RESEARCH

Open Access

Convergence criteria of Newton's method on Lie groups

Jinsu He[1*], Jinhua Wang[2,3] and Jen-Chih Yao[4,5]

*Correspondence: hejinsu@zjnu.cn
[1]Department of Mathematics, Zhejiang Normal University, Jinhua, P.R. China
Full list of author information is available at the end of the article

Abstract

In the present paper, we study Newton's method on Lie groups (independent of affine connections) for finding zeros of a mapping f from a Lie group to its Lie algebra. Under a generalized L-average Lipschitz condition of the differential of f, we establish a unified convergence criterion of Newton's method. As applications, we get the convergence criteria under the Kantorovich's condition and the γ-condition, respectively. Moreover, applications to optimization problems are also provided.

MSC: 65H10; 65D99

Keywords: Newton's method; Lie group; L-average Lipschitz condition

1 Introduction

Newton's method is one of the most important methods for finding the approximation solution of the equation $f(x) = 0$, where f is an operator from some domain D in a real or complex Banach space X to another Y. As is well known, one of the most important results on Newton's method is Kantorovich's theorem (*cf.* [1]). Under the mild condition that the second Fréchet derivative of F is bounded (or more general, the first derivative is Lipschitz continuous) on a proper open metric ball of the initial point x_0, Kantorovich's theorem provides a simple and clear criterion ensuring the quadratic convergence of Newton's method. Another important result on Newton's method is Smale's point estimate theory (*i.e.*, α-theory and γ-theory) in [2], where the notions of approximate zeros were introduced and the rules to judge an initial point x_0 to be an approximate zero were established, depending on the information of the analytic nonlinear operator at this initial point and at a solution x^*, respectively. There are a lot of works on the weakness and/or the extension of the Lipschitz continuity made on the mappings; see, for example, [3–7] and references therein. In particular, Zabrejko-Nguen parametrized in [7] the classical Lipschitz continuity. Wang introduced in [6] the notion of Lipschitz conditions with L-average to unify both Kantorovich's and Smale's criteria.

In a Riemannian manifold framework, an analogue of the well-known Kantorovich's theorem was given in [8] for Newton's method for vector fields on Riemannian manifolds while the extensions of the famous Smale's α-theory and γ-theory in [2] to analytic vector fields and analytic mappings on Riemannian manifolds were done in [9]. In the recent paper [10], the convergence criteria in [9] were improved by using the notion of the γ-condition for the vector fields and mappings on Riemannian manifolds. The radii of uniqueness balls of singular points of vector fields satisfying the γ-conditions were estimated in [11], while the local behavior of Newton's method on Riemannian manifolds was

studied in [12, 13]. Furthermore, in [14], Li and Wang extended the generalized *L*-average Lipschitz condition (introduced in [6]) to Riemannian manifolds and established a unified convergence criterion of Newton's method on Riemannian manifolds. Similarly, inspired by previous work of Zabrejko and Nguen in [7] on Kantorovich's majorant method, Alvarez *et al.* introduced in [15] a Lipschitz-type radial function for the covariant derivative of vector fields and mappings on Riemannian manifolds and established a unified convergence criterion of Newton's method on Riemannian manifolds.

Note also that Mahony used one-parameter subgroups of a Lie group to develop a version of Newton's method on an arbitrary Lie group in [16], where the algorithm presented is independent of affine connections on the Lie group. This means that Newton's method on Lie groups is different from the one defined on Riemannian manifolds. On the other hand, motivated by looking for approaches to solving ordinary differential equations on Lie groups, Owren and Welfert also studied in [17] Newton's method, independent of affine connections on the Lie group, and showed the local quadratical convergence. Recently, Wang and Li [18] established Kantorovich's theorem (independent of the connection) for Newton's method on the Lie group. More precisely, under the assumption that the differential of *f* satisfies the Lipschitz condition around the initial point (which is in terms of one-parameter semigroups and independent of the metric), the convergence criterion of Newton's method is presented. Extensions of Smale's point estimate theory for Newton's method on Lie groups were given in [19].

The purpose of the present paper is to establish a unified convergence criterion for Newton's method (independent of the connection) on Lie groups under a generalized *L*-average Lipschitz condition. As applications, we get the convergence criteria under the Kantorovich's condition and the γ-condition, respectively. Hence, our results extend the corresponding results in [18] and [19], respectively. Moreover, applications to optimization problems are also provided.

The remainder of the paper is organized as follows. Some preliminary results and notions are given in Section 2, while the main results about a unified convergence criterion are presented in Section 3. In Section 4, applications to optimization problems are explored. Theorems under the Kantorovich's condition and the γ-condition are provided in the final section.

2 Notions and preliminaries

Most of the notions and notations which are used in the present paper are standard; see, for example, [20, 21]. The Lie group $(G, \cdot)$ is a Hausdorff topological group with countable bases which also has the structure of an analytic manifold such that the group product and the inversion are analytic operations in the differentiable structure given on the manifold. The dimension of a Lie group is that of the underlying manifold, and we shall always assume that it is *m*-dimensional. The symbol *e* designates the identity element of *G*. Let $\mathcal{G}$ be the Lie algebra of the Lie group *G* which is the tangent space T_eG of *G* at *e*, equipped with Lie bracket $[\cdot, \cdot] : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$.

In the sequel we make use of the left translation of the Lie group *G*. We define, for each $y \in G$, the left translation $L_y : G \to G$ by

$$L_y(z) = y \cdot z \quad \text{for each } z \in G. \tag{2.1}$$

The differential of L_y at z is denoted by $(L'_y)_z$, which clearly determines a linear isomorphism from $T_z G$ to the tangent space $T_{(y \cdot z)} G$. In particular, the differential $(L'_y)_e$ of L_y at e determines a linear isomorphism from $\mathcal{G}$ to the tangent space $T_y G$. The exponential map $\exp : \mathcal{G} \to G$ is certainly the most important construction associated to G and $\mathcal{G}$, and is defined as follows. Given $u \in \mathcal{G}$, let $\sigma_u : \mathbb{R} \to G$ be a one-parameter subgroup of G determined by the left invariant vector field $X_u : y \mapsto (L'_y)_e(u)$; *i.e.*, σ_u satisfies that

$$\sigma_u(0) = e \quad \text{and} \quad \sigma'_u(t) = X_u\big(\sigma_u(t)\big) = \big(L'_{\sigma_u(t)}\big)_e(u) \quad \text{for each } t \in \mathbb{R}. \tag{2.2}$$

