function of the quasi-regular point q_0.

3. Semi-local Convergence Analysis

Let $q \in M$ be a quasi-regular point of the inclusion (2) with quasi-regular radius R_{q_0} and the quasi-regular bound function b_{q_0}. Let $\Omega \in [1, \infty)$ be fixed. Define

$$\mu := \Omega b_{q_0} d(F(q_0), K). \tag{26}$$

Moreover, for $R \in (0, R_{q_0}]$, we define

$$a_0(R) := \sup\Big\{\frac{\Omega b_{q_0}(t)}{\Omega b_{q_0}(t)\int_0^1 L_0(v)dv + 1} : \mu \le t \le R\Big\}. \tag{27}$$

By convention, we set $\sup \emptyset = -\infty$.

Next, we present a general convergence condition for the sequence $\{s_i\}$ given by the formula (12).

Proposition 3.1. *Suppose* (C_2) *and the following conditions hold:*

(C_5) *There exists* $\bar{s} \in [0, R_0)$ *such that for each* $i = 0, 1, 2, \ldots$

$$a\int_0^{s_i} L_0(v)dv < 1 \text{ and } s_i \le \bar{s}_0.$$

(C_6) $\mu \le \min\{b_a, \delta\}$.

Then, the following assertions hold:

$$0 \le s_i \le s_{i+1} \le \bar{s},$$

and there exists $s_* \in [0, \bar{s}]$ *such that* $\lim_{i\to\infty} s_i = s_*$.

Proof. It is well known that s_* is the unique least upper bound of the sequence $\{s_i\}$. It follows by the definition of $\{s_i\}$ and the condition (C_5) that the first assertion holds. So, the sequence $\{s_i\}$ is nondecreasing and bounded from above by $\bar{s}$. Thus, it converges to s_*. □

Next, we present the semi-local convergence analysis of the sequence $\{q_n\}$ generated by the Algorithm R (Ω, δ, q_0). In particular, we show that the sequence $\{s_i\}$ majorizes $\{q_i\}$.

Theorem 3.1. *Suppose that the conditions* $(C_1), (C_2), (C_3)$ *and* (C_5) *hold. Then, the sequence* $\{q_i\}$ *generated by the Algorithm R* (Ω, δ, q_0) *is well defined in* $B(q_0, s_*)$*, remains in* $B(q_0, s_*)$ *for each* $i = 0, 1, 2, \ldots$*, and converges to some* q_* *such that* $F(q_*) \in K$.

Moreover, the following assertions hold

$$d(q_i, q_*) \le s_* - s_i \tag{28}$$

and

$$F(q_i) + F'(q_i)q_i \in K. \tag{29}$$

Proof. Mathematical induction is used to show that for each $i = 0, 1, 2, \ldots$

$$\|q_{i+1} - q_i\| \leq s_{i+1} - s_i. \tag{30}$$

Notice that by (12), we have

$$\mu \leq s_i \leq s_* \leq R_0 \leq R_{q_0}. \tag{31}$$

It follows by the quasi-regularity condition that

$$A(q) \neq 0 \text{ and } d(0, A(q) \leq b_{q_0}(d(q_0, q))d(F(q), K) \text{ for each } q \in B(q_0, R). \tag{32}$$

So,

$$\begin{aligned} A(q_0) \neq 0 \text{ and } \Omega d(0, A(q_0) &\leq \Omega b_{q_0}(d(q_0, q_0))d(F(q_0), K) \\ &= \Omega b_{q_0}(0)d(F(q_0), K) = \mu = \delta. \end{aligned} \tag{33}$$

