Journal of Zhejiang University SCIENCE A
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Kantorovich's theorem for Newton's method on Lie groups*

WANG Jin-hua[1], LI Chong[2]
([1]Department of Mathematics, Zhejiang University of Technology, Hangzhou 310032, China)
([2]Department of Mathematics, Zhejiang University, Hangzhou 310027, China)
E-mail: wjh@zjut.edu.cn; cli@zju.edu.cn
Received Sept. 20, 2006; revision accepted Jan. 4, 2007

Abstract: The convergence criterion of Newton's method to find the zeros of a map f from a Lie group to its corresponding Lie algebra is established under the assumption that f satisfies the classical Lipschitz condition, and that the radius of convergence ball is also obtained. Furthermore, the radii of the uniqueness balls of the zeros of f are estimated. Owren and Welfert (2000) stated that if the initial point is close sufficiently to a zero of f, then Newton's method on Lie group converges to the zero; while this paper provides a Kantorovich's criterion for the convergence of Newton's method, not requiring the existence of a zero as a priori.

Key words: Newton's method, Lie group, Kantorovich's theorem, Lipschitz condition
doi:10.1631/jzus.2007.A0978 **Document code:** A **CLC number:** O242.23

INTRODUCTION

There is increased interest in numerical algorithms on manifolds as there are many numerical problems posed in manifolds arising in many natural contexts. Classical examples are given by eigenvalue problems, symmetric eigenvalue problems, invariant subspace computations, minimization problems with orthogonality constraints, optimization problems with equality constraints, etc. See for example (Gabay, 1982; Smith, 1993; 1994; Udriste, 1994; Edelman *et al.*, 1998; Adler *et al.*, 2002). For such problems, one often has to compute solutions of a system of equations or to find zeros of a vector field on a Riemannian manifold.

One of the most famous methods to approximately solve these problems is Newton's method. An analogue of the well-known Kantorovich theorem (Kantorovich, 1948; Kantorovich and Akilov, 1982) was given in (Ferreira and Svaiter, 2002) for Newton's method on Riemannian manifolds, while the extensions of the famous Smale's α-theory and γ-theory in (Smale, 1986) to analytic vector fields on Riemannian manifolds were made in (Dedieu *et al.*, 2003). To extend and improve the Smale's γ-theory and α-theory of Newton's method for operators in Banach spaces, Wang (1997) and Wang and Han (1997) proposed the notion of "γ-condition", which is weaker than the Smale assumption in (Smale, 1986) for analytic operators. In the recent paper (Li and Wang, 2006), we extended the notion of γ-condition to vector fields on Riemannian manifolds and then established the γ-theory and α-theory of Newton's method for the vector fields on Riemannian manifolds satisfying the γ-condition, which consequently improve the results in (Dedieu *et al.*, 2003). The radii of uniqueness balls of zeroes of vector fields satisfying the γ-conditions were studied in (Wang and Li, 2006), while the local behavior of Newton's method on Riemannian manifolds were studied in (Li and Wang, 2005), where we estimated the radii of convergence balls of Newton's method and uniqueness balls of zeros of vector fields on Riemannian manifolds under the assumption that the covariant derivatives of the vector fields satisfy some kind of general Lipschitz condition.

* Project supported by the National Natural Science Foundation of China (No. 10271025), and the Program for New Century Excellent Talents in University of China

On the other hand, the numerical problems such as minimization problems with orthogonality constraints, optimization problems with equality constraints, etc. can be actually considered as problems on Lie groups (Smith, 1993; 1994; Mahony, 1996; Adler *et al.*, 2002). As is well known, Lie group is a Hausdorff topological group, which has the structure of a smooth manifold such that the group product and the inversion are smooth operations in the differentiable structure given on the manifold.

Mahony (1996) used one-parameter subgroups of the Lie group to develop a version of Newton's method on an arbitrary Lie group, where the approach for solving eigenvalue problems as a constrained optimization problem on a Lie group via Newton's method has been explored and the local convergence was analyzed. In this manner the algorithm presented is independent of affine connections on the Lie group and differs from Newton's method mentioned above. Motivated by looking for approaches to solve ordinary differential equations on manifolds, Owren and Welfert (2000) introduced Newton's method independent of affine connections on the Lie group for solving the equation $f(x)=0$, with f being a map from a Lie group to its corresponding Lie algebra, and showed that, under classical assumptions on f, Newton's method converges quadratically.

This paper is aimed at studying the same problem. Under the classical Lipschitz condition, the convergence criterion of Newton's method independent of affined connections is established and the radius of convergence ball is obtained. Moreover, the radii of the uniqueness balls of the zeros of f are estimated. It should be remarked that most part of the results obtained in this paper are new. In particular, the convergence criterion of the semi-local behavior of Newton's method independent of affined connections is given on arbitrary Lie groups, which had not been found to be studied to the best of our knowledge.

NOTIONS AND PRELIMINARIES

Most of the notions and notations which we will use in this paper are standard, see for example (Warner, 1983; Varadarajan, 1984).

A Lie group $(G,\cdot)$ is a Hausdorff topological group with contable bases, which also has the structure of a smooth manifold such that the group product and the inversion are smooth 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 it is finite. 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 will use the left translation of the Lie group G. We define for each $y\in G$

$$L_y: G\to G,\ z\mapsto y\cdot z$$

the left multiplication in the group. The differential of L_y at e denoted by $(\mathrm{d}L_y)_e$ determines an isomorphism of $\mathcal{G}=T_eG$ with the tangent space T_yG via the relation

$$(\mathrm{d}L_y)_e(\mathcal{G})=T_yG,$$

or, equivalently,

$$\mathcal{G}=(\mathrm{d}L_y)_e^{-1}(T_yG)=(\mathrm{d}L_{y^{-1}})_y(T_yG).$$

The exponential map is a map

$$\exp:\ \mathcal{G}\to G,\ u\to\exp(u),$$

which is certainly the most important construct associated to G and $\mathcal{G}$. Given $u\in\mathcal{G}$, the left invariant vector field $X_u: y\mapsto(\mathrm{d}L_y)_e(u)$ determines a one-parameter subgroup of G $\sigma_u:\mathbb{R}\to G$ such that $\sigma_u(0)=0$, and