The value of the exponential map $\exp$ at u is then defined by

$$\exp(u) = \sigma_u(1).$$

Moreover, we have that

$$\exp(tu) = \sigma_{tu}(1) = \sigma_u(t) \quad \text{for each } t \in \mathbb{R} \text{ and } u \in \mathcal{G} \tag{2.3}$$

and

$$\exp(t + s)u = \exp(tu) \cdot \exp(su) \quad \text{for any } t, s \in \mathbb{R} \text{ and } u \in \mathcal{G}. \tag{2.4}$$

Note that the exponential map is not surjective in general. However, the exponential map is a diffeomorphism on an open neighborhood of $0 \in \mathcal{G}$. In the case when G is Abelian, $\exp$ is also a homomorphism from $\mathcal{G}$ to G, *i.e.*,

$$\exp(u + v) = \exp(u) \cdot \exp(v) \quad \text{for all } u, v \in \mathcal{G}. \tag{2.5}$$

In the non-abelian case, $\exp$ is not a homomorphism and, by the Baker-Campbell-Hausdorff (BCH) formula (*cf.* [21, p.114]), (2.5) must be replaced by

$$\exp(w) = \exp(u) \cdot \exp(v) \tag{2.6}$$

for all u, v in a neighborhood of $0 \in \mathcal{G}$, where w is defined by

$$w := u + v + \frac{1}{2}[u, v] + \frac{1}{12}\big([u, [u, v]] + [v, [v, u]]\big) + \cdots. \tag{2.7}$$

Let $f : G \to \mathcal{G}$ be a C^1-map and let $x \in G$. We use f'_x to denote the differential of f at x. Then, by [22, p.9] (the proof given there for a smooth mapping still works for a C^1-map), for each $\triangle_x \in T_x G$ and any nontrivial smooth curve $c : (-\varepsilon, \varepsilon) \to G$ with $c(0) = x$ and $c'(0) = \triangle_x$, one has that

$$f'_x \triangle_x = \left(\frac{\mathrm{d}}{\mathrm{d}t}(f \circ c)(t)\right)_{t=0}. \tag{2.8}$$

In particular,

$$f'_x \triangle_x = \left(\frac{\mathrm{d}}{\mathrm{d}t}f\big(x \cdot \exp\big(t\big(L'_{x^{-1}}\big)_x \triangle_x\big)\big)\right)_{t=0} \quad \text{for each } \triangle_x \in T_x G. \tag{2.9}$$

Define the linear map $df_x : \mathcal{G} \to \mathcal{G}$ by

$$df_x u = \left(\frac{d}{dt} f\left(x \cdot \exp(tu)\right) \right)_{t=0} \quad \text{for each } u \in \mathcal{G}. \tag{2.10}$$

Then, by (2.9),

$$df_x = f'_x \circ \left(L'_x\right)_e. \tag{2.11}$$

Also, in view of the definition, we have that for all $t \geq 0$,

$$\frac{d}{dt} f\left(x \cdot \exp(tu)\right) = df_{x \cdot \exp(tu)} u \quad \text{for each } u \in \mathcal{G} \tag{2.12}$$

and

$$f\left(x \cdot \exp(tu)\right) - f(x) = \int_0^t df_{x \cdot \exp(su)} u \, ds \quad \text{for each } u \in \mathcal{G}. \tag{2.13}$$

For the remainder of the present paper, we always assume that $\langle \cdot, \cdot \rangle$ is an inner product on $\mathcal{G}$ and $\| \cdot \|$ is the associated norm on $\mathcal{G}$. We now introduce the following distance on G which plays a key role in the study. Let $x, y \in G$ and define

$$\varrho(x,y) := \inf \left\{ \sum_{i=1}^k \|u_i\| \, \middle| \, \begin{array}{l} \text{there exist } k \geq 1 \text{ and } u_1, \ldots, u_k \in \mathcal{G} \text{ such that} \\ y = x \cdot \exp u_1 \cdot \ldots \cdot \exp u_k \end{array} \right\}, \tag{2.14}$$

where we adapt the convention that $\inf \emptyset = +\infty$. It is easy to verify that $\varrho(\cdot, \cdot)$ is a distance on G and the topology induced by this distance is equivalent to the original one on G.

Let $x \in G$ and $r > 0$. We denote the corresponding ball of radius r around x of G by $C_r(x)$, that is,

$$C_r(x) := \left\{ y \in G \,|\, \varrho(x,y) < r \right\}.$$

Let $\mathcal{L}(\mathcal{G})$ denote the set of all linear operators on $\mathcal{G}$. Below, we will modify the notion of the Lipschitz condition with L-average for mappings on Banach spaces to suit sections. Let L be a positive nondecreasing integrable function on $[0, R]$, where R is a positive number large enough such that $\int_0^R (R - s)L(s) \, ds \geq R$. The notion of Lipschitz condition in the inscribed sphere with the L average for operators from Banach spaces to Banach spaces was first introduced in [23] by Wang for the study of Smale's point estimate theory.

Definition 2.1 Let $r > 0$, $x_0 \in G$, and let T be a mapping from G to $\mathcal{L}(\mathcal{G})$. Then T is said to satisfy the L-average Lipschitz condition on $C_r(x_0)$ if

$$\left\| T(x \cdot \exp u) - T(x) \right\| \leq \int_{\rho(x_0,x)}^{\rho(x_0,x)+\|u\|} L(s) \, ds \tag{2.15}$$

holds for any $u, u_0, \ldots, u_k \in \mathcal{G}$ and $x \in C_r(x_0)$ such that $x = x_0 \exp u_0 \exp u_1 \cdots \exp u_k$ and $\|u\| + \rho(x, x_0) < r$, where $\rho(x, x_0) := \sum_{i=0}^k \|u_i\|$.