But from $d(0, A(q_0) \leq \delta$, since $\Omega \geq 1$ there exists $u \in T_q M$ with $\|u\| \leq \delta$ so that $F(q_0) + F'(q_0)u \in K$. Hence, we get

$$A_\delta(q_0) = \{u \in T_{q_0}M : \|u\| \leq \delta, F(q_0) + F'(q_0)u \in K\}$$

and

$$d(0, A_\delta(q_0)) = d(0, A(q_0)).$$

Hence, by the Algorithm $\mathrm{R}(\Omega, \delta, q_0)$ and (32), there exists $u_0 \in A_\delta(q_0)$ so that

$$\begin{aligned} \|u_0\| &\leq \Omega d(0, A_\delta(q_0)) = \Omega d(0, A(q_0) \\ &\leq \Omega b_{q_0}(0)(F(q_0), K) = \mu = s_1 - s_0, \end{aligned}$$

showing (30) and (31) for $i = 0$. Suppose that (30) and (31) hold for each $i = 0, 1, 2, \ldots, j-1$. Define the geodesic $\gamma : [0,1] \to M$ as

$$\gamma(\theta) = \exp_q \theta u_{j-1} \text{ for each } \theta \in [0,1]. \tag{34}$$

Moreover, we have by the inductive hypothesis that

$$\begin{aligned} d(q_0, q_j) &\leq \|q_j - q_{j-1}\| + \|q_{j-1} - q_{j-2}\| + \ldots + \|q_1 - q_0\| \\ &\leq s_j - s_{j-1} + s_{j-1} - s_{j-2} + \ldots + s_1 - s_0 = s_j, \end{aligned} \tag{35}$$

and

$$d(q_{j-1}, q_0) \leq s_{j-1} \leq s_j. \tag{36}$$

Then, by (34) $\gamma(\theta) \in B(q_0, s_*) \subseteq B(q_0, R)$, so we obtain

$$A(q_j) \neq 0 \text{ and } d(0, A(q_j)) \leq b_{q_0}(d(q_j, q_0))d(F(q_j), K). \tag{37}$$

Hence, (32) holds for $q = q_j$.

Next, we shall show that

$$\Omega d(0, A(q_j)) \leq s_{j+1} - s_j, \tag{38}$$

using the conditions (C_1) and (C_2) on $B(q_0, s_*)$. Indeed, it also follows by (37) in turn that

$$\begin{aligned}\Omega d(0, A(q_j)) &\leq \Omega b_{q_0}(d(q_0, q_j))d(F(q_j), K)\\ &\leq \Omega b_{q_0}(d(q_0, q_j))\|F(q_j) - F(q_{j-1} - F'(q_{j-1}u_{j-1}\|\\ &\leq \Omega b_{q_0}(d(q_0, q_j))\|\int_0^1 (F'(\gamma(\theta))P_{\gamma,\gamma(\theta),q_{j-1}} - F'(q_{j-1})u_{j-1}d\theta\|\\ &\leq \Omega b_{q_0}(d(q_0, q_j))\int_0^1 \Big(\int_{d(q_{j-1},q_j)}^{\theta\|u_{j-1}\|+d(q_{j-1},q_0)} L(v)dv\Big)\|u_{j-1}\|d\theta\\ &\leq \Omega b_{q_0}(d(q_0, q_j))\int_0^{\|u_{j-1}\|} L(d(q_{j-1}, q_0) + v)(\|u_{j-1}\| - v)dv\\ &\leq \Omega b_{q_0}(s_j)\int_0^{\|u_{j-1}\|} L(s_{j-1} + v)(\|u_{j-1}\| - v)dv\\ &\leq \Omega b_{q_0}(s_j)\int_0^{s_j - s_{j-1}} L(s_{j-1} + v)(s_j - s_{j-1} - v)dv.\end{aligned} \tag{39}$$

Furthermore, by the conditions (C_1), (26), and (27), we have

$$\frac{\Omega b_{q_0}(s_j)}{a_0(R)} \leq (1 - a_0(R)\int_0^{s_j} L_0(v)dv)^{-1}$$

and since $a \geq a_0(R)$, we get

$$\begin{aligned}\frac{\Omega b_{q_0}(s_j)}{a} &\leq (1 - a_0(R)\int_0^{s_j} L_0(v)dv)^{-1}\\ &= -(q_0'(s_j))^{-1}.\end{aligned} \tag{40}$$