$$\sigma_u'(t)=X_u(\sigma_u(t))=(\mathrm{d}L_{\sigma_u(t)})_e(u).$$

The exponential map is then defined by the relation $\exp(u)=\sigma_u(1)$. Note that the exponential map is not surjective in general. However, the exponential map is a diffeomorphism on an open neighborhood $\Gamma(0)$ of $0\in\mathcal{G}$. Let $N(e)=\exp(\Gamma(0))$, then for each $y\in N(e)$, there exists $v\in\Gamma(0)$ such that $y=\exp(v)$. Furthermore, if $\exp(u)=\exp(v)\in N(e)$ for some u, $v\in\Gamma(0)$, then $u=v$. If G is Abelian, exp is also a homomorphism from $\mathcal{G}$ to G, i.e.,

$$\exp(u+v)=\exp(u)\cdot\exp(v) \tag{1}$$

for all $u,v\in\mathcal{G}=T_eG$. In the non-Abelian case, exp is not a homomorphism and Eq.(1) must be replaced by

$$\exp(w)=\exp(u)\cdot\exp(v),$$

where w is given by the Baker-Campbell-Hausdorff (BCH) formula (Varadarajan, 1984)

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

for all u, v in an open neighborhood of $0\in\mathcal{G}$.

To analyse convergence, we need a Riemannian metric on the Lie group G. Following (Varadarajan, 1984), take an inner product $\langle\cdot,\cdot\rangle$ on $\mathcal{G}$ and define $\langle u,v\rangle_x=\langle(\mathrm{d}L_{x^{-1}})_x(u),(\mathrm{d}L_{x^{-1}})_x(v)\rangle_e$ for each $x\in G$ and $u,v\in T_xG$. This construct actually produces a Riemannian metric on the Lie group G (Varadarajan, 1984). Let $\|\cdot\|_x$ be associated norm, where the subscript x is sometimes omitted if there is no confusion. For any two distinct elements x, $y\in G$, let c:[0,1]→G be a piecewise smooth curve connecting x and y. Then the arc-length of c is defined by $l(c):=\int_0^1\|c'(t)\|\,\mathrm{d}t$, and the distance from x to y by $d(x,y):=\inf_c l(c)$, where the infimum is taken over all piecewise smooth curves c:[0,1]→G connecting x and y. Thus, we assume throughout the whole paper that G is connected and (G,d) is a complete metric space. Since we only deal with finite dimensional Lie algebras, every linear mapping $\varphi:\mathcal{G}\to\mathcal{G}$ is bounded and we define its norm by

$$\|\varphi\|=\sup_{u\neq 0}\big(\|\varphi(u)\|/\|u\|\big)=\sup_{\|u\|=1}\|\varphi(u)\|<\infty.$$

For r>0 we introduce the corresponding ball of radius r around $y\in G$ defined by one-parameter subgroups of G as

$$C_r(y)=\{z\in G: z=y\exp(u),\|u\|\le r\}.$$

Following (Owren and Welfert, 2000), we give the following definition:

Definition 1 Let $\{x_n\}_{n\ge 0}$ be a sequence of G and $x\in G$. Then $\{x_n\}_{n\ge 0}$ is said to be

(1) convergent to x if for any ε>0 there exists a natural number K such that $x^{-1}x_n\in N(e)$ and $\|\exp^{-1}(x^{-1}x_n)\|\le\varepsilon$ for all $n\ge K$;

(2) quadratically convergent to x if $\{\|\exp^{-1}(x^{-1}x_n)\|\}$ is quadratically convergent to 0; that is, $\{x_n\}_{n\ge 0}$ is convergent to x and there exists a constant q and a natural number K such that $\|\exp^{-1}(x^{-1}x_{n+1})\|\le q\|\exp^{-1}(x^{-1}x_n)\|^2$ for all $n\ge K$.

Note that convergence of a sequence $\{x_n\}_{n\ge 0}$ in G to x in the sense of Definition 1 is equivalent to that $\lim_{n\to\infty}d(x_n,x)=0$. In the remainder of this paper, let f: $G\to\mathcal{G}=T_eG$ be a C^1-mapping. The differential of f at a point $x\in G$ is a linear map $f'_x:T_xG\to\mathcal{G}$ defined by

$$f'_x(\Delta_x)=\frac{\mathrm{d}}{\mathrm{d}t}f\{x\exp[t(\mathrm{d}L_{x^{-1}})_x(\Delta_x)]\}\big|_{t=0} \quad (2)$$

for any $\Delta_x\in T_xG$. The differential f'_x can be expressed via a function $\mathrm{d}f_x:\mathcal{G}\to\mathcal{G}$ given by

$$\mathrm{d}f_x=(f\circ L_x)'_e=f'_x\circ(\mathrm{d}L_x)_e.$$

Thus, by Eq.(2), it follows that

$$\mathrm{d}f_x(u)=f'_x[(\mathrm{d}L_x)_e(u)]=\frac{\mathrm{d}}{\mathrm{d}t}f[x\exp(tu)]\big|_{t=0}$$

for any $u\in\mathcal{G}$. Therefore the following lemma is clear:

Lemma 1 Let $x\in G$, $u\in\mathcal{G}$ and $t\in\mathbb{R}$. Then

$$\frac{\mathrm{d}}{\mathrm{d}t}f[x\exp(-tu)]=-\mathrm{d}f_{x\exp(-tu)}(u), \quad (3)$$

and

$$f[x\exp(tu)]-f(x)=\int_0^t \mathrm{d}f_{x\exp(su)}(u)\mathrm{d}s. \quad (4)$$

Following (Owren and Welfert, 2000), Newton's method for f with initial point $x_0\in G$ is defined as

$$x_{n+1}=x_n\exp[-\mathrm{d}f_{x_n}^{-1}f(x_n)],\quad \forall n\in\mathbb{N}. \quad (5)$$

We end this section with some notions of different Lipschitz conditions and a useful lemma.