The majorizing function h defined in the following, which was first introduced and studied by Wang (*cf.* [23]), is a powerful tool in our study. Let $r_0 > 0$ and $b > 0$ be such that

$$\int_0^{r_0} L(s)\,ds = 1 \quad \text{and} \quad b = \int_0^{r_0} L(s)s\,ds. \tag{2.16}$$

For $\beta > 0$, define the majorizing function h by

$$h(t) = \beta - t + \int_0^t L(s)(t - s)\,ds \quad \text{for each } 0 \le t \le R. \tag{2.17}$$

Some useful properties are described in the following propositions, see [23].

Proposition 2.1 *The function h is monotonic decreasing on $[0, r_0]$ and monotonic increasing on $[r_0, R]$. Moreover, if $\beta \le b$, h has a unique zero respectively in $[0, r_0]$ and $[r_0, R]$, which are denoted by r_1 and r_2.*

Let $\{t_n\}$ denote the sequence generated by Newton's method with initial value $t_0 = 0$ for h, that is,

$$t_{n+1} = t_n - h'(t_n)^{-1}h(t_n) \quad \text{for each } n = 0, 1, \ldots. \tag{2.18}$$

Proposition 2.2 *Suppose that $\beta \le b$. Then the sequence $\{t_n\}$ generated by (2.18) is monotonic increasing and convergent to r_1.*

The following lemma will be useful in the proof of the main theorem.

Lemma 2.1 *Let $0 < r \le r_0$ and let $x_0 \in G$ be such that $df_{x_0}^{-1}$ exists. Suppose that $df_{x_0}^{-1}\,df$ satisfies the L-average Lipschitz condition on $C_r(x_0)$. Let $x \in C_r(x_0)$ be such that there exist $k \ge 1$ and $u_0, \ldots, u_k \in \mathcal{G}$ satisfying $x = x_0 \cdot \exp u_0 \cdot \ldots \cdot \exp u_k$ and $\rho(x, x_0) := \sum_{i=0}^{k} \|u_i\| < r$. Then df_x^{-1} exists and*

$$\left\| df_x^{-1}\,df_{x_0} \right\| \le \frac{1}{1 - \int_0^{\rho(x,x_0)} L(s)\,ds}. \tag{2.19}$$

Proof Write $y_0 = x_0$ and $y_{i+1} = y_i \cdot \exp u_i$ for each $i = 0, \ldots, k$. Since (2.15) holds with $T = df_{x_0}^{-1}\,df$, one has that

$$\left\| df_{x_0}^{-1}(df_{y_i \cdot \exp u_i} - df_{y_i}) \right\| \le \int_{\rho(y_i, x_0)}^{\rho(y_{i+1}, x_0)} L(s)\,ds \quad \text{for each } 0 \le i \le k. \tag{2.20}$$

Noting that $y_{k+1} = x$, we have that

$$
\begin{aligned}
\left\| df_{x_0}^{-1}\,df_x - \mathbf{I}_{\mathcal{G}} \right\| &= \left\| df_{x_0}^{-1}(df_{y_k \cdot \exp u_k} - df_{x_0}) \right\| \\
&\le \sum_{i=0}^{k} \left\| df_{x_0}^{-1}(df_{y_i \cdot \exp u_i} - df_{y_i}) \right\| \\
&= \int_0^{\rho(x, x_0)} L(s)\,ds
\end{aligned}
$$

$$< \int_0^{r_0} L(s)\,ds$$

$$= 1.$$

Thus the conclusion follows from the Banach lemma and the proof is complete. □

3 Convergence criteria

Following [17], we define Newton's method with initial point x_0 for f on a Lie group as follows:

$$x_{n+1} = x_n \cdot \exp\left(-df_{x_n}^{-1}f(x_n)\right) \quad \text{for each } n = 0, 1, \ldots.$$ (3.1)

Recall that $f : G \to \mathcal{G}$ is a C^1-mapping. In the remainder of this section, we always assume that $x_0 \in G$ is such that $df_{x_0}^{-1}$ exists and set $\beta := \|df_{x_0}^{-1}f(x_0)\|$. Let r_0 and b given by (2.16), and r_1 be given by Proposition 2.1.

Theorem 3.1 *Suppose that* $df_{x_0}^{-1}\,df$ *satisfies the L-average Lipschitz condition on* $C_{r_1}(x_0)$ *and that*

$$\beta = \left\|df_{x_0}^{-1}f(x_0)\right\| \leq b.$$ (3.2)

Then the sequence $\{x_n\}$ *generated by Newton's method* (3.1) *with initial point* x_0 *is well defined and converges to a zero* x^* *of* f. *Moreover, the following assertions hold for each* $n = 0, 1, \ldots$:

$$\varrho(x_{n+1}, x_n) \leq \left\|df_{x_n}^{-1}f(x_n)\right\| \leq t_{n+1} - t_n;$$ (3.3)

$$\varrho(x_n, x^*) \leq r_1 - t_n.$$ (3.4)

Proof Write $v_n = -df_{x_n}^{-1}f(x_n)$ for each $n = 0, 1, \ldots$. Below we shall show that each v_n is well defined and

$$\varrho(x_{n+1}, x_n) \leq \|v_n\| \leq t_{n+1} - t_n$$ (3.5)

holds for each $n = 0, 1, \ldots$. Granting this, one sees that the sequence $\{x_n\}$ generated by Newton's method (3.1) with initial point x_0 is well defined and converges to a zero x^* of f, because, by (3.1),

$$x_{n+1} = x_n \cdot \exp v_n \quad \text{for each } n = 0, 1, \ldots.$$

Furthermore, assertions (3.3) and (3.4) hold for each n and the proof of the theorem is completed.