Thus, the assertion (38) holds, and by (12), (38) - (40), we get $s_{j+1} - s_j \leq \mu \leq \delta$, so $d(0, A(q_j)) \leq \delta$. It follows that there exists $u \in T_{q_j}M$ so that

$$F(q_j) + F'(q_j)u \in K \text{ for } \|u\| \leq \delta,$$

$$A_\delta(q_j) = \{u \in T_{q_j}M : \|u\| \leq \delta, F(q_j) + F'(q_j)u \in K\},$$

and

$$d(0, A_\delta(q_j)) = d(0, A(q_j)).$$

Select $u_j \in A_\delta(q_j)$. It follows by the Algorithm R (Ω, δ, q_0) that (29) holds if $i = j$, $\|u_j\| \leq \Omega d(0, A_\delta(q_j)) = \Omega d(0, A(q_j))$ and (38) that (30) holds. Therefore, the sequence $\{s_j\}$ majorizing for $\{q_j\}$ and is Cauchy. It follows that the sequence $\{q_j\}$ is Cauchy too and it converges to some q_* such that $F(q_*) \in K$. Finally, by (40) and the triangle inequality, we have

$$\|q_{i+n} - q_i\| \leq s_{i+n} - s_i.$$

By letting $n \to \infty$ in (40), we show (28). □

Remark: The limit point s_* can be replaced by R_0 or $\bar{s}$ in the **Theorem 3.1**.

4. Comparisons and Examples

In this section, we compare the scalar sequences (12) and (13). First, we list the result related to the condition (C_4) and the sequence $\{t_i\}$ (see Theorem 3.1 in [36]).

(H_1) The condition (C_4) holds on $B(q_0, s_*)$.

(H_2) The equation $g_1(t) = 0$ has a smallest positive solution denoted by ρ_*.

(H_3) $\mu =$ minimum $\{b_a, \delta\}$ and $\rho_* \leq R$.

Theorem 4.1. *Suppose the conditions $(H_1) - (H_3)$ hold. Then, the sequence $\{q_i\}$ generated by the Algorithm R (Ω, δ, q_0) is well defined in $B(q_0, \rho_*)$, remains in $B(q_0, \rho_*)$ for each $i = 0, 1, 2, \ldots$ and converges to some q_* such that $F(q_*) \in K$.*

Moreover, the following assertions hold for each $i = 0, 1, 2, \ldots$

$$d(q_{i+1}, q_i) \leq t_{i+1} - t_i \tag{41}$$

$$d(q_i, q_*) \leq \rho_* - t_i \tag{42}$$

and

$$F(q_i) + F'(q_i)u_i \in K, \tag{43}$$

where $\rho_* = \lim\limits_{i\to\infty} t_i$ and the sequence $\{t_i\}$ is given by (13).
Next, we compare the sequence $\{s_i\}$ to $\{t_i\}$.

Proposition 4.1. *Suppose that the conditions of Theorem 4.1 hold. Then, the sequences $\{s_i\}, \{t_i\}$ and $\{q_i\}$ exist and the following error estimates hold*

$$0 \leq s_i \leq t_i, \tag{44}$$

$$\|q_{i+1} - q_i\| \leq s_{i+1} - s_i \leq t_{i+1} - t_i, \tag{45}$$

and

$$0 \leq s_* \leq \rho_*. \tag{46}$$

Proof. By Theorem 4.1, the conditions of Theorem 3.1 hold for $\bar{s} = \rho_*$, so the sequences $\{s_i\}, \{t_i\}$ and $\{q_i\}$ exist and

$$\|q_{i+1} - q_i\| \leq s_{i+1} - s_i \tag{47}$$

and

$$\|q_{i+1} - q_i\| \leq s_{i+1} - s_i \leq t_{i+1} - t_i, \tag{48}$$

Moreover, by (5), (6), (12), and (13) assertions (44) and (45) hold. Finally by letting $i \to \infty$ in (44), estimate (46) follows. □

Remark A: In Proposition 4.1, we showed that Conditions of Theorem 4.1 $\implies$ Conditions of Theorem 3.1. But this implication does not necessarily hold vice versa. Hence, the new results are heavier and with advantages as stated in the introduction (see also (41) - (47)). It is also worth noting that these advantages are obtained under the same computational cost, since in practice the computation of the function L_1 used in the old Theorem 4.1 requires that of L_0 and L used in our Theorem 3.1 as special cases.