Definition 2 Let L>0 be a constant and r>0. Let $x_0\in G$ be such that $\mathrm{d}f_{x_0}^{-1}$ exists. Then $\mathrm{d}f_{x_0}^{-1}\mathrm{d}f$ is said to satisfy:

(1) the center Lipschitz condition with constant L in $C_r(x_0)$ if

$$\| \mathrm{d}f_{x_0}^{-1}(\mathrm{d}f_{x_0\exp(u)} - \mathrm{d}f_{x_0}) \| \le L\|u\| \quad (6)$$

for each $u\in\mathcal{G}$ with $\|u\|\le r$.

(2) the radius Lipschitz condition with constant L in $C_r(x_0)$ if

$$\| \mathrm{d}f_{x_0}^{-1}(\mathrm{d}f_{x_0\exp(u)} - \mathrm{d}f_{x_0\exp(\tau u)}) \| \le (1-\tau)L\|u\| \quad (7)$$

holds for any $\tau\in[0,1]$ and $u\in\mathcal{G}$ with $\|u\|\le r$.

Clearly, the radius Lipschitz condition implies the center Lipschitz condition.

Lemma 2 Let $L>0$ be a constant and $0<r<1/L$. Suppose that $\mathrm{d}f_{x_0}^{-1}\mathrm{d}f$ satisfies the center Lipschitz condition with constant L in $C_r(x_0)$. Let $x\in C_r(x_0)$ be such that $x=x_0\exp(u)$ with $\|u\|\le r$. Then $\mathrm{d}f_x^{-1}$ exists and

$$\| \mathrm{d}f_x^{-1}\circ \mathrm{d}f_{x_0} \| \le \frac{1}{1-L\|u\|}. \quad (8)$$

Proof By Eq.(6), $\| \mathrm{d}f_{x_0}^{-1}(\mathrm{d}f_{x_0\exp(u)} - \mathrm{d}f_{x_0} \| \le L\|u\| < 1$. It follows from the Banach lemma that $(\mathrm{d}f_{x_0}^{-1}\circ \mathrm{d}f_{x_0\exp(u)})^{-1}$ exists and

$$\| \mathrm{d}f_x^{-1}\circ \mathrm{d}f_{x_0} \| \le \frac{1}{1-L\|u\|}.$$

The proof is complete.

CONVERGENCE CRITERIA

This section is devoted to two convergence criteria of Newton's method. The first one is based on the information around the zero; while the second one around the initial point. Recall that $f:G\to T_eG$ is a C^1-mapping.

Theorem 1 Assume that the Lie group G is also an Abelian group. Let $L>0$ be a constant and $0<r<2/(3L)$. Let $x^*\in G$ be such that $f(x^*)=0$ and $\mathrm{d}f_{x^*}^{-1}$ exists. Suppose that $\mathrm{d}f_{x^*}^{-1}\mathrm{d}f$ satisfies the radius Lipschitz condition with constant L in $C_r(x^*)$. Then Newton's method Eq.(5) is well-defined and quadratically convergent to x^* for each point $x_0\in C_r(x^*)$; that is, the sequence $\{x_n\}$ generated by Newton's method Eq.(5) with initial point $x_0\in C_r(x^*)$ is convergent, and there exists $u_n\in\mathcal{G}$ with $\|u\|\le r$ such that $x_n=x^*\exp(u_n)$ and

$$\|u_{n+1}\| \le q\|u_n\|^2, \quad \forall n\in\mathbb{N}, \quad (9)$$

where

$$q=\frac{L}{2(1-L\|u_0\|)}. \quad (10)$$

Proof Let $x_0\in C_r(x^*)$, and let $u_0\in\mathcal{G}$ be such that $x_0=x^*\exp(u_0)$ and $\|u_0\|\le r$. Let q be defined by Eq.(10) and set $q_0=q\|u_0\|$. The $q_0<1$ because $\|u_0\|\le r\le 2/(3L)$. It suffices to verify that, $\forall n\in\mathbb{N}$, x_n is well-defined and there exists $u_n\in\mathcal{G}$ with $\|u_n\|\le r$ such that $x_n=x^*\exp(u_n)$, and

$$\|u_{n+1}\| \le q\|u_n\|^2 \le q_0^{2^{n+1}-1}\|u_0\|. \quad (11)$$

We will proceed by induction. It is clear in the case when $n=0$. Now assume that, $\forall n\in\mathbb{N}$, x_n is well-defined and there exists $u_n\in\mathcal{G}$ with $\|u_n\|\le r$ such that Eq.(11) holds. We have to prove that x_{k+1} is well-defined and that there exists $u_{k+1}\in\mathcal{G}$ with $\|u_{k+1}\|\le r$ such that Eq.(11) holds for $n=k+1$. Note that, by the assumption that Eq.(11) holds for $n=k$ and Lemma 2, $\mathrm{d}f_{x_k}^{-1}$ exists and

$$\| \mathrm{d}f_{x_k}^{-1}\circ \mathrm{d}f_{x^*} \| \le \frac{1}{1-L\|u_k\|}. \quad (12)$$

This shows that x_{k+1} is well-defined. Let $u_{k+1}=u_k-\mathrm{d}f_{x_k}^{-1}[(f(x_k)]$. Then, by Lemma 1,

$$\begin{aligned}
\|u_{k+1}\| &= \|u_k - \mathrm{d}f_{x_k}^{-1}[f(x_k)-f(x^*)]\| \\
&= \|u_k - \mathrm{d}f_{x_k}^{-1}\int_0^1 \mathrm{d}f_{x^*\exp(tu_k)}(u_k)\mathrm{d}t\| \\
&\le \|\mathrm{d}f_{x_k}^{-1}\circ \mathrm{d}f_{x^*}\| \cdot \int_0^1 \|\mathrm{d}f_{x^*}^{-1}(\mathrm{d}f_{x_k} - \mathrm{d}f_{x^*\exp(tu_k)})u_k\|\mathrm{d}t \\
&\le \frac{1}{1-L\|u_k\|}\int_0^1 (1-t)L\|u_k\|^2\mathrm{d}t \\
&= \frac{L}{2(1-L\|u_k\|)}\|u_k\|^2,
\end{aligned}$$

where the last inequality holds because of Eqs.(7) and (12). This implies that

$$\| u_{k+1} \| \le q \| u_k \|^2 \le q_0^{2^{k+1}-1} \| u_0 \|,$$

hence $\| u_{k+1} \| \le r$.