Note that v_0 is well defined by assumption and $x_1 = x_0 \cdot \exp v_0$. Hence, $\varrho(x_1, x_0) \leq \|v_0\|$. Since $\|v_0\| = \|-df_{x_0}^{-1}(f(x_0))\| = \beta = t_1 - t_0$, it follows that (3.5) is true for $n = 0$. We now proceed by mathematical induction on n. For this purpose, assume that v_n is well defined and (3.5) holds for each $n \leq k - 1$. Then

$$\sum_{i=0}^{k-1} \|v_i\| \leq t_k - t_0 = t_k < r_1 \quad \text{and} \quad x_k = x_0 \cdot \exp v_0 \cdot \ldots \cdot \exp v_{k-1}.$$ (3.6)

Thus, we use Lemma 2.1 to conclude that $df_{x_k}^{-1}$ exists and

$$\| df_{x_k}^{-1} \, df_{x_0} \| \leq \frac{1}{1 - \int_0^{t_k} L(s) \, ds} = -h'(t_k)^{-1}. \tag{3.7}$$

Therefore, v_k is well defined. Observe that

$$f(x_k) = f(x_k) - f(x_{k-1}) - df_{x_{k-1}} v_{k-1}$$

$$= \int_0^1 df_{x_{k-1} \cdot \exp(t v_{k-1})} v_{k-1} \, dt - df_{x_{k-1}} v_{k-1}$$

$$= \int_0^1 [df_{x_{k-1} \cdot \exp(t v_{k-1})} - df_{x_{k-1}}] v_{k-1} \, dt,$$

where the second equality is valid because of (2.13). Therefore, applying (2.15), one has that

$$\| df_{x_0}^{-1} f(x_k) \| \leq \int_0^1 \| df_{x_0}^{-1} [df_{x_{k-1} \cdot \exp(t v_{k-1})} - df_{x_{k-1}}] \| \| v_{k-1} \| \, dt$$

$$\leq \int_0^1 \int_{\rho(x_{k-1}, x_0)}^{\rho(x_{k-1}, x_0) + t \| v_{k-1} \|} L(s) \, ds \| v_{k-1} \| \, dt$$

$$\leq \int_0^1 \int_{t_{k-1}}^{t_{k-1} + t(t_k - t_{k-1})} L(s) \, ds (t_k - t_{k-1}) \, dt$$

$$= \int_{t_{k-1}}^{t_k} L(s)(t_k - s) \, ds$$

$$= h(t_k) - h(t_{k-1}) - h'(t_{k-1})(t_k - t_{k-1})$$

$$= h(t_k), \tag{3.8}$$

where the first equality holds because $h(t_{k-1}) + h'(t_{k-1})(t_k - t_{k-1}) = 0$. Combining this with (3.7) yields that

$$\| v_k \| = \| -df_{x_k}^{-1} f(x_k) \|$$

$$\leq \| df_{x_k}^{-1} \, df_{x_0} \| \| df_{x_0}^{-1} f(x_k) \|$$

$$\leq -h'(t_k)^{-1} h(t_k)$$

$$= t_{k+1} - t_k. \tag{3.9}$$

Since $x_{k+1} = x_k \cdot \exp v_k$, we have $\varrho(x_{k+1}, x_k) \leq \| v_k \|$. This together with (3.9) gives that (3.5) holds for $n = k$, which completes the proof of the theorem. $\square$

4 Applications to optimization problems

Let $\phi : G \to \mathbb{R}$ be a C^2-map. Consider the following optimization problem:

$$\min_{x \in G} \phi(x). \tag{4.1}$$

Newton's method for solving (4.1) was presented in [16], where local quadratical convergence result was established for a smooth function ϕ.

Let $X \in \mathcal{G}$. Following [16], we use $\widetilde{X}$ to denote the left invariant vector field associated with X defined by

$$\widetilde{X}(x) = \left(L'_x\right)_e X \quad \text{for each } x \in G,$$

and $\widetilde{X}\phi$ the Lie derivative of ϕ with respect to the left invariant vector field $\widetilde{X}$, that is, for each $x \in G$,

$$(\widetilde{X}\phi)(x) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \phi(x \cdot \exp tX). \tag{4.2}$$

Let $\{X_1, \ldots, X_n\}$ be an orthonormal basis of $\mathcal{G}$. According to [24, p.356] (see also [16]), $\operatorname{grad}\phi$ is a vector field on G defined by

$$\operatorname{grad}\phi(x) = \left(\widetilde{X}_1, \ldots, \widetilde{X}_n\right)\left(\widetilde{X}_1\phi(x), \ldots, \widetilde{X}_n\phi(x)\right)^T = \sum_{j=1}^n \widetilde{X}_j\phi(x)\widetilde{X}_j \quad \text{for each } x \in G. \tag{4.3}$$

Then Newton's method with initial point $x_0 \in G$ considered in [16] can be written in a coordinate-free form as follows.

Algorithm 4.1 Find $X^k \in \mathcal{G}$ such that $\widetilde{X}^k = (L'_x)_e X^k$ and

$$\operatorname{grad}\phi(x_k) + \operatorname{grad}\left(\widetilde{X}^k\phi\right)(x_k) = 0;$$

Set $x_{k+1} = x_k \cdot \exp X^k$;
Set $k \leftarrow k + 1$ and repeat.

Let $f : G \to \mathcal{G}$ be a mapping defined by

$$f(x) = \left(L'_x\right)_e^{-1} \operatorname{grad}\phi(x) \quad \text{for each } x \in G. \tag{4.4}$$

Define the linear operator $H_x\phi : \mathcal{G} \to \mathcal{G}$ for each $x \in G$ by

$$(H_x\phi)X = \left(L'_x\right)_e^{-1} \operatorname{grad}(\widetilde{X}\phi)(x) \quad \text{for each } X \in \mathcal{G}. \tag{4.5}$$

Then $H_{(\cdot)}\phi$ defines a mapping from G to $\mathcal{L}(\mathcal{G})$. The following proposition gives the equivalence between $\mathrm{d}f_x$ and $H_x\phi$. The following proposition was given in [18].

Proposition 4.1 *Let $f(\cdot)$ and $H_{(\cdot)}\phi$ be defined respectively by* (4.4) *and* (4.5). *Then*

$$\mathrm{d}f_x = H_x\phi \quad \text{for each } x \in G. \tag{4.6}$$

Remark 4.1 One can easily see from Proposition 4.1 that, with the same initial point, the sequence generated by Algorithm 4.1 for ϕ coincides with the one generated by Newton's method (3.1) for f defined by (4.4).

Let $x_0 \in G$ be such that $(H_{x_0}\phi)^{-1}$ exists, and let $\beta_\phi := \|(H_{x_0}\phi)^{-1}(L'_{x_0})_e^{-1} \operatorname{grad}\phi(x_0)\|$. Recall that r_0 and b are given by (2.16), and r_1 is given by Proposition 2.1. Then the main theorem of this section is as follows.