Next, we complete this section by using three concrete examples on the vector fields from $\mathbb{R} \to \mathbb{R}, \mathbb{R}^2 \to \mathbb{R}^2$ and $\mathbb{R}^3 \to \mathbb{R}^3$, respectively provided that $K = \{0\}$. The convergence conditions are validated in these examples for the Gauss-Newton method.

Example 4.1. Let us consider the vector field F from $\mathbb{R}$ to $\mathbb{R}$ defined by

$$F(p) = p^3 - d,$$

where $F : D \subset \mathbb{R} \to \mathbb{R}, d \in (0, \frac{1}{2}\Big)$ and $D = B(p_0, 1-d)$ for $p_0 = 1$. Then, it follows by the Definitions 1.1, 1.2, 1.3, and majorizing functions g_0, g and g_1 given, respectively by (7) - (9) that since $\|F'(p_0)^{-1}\| = \frac{1}{3} = a, \|F'(p_0)^{-1}F(p_0)\| = \frac{1-d}{3} = \mu$, for $l_0 = 3-d, l = 2\left(1 + \frac{1}{3-d}\right)$ and $l_1 = 2(2-d)$, we can choose $L_0(t) = l_0, L(t) = l$and $L_1(t) = l_1$. Moreover, we have

$$g_0(t) = \mu - t + \frac{L_0}{2} t^2 \tag{49}$$

$$g(t) = \mu - t + \frac{L}{2} t^2 \tag{50}$$

$$g_1(t) = \mu - t + \frac{L_1}{2} t^2. \tag{51}$$

Furthermore the majorizing sequences corresponding to the functions g and g_1 are (see (12) - (14).

$$\begin{aligned} s_0 = 0, s_1 = \mu, s_2 &= s_1 + \frac{l_0 (s_1 - s_0)^2}{2(1 - l_0 - s_1)}, \\ s_{i+1} &= s_i + \frac{l (s_i - s_{i-1})^2}{2(1 - l_0 s_i)}, \quad i = 2, 3, \ldots, \end{aligned} \tag{52}$$

$$t_0 = 0, t_1 = \mu, t_{i+1} - t_i = \frac{l_1 (t_i - t_{i-1})^2}{2(1 - l_1 t_i)} = -\frac{g_1(t_i)}{g_1'(t_i)}, \quad i = 1, 2, \ldots. \tag{53}$$

The sequence $\{t_i\}$ appears as majorizing in the study of Newton-type methods [1]. The convergence condition for it is given by the celebrated Newton-Kantorovich semi-local convergence condition [1]

$$2 l_1 \mu \leq 1. \tag{54}$$

But this condition is not satisfied, since

$$2 l_1 \mu = \frac{4}{3}(2-d)(1-d) > 1 \text{ for each } d \in (0, \frac{1}{2}).$$

Thus, the results in [1, 36] cannot assure the convergence of the method to the singularity $p^* = \sqrt[3]{d}$. But the method converges to p^* say if $d = 0.47$. This is happening, since (54) is only a sufficient condition. These observations indicate that (54) can be weakened and this is achieved by switching (53) by (52). The convergence of (52) depends on the corresponding to (54) sufficient convergence condition which is

$$2 l \mu \leq 1. \tag{55}$$

Notice that this condition is satisfied since

$$2l\mu \leq 1 \qquad \text{for each } d \in [0.461983, 0.5).$$

It is worth noticing that (5) and (6) hold as strict inequalities, since

$$l_0 < l_1 \tag{56}$$

and

$$l < l_1 \qquad \text{for each } d \in (0, 1). \tag{57}$$

Hence, the applicability of the method is extended under the same computational effort, since the computation of l_1 requires that of l_0 and l as specializations. Moreover, the sequences $\{s_i\}$ and $\{s_i - s_{i-1}\}$ are tighter than $\{t_i\}, \{t_i - t_{i-1}\}$, respectively (see (15)-(17), (56) and (57)). Let us demonstrate this further, say for $d = 0.6$, when both (54) and (55)) are satisfied.