Since $x_{k+1} = x_k \exp\{-\mathrm{d}f_{x_k}^{-1}[f(x_k)]\}$, $x_k = x^* \exp(u_k)$ and G is an Abelian group, one has

$$x_{k+1} = x^* \exp(u_k) \cdot \exp\{-\mathrm{d}f_{x_k}^{-1}[f(x_k)]\} = x^* \exp(u_{k+1}).$$

Therefore, Eq.(11) is seen to hold for $n=k+1$ and the proof is complete.

For the second theorem, we need to introduce the quadratic majorizing function h, which was used in (Kantorovich and Akilov, 1982; Wang, 1999; Gutierrez and Hernandez, 2000) and defined by

$$h(t)=Lt^2/2-t-\beta \quad \text{for each } t\ge 0, \tag{13}$$

where $\beta>0$ and $L>0$. Write $\lambda=L\beta$, then if $\lambda\le 1/2$, h has two zeros r_1 and r_2:

$$r_1 = \frac{1-\sqrt{1-2\lambda}}{L}, \quad r_2 = \frac{1+\sqrt{1-2\lambda}}{L}. \tag{14}$$

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

$$t_{n+1}=t_n-[h'(t_n)]^{-1}h(t_n), \quad \forall n\in\mathbb{N}. \tag{15}$$

Then, if $\lambda\le 1/2$, $\{t_n\}$ is monotonically increasing and convergent to r_1. Moreover,

$$r_1 - t_n = \frac{(1-\zeta)\zeta^{2^n-1}}{1-\zeta^{2^n}} r_1, \quad \forall n\in\mathbb{N}, \tag{16}$$

where

$$\zeta = \left(1-\sqrt{1-2\lambda}\right)\Big/\left(1+\sqrt{1-2\lambda}\right). \tag{17}$$

We also need to introduce the metric closed ball of radius $r>0$ around $y\in G$, which is denoted by

$$B(y,r) = \{z \in G : d(z,y) \le r\}.$$

Clearly, $C_r(y)\subset B(y,r)$.

Definition 3 Let $r>0$ and let $x_0\in G$ be such that $\mathrm{d}f_{x_0}^{-1}$ exists. Then $\mathrm{d}f_{x_0}^{-1}\mathrm{d}f$ is said to satisfy

(1) the center Lipschitz condition with constant L in $B(x_0,r)$ if

$$\| \mathrm{d}f_{x_0}^{-1}(\mathrm{d}f_x - \mathrm{d}f_{x_0}) \| \le L d(x_0, x) \tag{18}$$

for each $x\in B(x_0,r)$;

(2) the Lipschitz condition with constant L in $B(x_0,r)$ if

$$\| \mathrm{d}f_{x_0}^{-1}(\mathrm{d}f_{x'} - \mathrm{d}f_x) \| \le L d(x', x) \tag{19}$$

holds for any $x',x\in B(x_0,r)$ such that $d(x_0,x)+d(x,x')\le r$.

Similar to Lemma 2, we have the following lemma:

Lemma 3 Let $L>0$ be a constant and $0<r<1/L$. Suppose that $\mathrm{d}f_{x_0}^{-1}\mathrm{d}f$ satisfies the center Lipschitz condition with constant L in $B(x_0,r)$. Then for each $x\in B(x_0,r)$, $\mathrm{d}f_x^{-1}$ exists and

$$\| \mathrm{d}f_x^{-1} \circ \mathrm{d}f_{x_0} \| \le \frac{1}{1-Ld(x_0,x)}. \tag{20}$$

Recall that $f{:}G\to T_eG$ is a C^1-mapping. In the remainder of this section, let $x_0\in G$ be such that $\mathrm{d}f_{x_0}^{-1}$ exists and set $\beta=\| \mathrm{d}f_{x_0}^{-1}[f(x_0)] \|$.

Theorem 2 Let $L>0$ be a constant. Suppose that $\mathrm{d}f_{x_0}^{-1}\mathrm{d}f$ satisfies the Lipschitz condition with constant L in $B(x_0,r_1)$. Let $\lambda=L\beta\le 1/2$. Then Newton's method Eq.(5) with initial point x_0 is well-defined and the generated sequence $\{x_n\}$ converges to a zero x^* of f in $\overline{B(x_0,r_1)}$. Moreover,

$$d(x_n, x^*) \le \frac{1-\zeta}{1-\zeta^{2^n}} \zeta^{2^n-1} r_1 \le \zeta^{2^n-1} r_1, \ \forall n\in\mathbb{N}. \tag{21}$$

Furthermore, if G is also an Abelian group, then there is a zero x^* of f in $C_{r_1}(x_0)$ such that $\forall n\ge 0$, there exists $u_n\in\mathcal{G}$ satisfying $x_n=x^*\exp(u_n)$,

$$\frac{\| u_n \|}{r_1 - t_n} \le \left(\frac{\| u_{n-1} \|}{r_1 - t_{n-1}}\right)^2 \le \cdots \le \left(\frac{\| u_0 \|}{r_1 - t_0}\right)^{2^n}, \tag{22}$$

and

$$\| u_n \| \le \frac{1-\zeta}{1-\zeta^{2^n}} \zeta^{2^n-1} \| u_0 \| \le \zeta^{2^n-1} \| u_0 \|. \tag{23}$$