Theorem 4.1 *Suppose that*

$$\beta_\phi = \left\| (H_{x_0}\phi)^{-1} \left(L'_{x_0} \right)_e^{-1} \operatorname{grad}\phi(x_0) \right\| \le b, \tag{4.7}$$

and that $(H_{x_0}\phi)^{-1}(H_{(\cdot)}\phi)$ *satisfies the L-average Lipschitz condition on* $C_{r_1}(x_0)$. *Then the sequence generated by Algorithm* 4.1 *with initial point* x_0 *is well defined and converges to a critical point* x^* *of* ϕ: $\operatorname{grad}\phi(x^*) = 0$.

Furthermore, if $H_{x_0}\phi$ *is additionally positive definite and the following Lipschitz condition is satisfied*:

$$\left\| (H_{x_0}\phi)^{-1} \right\| \|H_{x\cdot\exp u}\phi - H_x\phi\|$$
$$\le \int_{\rho(x,x_0)}^{\rho(x,x_0)+\|u\|} L(s)\,ds \quad \text{for } x \in G \text{ and } u \in \mathcal{G} \text{ with } \rho(x_0,x) + \|u\| < r_1. \tag{4.8}$$

Then x^* *is a local solution of* (4.1).

Proof Recall that f is defined by (4.4). Then by Proposition 4.1, $df_x = H_x\phi$ for each $x \in G$. Hence, by assumptions, $df_{x_0}^{-1} df$ satisfies the L-average Lipschitz condition on $C_{r_1}(x_0)$ and condition (3.2) is satisfied because $\beta_\phi \le b$. Thus, Theorem 3.1 is applicable; hence the sequence generated by Newton's method for f with initial point x_0 is well defined and converges to a zero x^* of f. Consequently, by Remark 4.1, one sees that the first assertion holds.

To prove the second assertion, we assume that $H_{x_0}\phi$ is additionally positive definite and the Lipschitz condition (4.8) is satisfied. It is sufficient to prove that $H_{x^*}\phi$ is positive definite. Let λ^* and λ^0 be the minimum eigenvalues of $H_{x^*}\phi$ and $H_{x_0}\phi$, respectively. Then $\lambda^0 > 0$. We have to show that $\lambda^* > 0$. To do this, let $\{x_n\}$ be the sequence generated by Algorithm 4.1 and write $v_n = df_{x_n}^{-1}f(x_n)$ for each $n = 0, 1, \dots$. Then, by Remark 4.1,

$$x_{n+1} = x_n \cdot \exp(-v_n) \quad \text{for each } n = 0, 1, \dots, \tag{4.9}$$

and by Theorem 3.1,

$$\|v_n\| \le t_{n+1} - t_n \quad \text{for each } n = 0, 1, \dots. \tag{4.10}$$

Therefore, for each $n = 0, 1, \dots$,

$$\left\| H_{x_0}\phi^{-1} \right\| \left\| (H_{x_{n+1}}\phi - H_{x_0}\phi) \right\| = \left\| H_{x_0}\phi^{-1} \right\| \left\| (H_{x_n\cdot\exp(-v_n)}\phi - H_{x_0}\phi) \right\|$$
$$= \sum_{j=0}^{n} \left\| H_{x_0}\phi^{-1} \right\| \|H_{x_j\cdot\exp(-v_n)}\phi - H_{x_j}\phi\|$$
$$\le \sum_{j=0}^{n} \int_{\rho(x_j,x_0)}^{\rho(x_j,x_0)+\|v_n\|} L(s)\,ds$$
$$\le \int_0^{t_k} L(s)\,ds$$
$$< \int_0^{r_0} L(s)\,ds$$
$$= 1 \tag{4.11}$$

thanks to (4.8)-(4.10). Since

$$\left| \frac{\lambda^*}{\lambda^0} - 1 \right| = \frac{1}{\lambda^0} \left| \min_{v \in \mathcal{G}, \|v\|=1} \langle (H_{x^*}\phi)v, v \rangle - \min_{v \in \mathcal{G}, \|v\|=1} \langle (H_{x_0}\phi)v, v \rangle \right|$$

$$\leq \left\| H_{x_0}\phi^{-1} \right\| \|H_{x^*}\phi - H_{x_0}\phi\|,$$

it follows that

$$\left| \frac{\lambda^*}{\lambda^0} - 1 \right| \leq \lim_{n \to \infty} \left\| H_{x_0}\phi^{-1} \right\| \|H_{x_{n+1}}\phi - H_{x_0}\phi\| < 1$$

thanks to (4.11). This implies that $\lambda^* > 0$ and completes the proof. $\qquad\square$

5 Theorems under the Kantorovich's condition and the γ-condition

If $L(\cdot)$ is a constant, then the L-average Lipschitz condition is reduced to the classical Lipschitz condition.

Let $r > 0$, $x_0 \in G$, and let T be a mapping from G to $\mathcal{L}(\mathcal{G})$. Then T is said to satisfy the L Lipschitz condition on $C_r(x_0)$ if

$$\left\| T(x \cdot \exp u) - T(x) \right\| \leq L \|u\|$$

holds for any $u, u_0, \ldots, u_k \in \mathcal{G}$ and $x \in C_r(x_0)$ such that $x = x_0 \exp u_0 \exp u_1 \cdots \exp u_k$ and $\|u\| + \rho(x, x_0) < r$, where $\rho(x, x_0) = \sum_{i=0}^{k} \|u_i\|$.

Let $\beta > 0$ and $L > 0$. The quadratic majorizing function h is reduced to

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

Let $\{t_n\}$ denote the sequence generated by Newton's method with initial value $t_0 = 0$ for h, that is,

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

Assume that $\lambda := L\beta \leq \frac{1}{2}$. Then h has two zeros r_1 and r_2:

$$r_1 = \frac{1 - \sqrt{1 - 2\lambda}}{L} \quad \text{and} \quad r_2 = \frac{1 + \sqrt{1 - 2\lambda}}{L}; \tag{5.1}$$

moreover, $\{t_n\}$ is monotonic increasing and convergent to r_1, and satisfies that

$$r_1 - t_n = \frac{q^{2^n - 1}}{\sum_{j=0}^{2^n - 1} q^j} r_1 \quad \text{for each } n = 0, 1, \ldots,$$

where

$$q = \frac{1 - \sqrt{1 - 2\lambda}}{1 + \sqrt{1 - 2\lambda}}.$$

Recall that $f : G \to \mathcal{G}$ is a C^1-mapping. As in the previous section, we always assume that $x_0 \in G$ is such that $\mathrm{d}f_{x_0}^{-1}$ exists and set $\beta := \|\mathrm{d}f_{x_0}^{-1}f(x_0)\|$. Then, by Theorem 3.1, we obtain the following results, which were given in [18].