i	s_i	t_i	$s_i - s_{i-1}$	$t_i - t_{i-1}$
0	0	0		
1	0.1333333333	0.1333333333	0.1333333333	0.1333333333
2	0.1703703703	0.1730496453	0.0370370370	0.0397163120
3	0.1736579102	0.1773338485	0.0032875398	0.0042842031
4	0.1736841630	0.1773848871	0.0000262528	0.0000510385
5	0.1736841647	0.1773848943	0.0000000016	0.0000000072
6	0.1736841647	0.1773848943	0	0

TABLE 1. Comparison of majorizing sequences.

Python programming language was used for the computation above. The code could be run in any Python compiler, in particular, Visual Studio Code application was used to run the given Python code (Python version 3.13.1), which took less than a second to execute the code.

The table justifies the claims made above. Notice also that

$$s^* = \lim_{i \to \infty} s_i \leq t^* = \lim_{i \to \infty} t_i,$$

where

$$s^* = \frac{1 - \sqrt{1 - 2l\mu}}{l} = 0.1736841647,$$

and

$$t^* = \frac{1 - \sqrt{1 - 2l_1\mu}}{l_1} = 0.1773848943,$$

so

$$s^* < t^*$$

as expected.

Example 4.2. Let us select the vector field F from $\mathbb{R}^2$ to $\mathbb{R}^2$ defined by

$$F(q) = F\left(q_1, q_2\right)^{tr} = \left(\frac{\cos(q_1) + 20q_1}{20}, q_2\right)^{tr} \tag{58}$$

and equipped with the norm $\|\cdot\|_\infty$. Then, by (58), we have

$$F'(q) = \begin{bmatrix} \dfrac{-\sin q_1 + 20}{20} & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$F''(q) = \left[\begin{array}{cc|cc} \dfrac{-\cos q_1}{20} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Select $q_0 = (0,0)^{tr}$. Then, as in Example 4.1, we have

$$\|F'\left(q_0\right)^{-1}\| = 1 = a, \|F'\left(q_0\right)^{-1} F\left(q_0\right)\| = 0.05 = \mu$$

and

$$\begin{aligned} \|F''(q)\| &= \max\left(\left|\frac{-\cos q_1}{20}\right|, 0\right) = \left|\frac{\cos q_1}{20}\right| \leq \frac{1}{20} \\ &= l_1. \end{aligned}$$

We can also take $l_0 = l = \frac{1}{20}$. Clearly, with these values $2l_1\mu \leq 1$ and $2l\mu \leq 1$ hold as strict inequalities. Hence, there exists a $p^* \in B(p_0, t^*)$ solving (4).

Example 4.3. Let us choose the vector field, V from $\mathbb{R}^3 \to \mathbb{R}^3$ as

$$V(q) = V\begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} = \begin{bmatrix} -q_2 \\ q_1 - q_2q_3 \\ q_1q_2q_3 \end{bmatrix}$$

equipped with the Frobenius norm. Set $F = V/S^2$. Clearly, we have $V/S^2(q) \in T_qS^2$ for each $q \in S^2$. Then $F'(q)$ in the basis

$$\mathrm{B} = \left\{\begin{pmatrix} -q_3 \\ 0 \\ q_1 \end{pmatrix}, \begin{pmatrix} 0 \\ -q_3 \\ q_2 \end{pmatrix}\right\}$$