Proof Recall that by Eq.(5), $x_{n+1}=x_n\exp\{-\mathrm{d}f_{x_n}^{-1}[f(x_n)]\}$ $\forall n\in\mathbb{N}$. Write $v_n=-\mathrm{d}f_{x_n}^{-1}[f(x_n)]$ for each n. We claim that

$$d(x_{n+1},x_n)\le\|v_n\|\le t_{n+1}-t_n,\quad \forall n\in\mathbb{N}. \tag{24}$$

Define the curve c_0 by $c_0(t):=x_0\exp(tv_0)$ for each $t\in[0,1]$, then c_0 is a smooth curve connecting x_0 and x_1, and $l(c_0)=\|v_0\|$. Hence, $d(x_1,x_0)\le l(c_0)=\|v_0\|$. Since $\|v_0\|=\|-\mathrm{d}f_{x_0}^{-1}[f(x_0)]\|=\beta\le t_1-t_0$, Eq.(24) is true for $n=0$. We now proceed by mathematical induction on n. For this purpose, assume that Eq.(24) holds for $n=0,1,\cdots,k-1$. Then

$$d(x_k,x_0)\le\sum_{i=0}^{k-1}d(x_{i+1},x_i)\le\sum_{i=0}^{k-1}\|v_i\|\le t_k<r_1. \tag{25}$$

It follows that $x_k\in B(x_0, r_1)$. Thus, we use Lemma 3 to conclude that $\mathrm{d}f_{x_k}^{-1}$ exists and

$$\|\mathrm{d}f_{x_k}^{-1}\circ\mathrm{d}f_{x_0}\|\le\frac{1}{1-Lt_k}=-[h'(t_k)]^{-1}. \tag{26}$$

Therefore, x_{k+1} is well-defined. Observe that

$$\begin{aligned}f(x_k)&=f(x_k)-f(x_{k-1})-\mathrm{d}f_{x_{k-1}}v_{k-1}\\&=\int_0^1\mathrm{d}f_{x_{k-1}\exp(tv_{k-1})}(v_{k-1})\mathrm{d}t-\mathrm{d}f_{x_{k-1}}v_{k-1}\text{(by Lemma 1)}\\&=\int_0^1[\mathrm{d}f_{x_{k-1}\exp(tv_{k-1})}-\mathrm{d}f_{x_{k-1}}]v_{k-1}\mathrm{d}t,\end{aligned}$$

Therefore, applying Eq.(19), one has that

$$\begin{aligned}&\|\mathrm{d}f_{x_0}^{-1}\circ f(x_k)\|\\&\le\int_0^1\|\mathrm{d}f_{x_0}^{-1}[\mathrm{d}f_{x_{k-1}\exp(tv_{k-1})}-\mathrm{d}f_{x_{k-1}}]\|\,\|v_{k-1}\|\,\mathrm{d}t\\&\le\int_0^1Ld(x_{k-1},x_{k-1}\exp(tv_{k-1}))\|v_{k-1}\|\,\mathrm{d}t\\&\le\int_0^1L\|tv_{k-1}\|\,\|v_{k-1}\|\,\mathrm{d}t\\&\le\frac{L}{2}(t_k-t_{k-1})^2=h(t_k).\end{aligned} \tag{27}$$

Consequently, combining Eqs.(26) and (27) yields

$$\begin{aligned}\|v_k\|&=\|-\mathrm{d}f_{x_k}^{-1}\circ f(x_k)\|\\&\le\|\mathrm{d}f_{x_k}^{-1}\circ\mathrm{d}f_{x_0}\|\,\|\mathrm{d}f_{x_0}^{-1}\circ f(x_k)\|\\&\le-[h'(t_k)]^{-1}h(t_k)=t_{k+1}-t_k.\end{aligned}$$

Define the curve c_k by $c_k(t):=x_k\exp(tv_k)$ for each $t\in[0,1]$. Then c_k is a smooth curve connecting x_k and x_{k+1}, and $l(c_k)=\|v_k\|$. Hence, $d(x_{k+1},x_k)\le l(c_k)=\|v_k\|$. Consequently, Eq.(24) holds for $n=k$ and the claim stands. Thus, Newton's method Eq.(5) with initial point x_0 is well-defined and the generated sequence $\{x_n\}$ converges to a zero x^* of f in $B(x_0, r_1)$. Eq.(21) follows from Eq.(16).

Now assume that G is an Abelian group. Let

$$u_n=-\sum_{k=n}^{\infty}v_k,\quad\forall n\in\mathbb{N}. \tag{28}$$

Then, by Eq.(24),

$$\|u_n\|\le r_1-t_n,\quad\forall n\in\mathbb{N}. \tag{29}$$

Set $x^*=x_0\exp(-u_0)$. Then $x^*\in C_{r_1}(x_0)$. Furthermore,

$$x_k=x_0\cdot\prod_{i=0}^{k-1}\exp(v_i)=x_0\exp\left(\sum_{i=0}^{k-1}v_i\right).$$

Therefore, by mathematical induction, it is easy to show that $x_n=x^*\exp(u_n)$ holds $\forall n\in\mathbb{N}$. Consequently, $\{x_n\}$ converges to x^* and x^* is a zero of f in $C_{r_1}(x_0)$. It remains to show that Eqs.(22) and (23) hold $\forall n\in\mathbb{N}$. Since Eq.(23) is a direct consequence of Eqs.(16) and (22), we only need to show Eq.(22) $\forall n\in\mathbb{N}$. We shall proceed by mathematical induction. Clearly Eq.(22) is trivial in the case when $n=0$. Now assume that Eq.(22) holds for $n=k$. Recall that $x_k=x^*\exp(u_k)$, $u_{k+1}=u_k+v_k$ and $v_k=-\mathrm{d}f_{x_k}^{-1}\circ f(x_k)$. It follows from Lemma 1 that

$$\begin{aligned}\|u_{k+1}\|&=\|u_k-\mathrm{d}f_{x_k}^{-1}\circ f(x_k)\|\\&=\|u_k-\mathrm{d}f_{x_k}^{-1}[f(x_k)-f(x^*)]\|\\&=\|\mathrm{d}f_{x_k}^{-1}\int_0^1[\mathrm{d}f_{x_k}-\mathrm{d}f_{x^*\exp(tu_k)}]u_k\mathrm{d}t\|.\end{aligned} \tag{30}$$