Theorem 5.1 *Suppose that* $\mathrm{d}f_{x_0}^{-1}\mathrm{d}f$ *satisfies the L-Lipschitz condition on* $C_{r_1}(x_0)$ *and that* $\lambda = L\beta \leq \frac{1}{2}$. *Then the sequence* $\{x_n\}$ *generated by Newton's method* (3.1) *with initial point* x_0 *is well defined and converges to a zero* x^* *of* f. *Moreover, the following assertions hold for each* $n = 0, 1, \ldots$:

$$\varrho(x_{n+1}, x_n) \leq \left\| \mathrm{d}f_{x_n}^{-1}f(x_n) \right\| \leq t_{n+1} - t_n;$$

$$\varrho(x_n, x^*) \leq \frac{q^{2^n - 1}}{\sum_{j=0}^{2^n - 1} q^j} r_1.$$

Let $x_0 \in G$ be such that $(H_{x_0}\phi)^{-1}$ exists, and let $\beta_\phi = \|(H_{x_0}\phi)^{-1}(L'_{x_0})_e^{-1} \operatorname{grad} \phi(x_0)\|$. Recall that r_1 is defined by (5.1). Then, by Theorem 4.1, we get the following results, which were given in [18].

Theorem 5.2 *Suppose that* $\lambda = L\beta_\phi \leq \frac{1}{2}$, *and that* $(H_{x_0}\phi)^{-1}(H_{(\cdot)}\phi)$ *satisfies the L-Lipschitz condition on* $C_{r_1}(x_0)$. *Then the sequence generated by Algorithm* 4.1 *with initial point* x_0 *is well defined and converges to a critical point* x^* *of* ϕ: $\operatorname{grad}\phi(x^*) = 0$.

Furthermore, if $H_{x_0}\phi$ *is additionally positive definite and the following Lipschitz condition is satisfied:*

$$\left\| (H_{x_0}\phi)^{-1} \right\| \|H_{x\cdot\exp u}\phi - H_x\phi\| \leq L\|u\| \quad \text{for } x \in G \text{ and } u \in \mathcal{G} \text{ with } \varrho(x_0, x) + \|u\| < r_1.$$

Then x^* *is a local solution of* (4.1).

Let k be a positive integer and assume further that $f : G \to \mathcal{G}$ is a C^k-map. Define the map $\mathrm{d}^k f_x : \mathcal{G}^k \to \mathcal{G}$ by

$$\mathrm{d}^k f_x u_1 \cdots u_k = \left(\frac{\partial^k}{\partial t_k \cdots \partial t_1} f(x \cdot \exp t_k u_k \cdots \exp t_1 u_1) \right)_{t_k = \cdots = t_1 = 0}$$

for each $(u_1, \ldots, u_k) \in \mathcal{G}^k$. In particular,

$$\mathrm{d}^k f_x u^k = \left(\frac{\mathrm{d}^k}{\mathrm{d}t^k} f(x \cdot \exp tu) \right)_{t=0} \quad \text{for each } u \in \mathcal{G}.$$

Let $1 \leq i \leq k$. Then, in view of the definition, one has that

$$\mathrm{d}^k f_x u_1 \cdots u_k = \mathrm{d}^{k-i} \left(\mathrm{d}^i f_\cdot (u_1 \cdots u_i) \right)_x u_{i+1} \cdots u_k \quad \text{for each } (u_1, \ldots, u_k) \in \mathcal{G}^k.$$

In particular, for fixed $u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_k \in \mathcal{G}$,

$$\mathrm{d}^i f_x u_1 \cdots u_{i-1} = \mathrm{d} \left(\mathrm{d}^{i-1} f_\cdot (u_1 \cdots u_{i-1}) \right)_x (\cdot).$$

This implies that $\mathrm{d}^i f_x u_1 \cdots u_{i-1} u$ is linear with respect to $u \in \mathcal{G}$ and so is $\mathrm{d}^k f_x u_1 \cdots u_{i-1} \times uu_{i+1} \cdots u_k$. Consequently, $\mathrm{d}^k f_x$ is a multilinear map from $\mathcal{G}^k$ to $\mathcal{G}$ because $1 \leq i \leq k$ is arbitrary. Thus we can define the norm of $\mathrm{d}^k f_x$ by

$$\left\| \mathrm{d}^k f_x \right\| = \sup \left\{ \left\| \mathrm{d}^k f_x u_1 u_2 \cdots u_k \right\| : (u_1, \ldots, u_k) \in \mathcal{G}^k \text{ with each } \|u_j\| = 1 \right\}.$$

For the remainder of the paper, we always assume that f is a C^2-map from G to $\mathcal{G}$. Then, taking $i = 2$, we have

$$\mathrm{d}^2 f_z v u = \mathrm{d}(\mathrm{d} f.v)_z u \quad \text{for any } u, v \in \mathcal{G} \text{ and each } z \in G.$$

Thus, (2.13) is applied (with $\mathrm{d} f.v$ in place of $f(\cdot)$ for each $v \in \mathcal{G}$) to conclude the following formula:

$$\mathrm{d} f_{x \cdot \exp(tu)} - \mathrm{d} f_x = \int_0^t \mathrm{d}^2 f_{x \cdot \exp(su)} u \, \mathrm{d}s \quad \text{for each } u \in \mathcal{G} \text{ and } t \in \mathbb{R}. \tag{5.2}$$

The γ-conditions for nonlinear operators in Banach spaces were first introduced and explored by Wang [25, 26] to study Smale's point estimate theory, which was extended in [19] for a map f from a Lie group to its Lie algebra in view of the map $\mathrm{d}^2 f$ as given in Definition 5.1 below. Let $r > 0$ and $\gamma > 0$ be such that $\gamma r \leq 1$.