is defined for $q_3 \neq 0$ by

$$F'(q) = -\frac{1}{q_3}\begin{pmatrix} \gamma_{1,1}(q) & \gamma_{1,2}(q) \\ \gamma_{2,1}(q) & \gamma_{2,2}(q) \end{pmatrix},$$

where

$$\gamma_{ij}(q) = q_j\left(F_{i,q_3}(q) - \sum_{k=1}^{3} q_kF_{k,q_3}(q)q_i\right) - q_3\left(F_i, q_j(q) - \sum_{k=1}^{3} q_kF_{k,q_j}(q)q_i\right), \quad i,j = 1,2$$

and F_{i,q_j} is the partial derivative of F_i with respect to q_j. Thus, we get

$$F'(q) = -\begin{bmatrix} q_1q_2\left(q_1^2+1\right) & 1 + q_1^2\left(q_2^2 + q_3^2 - 1\right) \\ q_2^2 + q_1^2\left(q_2^2 - 2\right) + q_3^2 - 1 & q_1q_2\left(q_2^2 + q_3^2 - 3\right) \end{bmatrix}.$$

Hence, by Lagrange's multiplier method, we obtain that the Definitions 1.1, 1.2,1.3 hold if $L_0(t) = L(t) = L_1(t) = l$, where

$$l = \sup\left\{F'(q) : q_1^2 + q_2^2 + q_s^2 = 5\right\} = 11.$$

Choose $q_0 = (2, -0.0013091)^{tr}$ to obtain

$$\|F'(q_0)^{-1}\| = 1.00779 = a,$$

and

$$\|F'(q_0)^{-1} F(q_0)\| = 0.0013193 = \mu.$$

So, again $2l_1\mu \leq 1$ and $2l\mu \leq 1$ hold as strict inequalities. Therefore, there exists $p^* \in B(q_0, t^*)$ solving (4).

5. Conclusion

A new semi-local convergence analysis is presented for the Gauss-Newton method defined on Riemannian manifolds to approximate a zero of a vector field. The analysis is based on more accurate majorant functions which are used to control the covariant derivative. The sufficient convergence conditions are weaker than before [36]. This way the resulting majorizing sequences are tighter resulting in at least as few iterations to reach a predetermined error tolerance. The information on the location of the zero is more accurate. The majorant functions used are specializations of earlier ones. Hence, the advantages are obtained without additional conditions or computational effort. The same technique can be used to extend the applicability of other methods in an analogous way [1, 6-28, 30-35, 37]. This is the direction of future research.

Competing interests: The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] R.L. Adler, J.P. Dedieu, J. Margulies, M. Martens, M. Shub, Newton's method on Riemannian manifolds and a geometric model for the human spine, IMA J. Numer. Anal. 22 (2002), 359–390. https://doi.org/10.1093/imanum/22.3.359.

[2] S. Amat, I.K. Argyros, S. Busquier, R. Castro, S. Hilout, S. Plaza, Traub-type high order iterative procedures on Riemannian manifolds, SeMA J. 63 (2014), 27–52. https://doi.org/10.1007/s40324-014-0010-0.

[3] I.K. Argyros, An improved unifying convergence analysis of Newton's method in Riemannian manifolds, J. Appl. Math. Comput. 25 (2007), 345–351. https://doi.org/10.1007/BF02832359.

[4] I.K. Argyros, Newton's method on Lie groups, J. Appl. Math. Comput. 31 (2009), 217–228. https://doi.org/10.1007/s12190-008-0203-8.

[5] I.K. Argyros, The theory and applications of iteration methods, CRC Press (2022). https://doi.org/10.1201/9781003128915.

[6] J. Balogh, T. Csendes, T. Rapcsák, Some global optimization problems on Stiefel manifolds, J. Global Optim. 30 (2004), 91–101. https://doi.org/10.1007/s10898-004-0574-9.

[7] J.V. Burke, M.C. Ferris, A Gauss–Newton method for convex composite optimization, Math. Program. 71 (1995), 179–194. https://doi.org/10.1007/BF01585997.