Hence, we conclude from Eq.(8) and Eq.(26) that

$$\|u_{k+1}\| \le \| \mathrm{d}f_{x_k}^{-1} \circ \mathrm{d}f_{x_0} \| \cdot \int_0^1 \| \mathrm{d}f_{x_0}^{-1}[\mathrm{d}f_{x_k} - \mathrm{d}f_{x^* \exp(tu_k)}]u_k \mathrm{d}t \|$$
$$\le \frac{1}{-h'(t_k)}\int_0^1 L(1-t)\|v_k\|^2 \,\mathrm{d}t = \frac{L}{2}\frac{(r_1-t_k)^2}{-h'(t_k)}\left(\frac{\|u_k\|}{r_1-t_k}\right)^2$$
$$= \frac{-h(t_k)+h'(t_k)(t_k-r_1)}{-h'(t_k)}\left(\frac{\|u_k\|}{r_1-t_k}\right)^2$$
$$= (r_1-t_{k+1})[\|u_k\|/(r_1-t_k)]^2,$$

where the last equality holds because of Eq.(15) (with $n=k$). Therefore, Eq.(22) holds for $n=k+1$ and the proof is complete.

Remark 1 The main result given in (Owren and Welfert, 2000) states that if the initial point is close sufficiently to a zero of f, then Newton's method on Lie group converges to the zero; while Theorem 2 above provides a Kantorovich's criterion for the convergence of Newton's method, which is not required to suppose the existence of a zero as a priori. Below we provide a simple example that shows the advantage of Theorem 2.

Example 1 Let N be a positive integer and let G be the special orthogonal group under standard matrix multiplication (Warner, 1983; Varadarajan, 1984), i.e.,

$$G = SO(N,\mathbb{R}) = \{\boldsymbol{x} \in \mathbb{R}^{N\times N} \mid \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{I}_N, \det \boldsymbol{x} = 1\},$$

where $\boldsymbol{I}_N$ is the $N\times N$ identity matrix. Then its Lie algebra is the set of all $N\times N$ skew-symmetric matrices

$$\mathcal{G} = so(N,\mathbb{R}) = \{\boldsymbol{v} \in \mathbb{R}^{N\times N} \mid \boldsymbol{v}^{\mathrm{T}} + \boldsymbol{v} = 0\}.$$

We endow $\mathcal{G}$ with the standard inner product

$$\langle \boldsymbol{u},\boldsymbol{v}\rangle = \mathrm{tr}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}) \text{ for each } \boldsymbol{u},\boldsymbol{v}\in\mathcal{G}. \tag{31}$$

Then the corresponding norm is the Frobenius norm $\|\cdot\|_{\mathrm{F}}$ defined by $\|\boldsymbol{u}\|_{\mathrm{F}} = \sqrt{\mathrm{tr}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u})}$ for each $\boldsymbol{u}\in\mathcal{G}$.

Let $\boldsymbol{A}\in\mathcal{G}$ be such that $\|\boldsymbol{A}\|_{\mathrm{F}}\le 1$ and define f: $G\to\mathcal{G}$ by $f(\boldsymbol{x})=\boldsymbol{x}-\boldsymbol{x}^{\mathrm{T}}+\boldsymbol{A}$ for each $\boldsymbol{x}\in G$. Let $\boldsymbol{x}\in G$, we have

$$f[\boldsymbol{x}\exp(t\boldsymbol{u})] = \boldsymbol{x}\sum_{n\ge 0}\frac{t^n}{n!}\boldsymbol{u}^n - \left(\boldsymbol{x}\sum_{n\ge 0}\frac{t^n}{n!}\boldsymbol{u}^n\right)^{\mathrm{T}} + \boldsymbol{A}$$

for each $\boldsymbol{u}\in\mathcal{G}$. Hence,

$$\mathrm{d}f_{\boldsymbol{x}}(\boldsymbol{u}) = \frac{\mathrm{d}}{\mathrm{d}t}f[\boldsymbol{x}\exp(t\boldsymbol{u})]|_{t=0} = \boldsymbol{x}\boldsymbol{u} - (\boldsymbol{x}\boldsymbol{u})^{\mathrm{T}} \tag{32}$$

for each $\boldsymbol{u}\in\mathcal{G}$. Taking $\boldsymbol{x}_0=\boldsymbol{I}_N$, it follows that $\mathrm{d}f_{\boldsymbol{x}_0} = 2\boldsymbol{I}_{N\times N}$ and so $\mathrm{d}f_{\boldsymbol{x}_0}^{-1} = \boldsymbol{I}_{N\times N}/2$. We claim that $\mathrm{d}f_{\boldsymbol{x}_0}^{-1}\mathrm{d}f$ satisfies the Lipschitz condition with constant $L=1$ in G, i.e.,

$$\|\mathrm{d}f_{\boldsymbol{x}_0}^{-1}(\mathrm{d}f_{\boldsymbol{x}'} - \mathrm{d}f_{\boldsymbol{x}})\|_{\mathrm{F}} \le d(\boldsymbol{x}',\boldsymbol{x}) \text{ for each } \boldsymbol{x}',\boldsymbol{x}\in G. \tag{33}$$

In fact, for any $\boldsymbol{x}', \boldsymbol{x}\in G$, since G is a simple connected compact Lie group (Warner, 1983; Varadarajan, 1984; Smith, 1993), there exists $\boldsymbol{v}\in\mathcal{G}=so(N, \mathbb{R})$ such that

$$\boldsymbol{x}'=\boldsymbol{x}\exp(\boldsymbol{v}) \text{ and } d(\boldsymbol{x}', \boldsymbol{x})=\|\boldsymbol{v}\|_{\mathrm{F}}. \tag{34}$$