Definition 5.1 Let $x_0 \in G$ be such that $\mathrm{d} f_{x_0}^{-1}$ exists. f is said to satisfy the γ-condition at x_0 on $C_r(x_0)$ if, for any $x \in C_r(x_0)$ with $x = x_0 \exp u_0 \exp u_1 \cdots \exp u_k$ such that $\rho(x, x_0) := \sum_{i=0}^k \|u_i\| < r$,

$$\left\| \mathrm{d} f_{x_0}^{-1} \mathrm{d}^2 f_x \right\| \leq \frac{2\gamma}{(1 - \gamma \rho(x, x_0))^3}.$$

As shown in Proposition 5.3, if f is analytic at x_0, then f satisfies the γ-condition at x_0. Let $\gamma > 0$ and let L be the function defined by

$$L(s) = \frac{2\gamma}{(1 - \gamma s)^3} \quad \text{for each } 0 < s < \frac{1}{\gamma}. \tag{5.3}$$

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

Proposition 5.1 *Suppose that f satisfies the γ-condition at x_0 on $C_r(x_0)$. Then $\mathrm{d} f_{x_0}^{-1} \mathrm{d} f$ satisfies the L-average Lipschitz condition on $C_r(x_0)$ with L defined by (5.3).*

Proof Let $x \in C_r(x_0)$ and let $u, u_0, \dots, u_k \in \mathcal{G}$ be such that $x = x_0 \exp u_0 \exp u_1 \cdots \exp u_k$ and $\sum_{i=0}^k \|u_i\| + \|u\| < r$. Write $\rho(x, x_0) := \sum_{i=0}^k \|u_i\|$. Observe from (5.2) that

$$\mathrm{d} f_{x \cdot \exp u} - \mathrm{d} f_x = \int_0^1 \mathrm{d}^2 f_{x \cdot \exp(su)} u \, \mathrm{d}s.$$

Combining this with the assumption yields that

$$\left\| \mathrm{d} f_{x_0}^{-1} (\mathrm{d} f_{x \cdot \exp u} - \mathrm{d} f_x) \right\| \leq \int_0^1 \left\| \mathrm{d} f_{x_0}^{-1} \mathrm{d}^2 f_{x \cdot \exp(su)} \right\| \|u\| \, \mathrm{d}s$$

$$\leq \int_0^1 \frac{2\gamma}{(1 - \gamma(\rho(x, x_0) + s\|u\|))^3} \|u\| \, \mathrm{d}s$$

$$= \int_{\rho(x, x_0)}^{\rho(x, x_0) + \|u\|} \frac{2\gamma}{(1 - \gamma t)^3} \, \mathrm{d}t.$$

Hence, $\mathrm{d}f_{x_0}^{-1}\,\mathrm{d}f$ satisfies the L-average Lipschitz condition on $C_r(x_0)$ with L defined by
(5.3). $\qquad\qquad\square$

Corresponding to the function L defined by (5.3), r_0 and b in (2.16) are $r_0 = (1 - \frac{\sqrt{2}}{2})\frac{1}{\gamma}$
and $b = (3 - 2\sqrt{2})\frac{1}{\gamma}$, and the majorizing function given in (2.17) reduces to

$$h(t) = \beta - t + \frac{\gamma t^2}{1 - \gamma t} \quad \text{for each } 0 \le t \le R.$$

Hence the condition $\beta \le b$ is equivalent to $\alpha = \gamma\beta \le 3 - 2\sqrt{2}$. Let $\{t_n\}$ denote the se-
quence generated by Newton's method with the initial value $t_0 = 0$ for h. Then the follow-
ing proposition was proved in [27], see also [10] and [6].

Proposition 5.2 *Assume that $\alpha = \gamma\beta \le 3 - 2\sqrt{2}$. Then the zeros of h are*

$$r_1 = \frac{1 + \alpha - \sqrt{(1+\alpha)^2 - 8\alpha}}{4\gamma}, \qquad r_2 = \frac{1 + \alpha + \sqrt{(1+\alpha)^2 - 8\alpha}}{4\gamma}$$

and

$$\beta \le r_1 \le \left(1 + \frac{1}{\sqrt{2}}\right)\beta \le \left(1 - \frac{1}{\sqrt{2}}\right)\frac{1}{\gamma} \le r_2 \le \frac{1}{2\gamma}.$$

Moreover, the following assertions hold:

$$t_{n+1} - t_n = \frac{(1 - \mu^{2^n})\sqrt{(1+\alpha)^2 - 8\alpha}}{2\alpha(1 - \nu\mu^{2^n-1})(1 - \nu\mu^{2^{n+1}-1})}\nu\mu^{2^n-1}\beta \le \mu^{2^n-1}\beta \quad \text{for each } n = 0, 1, \ldots,$$

where

$$\mu = \frac{1 - \alpha - \sqrt{(1+\alpha)^2 - 8\alpha}}{1 - \alpha + \sqrt{(1+\alpha)^2 - 8\alpha}} \quad \text{and} \quad \nu = \frac{1 + \alpha - \sqrt{(1+\alpha)^2 - 8\alpha}}{1 + \alpha + \sqrt{(1+\alpha)^2 - 8\alpha}}. \tag{5.4}$$

Recall that $x_0 \in G$ is such that $\mathrm{d}f_{x_0}^{-1}$ exists, and let $\beta := \|\mathrm{d}f_{x_0}^{-1}f(x_0)\|$. Then, by Theorem 3.1
and Proposition 5.2, we get the following results, which were given in [19].

Theorem 5.3 *Suppose that*

$$\alpha := \beta\gamma \le 3 - 2\sqrt{2}$$

*and that f satisfies the γ-condition at x_0 on $C_{r_1}(x_0)$. Then Newton's method (3.1) with ini-
tial point x_0 is well defined, and the generated sequence $\{x_n\}$ converges to a zero x^* of f.
Moreover, if $\alpha < 3 - 2\sqrt{2}$, then for each $n = 0, 1, \ldots,$*

$$\varrho(x_{n+1}, x_n) \le \nu^{2^n-1}\beta,$$

where ν is given by (5.4).

Below, we always assume that f is analytic on G. For $x \in G$ such that $\mathrm{d}f_x^{-1}$ exists, we define

$$\gamma_x := \gamma(f, x) = \sup_{i \geq 2} \left\| \frac{\mathrm{d}f_x^{-1} \mathrm{d}^i f_x}{i!} \right\|^{\frac{1}{i-1}}.$$

Also, we adopt the convention that $\gamma(f, x) = \infty$ if $\mathrm{d}f_x$ is not invertible. Note that this definition is justified and, in the case when $\mathrm{d}f_x$ is invertible, $\gamma(f, x)$ is finite by analyticity.