[8] R.A. Castro, S. Amat, I.K. Argyros, S. Busquier, Chebyshev–Halley's method on Riemannian manifolds, J. Comput. Appl. Math. 336 (2018), 30–53. https://doi.org/10.1016/j.cam.2017.12.019.

[9] J.P. Dedieu, P. Priouret, G. Malajovich, Newton's method on Riemannian manifolds: Covariant alpha theory, IMA J. Numer. Anal. 23 (2003), 395–419. https://doi.org/10.1093/imanum/23.3.395.

[10] M.P. do Carmo, J.F. Francis, Riemannian geometry, Vol. 2, Birkhäuser (1992).

[11] A. Edelman, R. Lippert, Z. Bai, Nonlinear eigenvalue problems with orthogonality constraints, in: Templates for the Solution of Algebraic Eigenvalue Problems, 2000, 290–314.

[12] O.P. Ferreira, P.R. Oliveira, Proximal point algorithm on Riemannian manifolds, Optim. 51 (2002), 257–270. https://doi.org/10.1080/02331930290019413.

[13] O.P. Ferreira, B.F. Svaiter, Kantorovich's theorem on Newton's method in Riemannian manifolds, J. Complexity 18 (2002), 304–329. https://doi.org/10.1006/jcom.2001.0582.

[14] S. Helgason, Differential geometry, Lie groups, and symmetric spaces, Academic Press (1979).

[15] U. Helmke, K. Hüper, J.B. Moore, Quadratically convergent algorithms for optimal dextrous hand grasping, IEEE Trans. Robot. Autom. 18 (2002), 138–146. https://doi.org/10.1109/TRA.2002.999643.

[16] U. Helmke, J.B. Moore, Optimization and dynamical systems, Springer (2012).

[17] J. Hiriart-Urruty, C. Lemaréchal, Convex analysis and minimization algorithms, Springer (1993). https://doi.org/10.1007/978-3-662-02796-7.

[18] K. Jittorntrum, M.R. Osborne, Strong uniqueness and second order convergence in nonlinear discrete approximation, Numer. Math. 34 (1980), 439–455. https://doi.org/10.1007/BF01403680.

[19] J. Kou, Y. Li, Modified Chebyshev's method free from second derivative for non-linear equations, Appl. Math. Comput. 187 (2007), 1027–1032. https://doi.org/10.1016/j.amc.2006.09.021.

[20] C. Li, G. López, V. Martín-Márquez, Monotone vector fields and the proximal point algorithm on Hadamard manifolds, J. Lond. Math. Soc. 79 (2009), 663–683. https://doi.org/10.1112/jlms/jdn087.

[21] 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. https://doi.org/10.1137/06065622X.

[22] C. Li, J. Wang, Newton's method for sections on Riemannian manifolds: Generalized covariant α-theory, J. Complexity 24 (2008), 423–451. https://doi.org/10.1016/j.jco.2007.12.003.

[23] C. Li, J. Wang, Newton's method on Riemannian manifolds: Smale's point estimate theory under the γ-condition, IMA J. Numer. Anal. 26 (2006), 228–251. https://doi.org/10.1093/imanum/dri039.

[24] C. Li, X. Wang, On the convergence of the Gauss–Newton method for convex composite optimization, Math. Program. 91 (2002), 349–356. https://doi.org/10.1007/s101070100249.

[25] Y. Ma, J. Košecká, S. Sastry, Optimization criteria and geometric algorithms for motion and structure estimation, Int. J. Comput. Vision 44 (2001), 219–249. https://doi.org/10.1023/A:1012276232049.

[26] K. Madsen, Minimization of non-linear approximation functions, Ph.D. Thesis, Technical Univ. Denmark (1985).

[27] Y. Nishimori, S. Akaho, Learning algorithms utilizing quasi-geodesic flows on the Stiefel manifold, Neurocomputing 67 (2005), 106–135. https://doi.org/10.1016/j.neucom.2004.11.035.