Noting that $\mathrm{d}f_{\boldsymbol{x}_0}^{-1} = \boldsymbol{I}_{N\times N}/2$, one has that, for each $\boldsymbol{u}\in\mathcal{G}$,

$$\|\mathrm{d}f_{\boldsymbol{x}_0}^{-1}(\mathrm{d}f_{\boldsymbol{x}'} - \mathrm{d}f_{\boldsymbol{x}})(\boldsymbol{u})\|_{\mathrm{F}} = \frac{1}{2}\|(\mathrm{d}f_{\boldsymbol{x}_0\exp(\boldsymbol{v})} - \mathrm{d}f_{\boldsymbol{x}})(\boldsymbol{u})\|_{\mathrm{F}}. \tag{35}$$

Fix $\boldsymbol{u}\in\mathcal{G}$ and define the map g:$G\to\mathcal{G}$ by $g(\boldsymbol{x})=\boldsymbol{x}\boldsymbol{u}-(\boldsymbol{x}\boldsymbol{u})^{\mathrm{T}}$ for each $\boldsymbol{x}\in G$. Then, for each $s\in[0,1]$,

$$\mathrm{d}g_{\boldsymbol{x}\exp(s\boldsymbol{v})}(\boldsymbol{v})=\boldsymbol{x}\exp(s\boldsymbol{v})\boldsymbol{v}\boldsymbol{u}-[\boldsymbol{x}\exp(s\boldsymbol{v})\boldsymbol{v}\boldsymbol{u}]^{\mathrm{T}}. \tag{36}$$

Since $\boldsymbol{x}\exp(s\boldsymbol{v})[\boldsymbol{x}\exp(s\boldsymbol{v})]^{\mathrm{T}}=\boldsymbol{I}_N$, it follows that

$$\|\boldsymbol{x}\exp(s\boldsymbol{v})\boldsymbol{v}\boldsymbol{u}\|_{\mathrm{F}} \le \|\boldsymbol{v}\boldsymbol{u}\|_{\mathrm{F}},$$

and

$$\|[\boldsymbol{x}\exp(s\boldsymbol{v})\boldsymbol{v}\boldsymbol{u}]^{\mathrm{T}}\|_{\mathrm{F}} \le \|\boldsymbol{v}\boldsymbol{u}\|_{\mathrm{F}}.$$

Hence

$$\|\mathrm{d}g_{\boldsymbol{x}\exp(s\boldsymbol{v})}(\boldsymbol{v})\|_{\mathrm{F}} \le 2\|\boldsymbol{v}\boldsymbol{u}\|_{\mathrm{F}} \le 2\|\boldsymbol{v}\|_{\mathrm{F}}\|\boldsymbol{u}\|_{\mathrm{F}}.$$

This together with Eq.(4) implies that

$$\begin{aligned}&\|g(\boldsymbol{x}\exp(\boldsymbol{v})) - g(\boldsymbol{x})\|_{\mathrm{F}}\\ &\le \int_0^1 \|\mathrm{d}g_{\boldsymbol{x}\exp(s\boldsymbol{v})}(\boldsymbol{v})\|_{\mathrm{F}}\,\mathrm{d}s \le 2\|\boldsymbol{v}\|_{\mathrm{F}}\|\boldsymbol{u}\|_{\mathrm{F}}.\end{aligned} \tag{37}$$

It follows from Eqs.(32) and (35) that

$$\| \mathrm{d}f_{\boldsymbol{x}_0}^{-1}(\mathrm{d}f_{\boldsymbol{x}'} - \mathrm{d}f_{\boldsymbol{x}})(\boldsymbol{u}) \|_{\mathrm{F}} = \frac{1}{2} \| g(\boldsymbol{x}\exp(\boldsymbol{v})) - g(\boldsymbol{x}) \|_{\mathrm{F}} \leq \| \boldsymbol{v} \|_{\mathrm{F}} \| \boldsymbol{u} \|_{\mathrm{F}} . \quad (38)$$

As $\boldsymbol{u}\in\mathcal{G}$ is arbitrary, Eq.(33) is seen to hold by Eq.(34).

On the other hand, since $f(\boldsymbol{x}_0)=\boldsymbol{A}$, it follows that

$$L\beta = \| \mathrm{d}f_{\boldsymbol{x}_0}^{-1} f(\boldsymbol{x}_0) \|_{\mathrm{F}} = \frac{1}{2} \| \boldsymbol{A} \|_{\mathrm{F}} \leq \frac{1}{2}.$$

Therefore, Theorem 2 is applicable for concluding that the sequence generated by Eq.(5) with initial point $\boldsymbol{x}_0=\boldsymbol{I}_N$ converges to a zero $\boldsymbol{x}^*$ of f. Note that the convergence theorem given by Owren and Welfert (2000) is not applicable because we do not know whether $\boldsymbol{x}_0$ is close sufficiently to the zero $\boldsymbol{x}^*$.

UNIQUENESS OF THE ZERO OF f

Recall that f:$G\rightarrow T_eG$ is a C^1-mapping. This section is devoted to estimating the radii of the uniqueness balls of the zeros of f. As in the previous section, the first theorem is concerned with an estimate of the radii of the uniqueness balls around the zero, while the second theorem around the initial point.

Theorem 3 Let $L>0$ be a constant and $0<r<2/L$. Suppose that $f(x^*)=0$ and $\mathrm{d}f_{x^*}^{-1}\mathrm{d}f$ satisfies the center Lipschitz condition with constant L in $C_r(x^*)$. Then x^* is the unique zero of f in $C_r(x^*)$.

Proof Suppose on the contrary that y^* is another zero of f in $C_r(x^*)$. Then there exists $u\in\mathcal{G}$ such that $y^*=x^*\exp(u)$ and $\| u \|\leq r$. Note that by Lemma 1 and Eq.(6),

$$\begin{aligned} \| u \| &= \| -\mathrm{d}f_{x^*}^{-1}[f(y^*) - f(x^*)] + u \| \\ &= \| -\mathrm{d}f_{x^*}^{-1} \int_0^1 \mathrm{d}f_{x^*\exp(tu)}(u)\mathrm{d}t + u \| \\ &= \| -\mathrm{d}f_{x^*}^{-1} \int_0^1 (\mathrm{d}f_{x^*\exp(tu)} - \mathrm{d}f_{x^*})u\mathrm{d}t \| \\ &\leq \int_0^1 tL \| u \|^2 \,\mathrm{d}t = \frac{L}{2} \| u \|^2 . \end{aligned}$$

Hence, $\|u\|\geq 2/L$, which contradicts that $\|u\|\leq r<2/L$. The proof is complete.