The following proposition is taken from [19].

Proposition 5.3 *Let $\gamma_{x_0} := \gamma(f, x_0)$ and let $r = \frac{2 - \sqrt{2}}{2\gamma_{x_0}}$. Then f satisfies the γ_{x_0}-condition at x_0 on $C_r(x_0)$.*

Thus, by Theorem 5.3 and Proposition 5.3, we get the following corollary, which was given in [19].

Corollary 5.1 *Suppose that*

$$\alpha := \beta \gamma_{x_0} \leq 3 - 2\sqrt{2}.$$

Then Newton's method (3.1) with initial point x_0 is well defined and the generated sequence $\{x_n\}$ converges to a zero x^* of f. Moreover, if $\alpha < 3 - 2\sqrt{2}$, then for each $n = 0, 1, \ldots$,

$$\varrho(x_{n+1}, x_n) \leq \nu^{2^n - 1} \beta,$$

where ν is given by (5.4).

Authors' contributions
All authors participated in its construction and drafted the manuscript. All authors read and approve the final manuscript.

Author details
[1] Department of Mathematics, Zhejiang Normal University, Jinhua, P.R. China. [2] Department of Mathematics, Zhejiang University of Technology, Hangzhou, 310032, P.R. China. [3] Department of Mathematics, National Sun Yat-sen University, Kaohsiung, Taiwan. [4] Center for Fundamental Science, Kaohsiung Medical University, Kaohsiung, 80702, Taiwan. [5] Department of Mathematics, King Abdulaziz University, P.O. Box 80203, Jeddah, 21589, Saudi Arabia.

References
1. Kantorovich, LV, Akilov, GP: Functional Analysis. Pergamon, Oxford (1982)
2. Smale, S: Newton's method estimates from data at one point. In: Ewing, R, Gross, K, Martin, C (eds.) The Merging of Disciplines: New Directions in Pure, Applied and Computational Mathematics, pp. 185-196. Springer, New York (1986)
3. Ezquerro, JA, Hernández, MA: Generalized differentiability conditions for Newton's method. IMA J. Numer. Anal. **22**, 187-205 (2002)
4. Ezquerro, JA, Hernández, MA: On an application of Newton's method to nonlinear operators with w-conditioned second derivative. BIT Numer. Math. **42**, 519-530 (2002)
5. Gutiérrez, JM, Hernández, MA: Newton's method under weak Kantorovich conditions. IMA J. Numer. Anal. **20**, 521-532 (2000)

6. Wang, XH: Convergence of Newton's method and uniqueness of the solution of equations in Banach space. IMA J. Numer. Anal. **20**(1), 123-134 (2000)
7. Zabrejko, PP, Nguen, DF: The majorant method in the theory of Newton-Kantorovich approximates and the Ptak error estimates. Numer. Funct. Anal. Optim. **9**, 671-674 (1987)
8. Ferreira, OP, Svaiter, BF: Kantorovich's theorem on Newton's method in Riemannian manifolds. J. Complex. **18**, 304-329 (2002)
9. Dedieu, JP, Priouret, P, Malajovich, G: Newton's method on Riemannian manifolds: covariant alpha theory. IMA J. Numer. Anal. **23**, 395-419 (2003)
10. Li, C, Wang, JH: Newton's method on Riemannian manifolds: Smale's point estimate theory under the γ-condition. IMA J. Numer. Anal. **26**, 228-251 (2006)
11. Wang, JH, Li, C: Uniqueness of the singular points of vector fields on Riemannian manifolds under the γ-condition. J. Complex. **22**, 533-548 (2006)
12. Li, C, Wang, JH: Convergence of Newton's method and uniqueness of zeros of vector fields on Riemannian manifolds. Sci. China Ser. A **48**, 1465-1478 (2005)
13. Wang, JH: Convergence of Newton's method for sections on Riemannian manifolds. J. Optim. Theory Appl. **148**, 125-145 (2011)
14. Li, C, Wang, JH: Newton's method for sections on Riemannian manifolds: generalized covariant α-theory. J. Complex. **24**, 423-451 (2008)
15. Alvarez, F, Bolte, J, Munier, J: A unifying local convergence result for Newton's method in Riemannian manifolds. Found. Comput. Math. **8**, 197-226 (2008)
16. Mahony, RE: The constrained Newton method on a Lie group and the symmetric eigenvalue problem. Linear Algebra Appl. **248**, 67-89 (1996)
17. Owren, B, Welfert, B: The Newton iteration on Lie groups. BIT Numer. Math. **40**(2), 121-145 (2000)
18. Wang, JH, Li, C: Kantorovich's theorems for Newton's method for mappings and optimization problems on Lie groups. IMA J. Numer. Anal. **31**, 322-347 (2011)
19. Li, C, Wang, JH, Dedieu, JP: Smale's point estimate theory for Newton's method on Lie groups. J. Complex. **25**, 128-151 (2009)
20. Helgason, S: Differential Geometry, Lie Groups, and Symmetric Spaces. Academic Press, New York (1978)
21. Varadarajan, VS: Lie Groups, Lie Algebras and Their Representations. Graduate Texts in Mathematics, vol. 102. Springer, New York (1984)
22. DoCarmo, MP: Riemannian Geometry. Birkhäuser Boston, Cambridge (1992)
23. Wang, XH: Convergence of Newton's method and inverse function theorem in Banach spaces. Math. Comput. **68**, 169-186 (1999)
24. Helmke, U, Moore, JB: Optimization and Dynamical Systems. Commun. Control Eng. Ser. Springer, London (1994)
25. Wang, XH, Han, DF: Criterion α and Newton's method in weak condition. Chin. J. Numer. Math. Appl. **19**, 96-105 (1997)
26. Wang, XH: Convergence on the iteration of Halley family in weak conditions. Chin. Sci. Bull. **42**, 552-555 (1997)
27. Wang, XH, Han, DF: On the dominating sequence method in the point estimates and Smale's theorem. Sci. Sin., Ser. A, Math. Phys. Astron. Tech. Sci. **33**, 135-144 (1990)