[28] P.K. Parida, D.K. Gupta, Recurrence relations for a Newton-like method in Banach spaces, J. Comput. Appl. Math. 206 (2007), 873–887. https://doi.org/10.1016/j.cam.2006.08.027.

[29] T. Rapcsák, Smooth nonlinear optimization in $\mathbb{R}^n$, Springer (2013).

[30] T. Rapcsák, Sectional curvatures in nonlinear optimization, J. Global Optim. 40 (2008), 375–388. https://doi.org/10.1007/s10898-007-9212-7.

[31] S.M. Robinson, Normed convex processes, Trans. Amer. Math. Soc. 174 (1972), 127–140.

[32] S.M. Robinson, Extension of Newton's method to nonlinear functions with values in a cone, Numer. Math. 19 (1972), 341–347. https://doi.org/10.1007/BF01404880.

[33] R.T. Rockafellar, Convex analysis, Princeton Univ. Press (1970). https://doi.org/10.1515/9781400873173.

[34] S.T. Smith, Optimization techniques on Riemannian manifolds, arXiv:1407.5965 (2014). https://doi.org/10.48550/arXiv.1407.5965.

[35] C. Udrişte, Convex functions and optimization methods on Riemannian manifolds, Springer (2013). https://doi.org/10.1007/978-94-015-8390-9.

[36] J.H. Wang, J.C. Yao, C. Li, Gauss–Newton method for convex composite optimizations on Riemannian manifolds, J. Global Optim. 53 (2012), 5–28.

[37] R.S. Womersley, Local properties of algorithms for minimizing nonsmooth composite functions, Math. Program. 32 (1985), 69–89. https://doi.org/10.1007/BF01585659.

Appendix

The Python code used for the computation is given below.

```python
# %%
def compute_s_next(s_prev, s_curr, l0, l):
    # Compute s_{i+1} based on the given recurrence relation.
    print(f"s_curr - s_prev: {s_curr - s_prev}")
    numerator = l * (s_curr - s_prev) ** 2
    denominator = 2 * (1 - l0 * s_curr)
    s_next = s_curr + numerator / denominator
    return s_next

def compute_t_next(t_prev, t_curr, l1):
    # Compute t_{i+1} based on the given recurrence relation.
    print(f"t_curr - t_prev: {t_curr - t_prev}")
    numerator = l1 * (t_curr - t_prev) ** 2
    denominator = 2 * (1 - l1 * t_curr)
    t_next = t_curr + numerator / denominator
    return t_next

# Define the constants based on the given equations
def constants(d):
    # Compute l0, l, l1, and mu based on parameter d.

    l0 = 3 - d
    l = 2 * (1 + 1 / (3 - d))
    l1 = 2 * (2 - d)
    mu = (1 - d) / 3
    return l0, l, l1, mu

# Example usage
# Define the parameter d
d = 0.6

# Get constants
l0, l, l1, mu = constants(d)

# Initial values for s and t
s0, s1 = 0, mu
t0, t1 = 0, mu

# Iteratively compute s_{i+1} and t_{i+1}
iterations = 10
s_values = [s0, s1]
t_values = [t0, t1]

for i in range(2, iterations):
    s_next = compute_s_next(s_values[-2], s_values[-1], l0, l)
    t_next = compute_t_next(t_values[-2], t_values[-1], l1)
    s_values.append(s_next)
    t_values.append(t_next)

print(f"s values: {s_values}")
print(f"t values: {t_values}")

# Print differences s_i - s_{i-1} and t_i - t_{i-1}
for i in range(1, len(s_values)):
    print(f"s_{i} - s_{i-1}: {s_values[i] - s_values[i-1]}")

for i in range(1, len(t_values)):
    print(f"t_{i} - t_{i-1}: {t_values[i] - t_values[i-1]}")

# Compute s* and t* using the given expressions
s_star = (1 - (1 - 2 * l * mu) ** 0.5) / l
t_star = (1 - (1 - 2 * l1 * mu) ** 0.5) / l1

print(f"s* (s_star): {s_star}")
print(f"t* (t_star): {t_star}")
```