Recall that the majorizing function h is defined by Eq.(13) and that h has two zeros r_1, r_2 given by Eq.(14). Given $t_0\in[0, r_2]$, we define

$$t_{n+1}=t_n+h(t_n), \quad \forall n\in\mathbb{N}. \quad (39)$$

Then the following proposition about the convergence property of the sequence $\{t_n\}$ is immediate.

Proposition 1 Let $t_0\in[0, r_2]$. Then the following assertions hold:

(1) $\forall n\in\mathbb{N}$: $0\leq t_n\leq r_1$ if $t_0\in[0, r_1]$; and $r_1 \leq t_n < r_2$ if $t_0\in[r_1,r_2)$.

(2) $\{t_n\}$ converges to r_1 monotonically.

Let $x_0\in G$ be such that $\mathrm{d}f_{x_0}^{-1}$ exists and consider the sequence $\{x_n\}$ defined by

$$x_{n+1} = x_n \exp[-\mathrm{d}f_{x_0}^{-1} \circ f(x_n)], \quad \forall n\in\mathbb{N}. \quad (40)$$

The following lemma describes its convergence property, the proof of which is almost the same as that in Theorem 2 for the sequence defined by Eq.(5).

Lemma 4 Assume that the Lie group G is also an Abelian group. Let $L>0$ be a constant and $x_0\in G$ be such that $\mathrm{d}f_{x_0}^{-1}$ exists. Suppose that $\mathrm{d}f_{x_0}^{-1}\mathrm{d}f$ satisfies the center Lipschitz condition with constant L in $C_{r_1}(x_0)$. Let

$$\lambda=L\beta\leq 1/2. \quad (41)$$

Then the sequence $\{x_n\}$ generated by Eq.(40) with initial point x_0 is well-defined and converges to a zero x^* of f in $C_{r_1}(x_0)$.

Theorem 4 Assume that the Lie group G is also an Abelian group. Let $L>0$ be a constant and $x_0\in G$ be such that $\mathrm{d}f_{x_0}^{-1}$ exists. Let $\lambda=L\beta\leq 1/2$. Let $r_1\leq r<r_2$ if $\lambda<1/2$ and $r=r_1$ if $\lambda=1/2$. Suppose that $\mathrm{d}f_{x_0}^{-1}\mathrm{d}f$ satisfies the center Lipschitz condition with constant L in $C_r(x_0)$. Then there exists a unique zero of f in $C_r(x_0)$.

Proof Let $\{x_n\}$ denote the sequence generated by Eq.(40), i.e., $x_{n+1} = x_n \exp[-\mathrm{d}f_{x_0}^{-1} \circ f(x_n)]$, $\forall n\in\mathbb{N}$. Then, by Lemma 4, $\{x_n\}$ converges to a zero x^* of f in $C_{r_1}(x_0)$. Thus, to complete the proof, it is sufficient to show that x^* is the unique zero of f in $C_r(x_0)$. To this end, we assume that y^* is another zero of f in $C_r(x_0)$.

Then, there exists $u_0 \in \mathcal{G}$ such that

$$y^* = x_0 \exp(-u_0) \text{ and } \| u_0 \| \le r. \tag{42}$$

Let $\{t_n\}$ and $\{t'_n\}$ denote the sequences generated by Eq.(39) with initial point $t_0=0$ and $\{t'_0\}=\|u_0\|$, respectively. Write $v_n = -\mathrm{d}f_{x_0}^{-1} \circ f(x_n)$ and $u_{n+1}=u_n+v_n$ $\forall n\in \mathbb{N}$. Note that Eq.(42) implies $x_0=y^*\exp(u_0)$. Moreover, applying Eq.(42), Lemma 1 and Eq.(6), one gets that

$$\begin{aligned} \|u_1\| &= \left\|u_0 - \mathrm{d}f_{x_0}^{-1} \circ f(x_0)\right\| \\ &= \left\|\mathrm{d}f_{x_0}^{-1}[f(y^*) - f(x_0) + \mathrm{d}f_{x_0}(u_0)]\right\| \\ &= \left\|\mathrm{d}f_{x_0}^{-1}\{f[x_0 \exp(-u_0)] - f(x_0) + \mathrm{d}f_{x_0}(u_0)\}\right\| \\ &= \left\|\mathrm{d}f_{x_0}^{-1}\int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} f[x_0 \exp(-tu_0)]\mathrm{d}t + \mathrm{d}f_{x_0}(u_0)\right\| \\ &= \int_0^1 \left\|\mathrm{d}f_{x_0}^{-1}(-\mathrm{d}f_{x_0 \exp(-tu_0)} + \mathrm{d}f_{x_0})\right\| \| u_0 \| \,\mathrm{d}t \\ &\le \int_0^1 L(t \| u_0 \|) \| u_0 \| \,\mathrm{d}t = t'_1 - t_1. \end{aligned} \tag{43}$$

Since G is also an Abelian group, using mathematical induction, one can easily verify that $x_n=y^*\exp(u_n)$ and $\|u_n\| \le t'_n - t_n$. This, together with Proposition 1, implies that $y^*=\lim_{n\to\infty}x_n$; hence $y^*=x^*$ and this completes the proof.

References

Adler, R.L., Dedieu, J.P., Margulies, J.Y., Martens, M., Shub, M., 2002. Newton's method on Riemannian manifolds and a geometric model for the human spine. *IMA J. Numer. Anal.*, **22**(3):359-390. [doi:10.1093/imanum/22.3.359]

Dedieu, J.P., Priouret, P., Malajovich, G., 2003. Newton's method on Riemannian manifolds: covariant alpha theory. *IMA J. Numer. Anal.*, **23**:395-419. [doi:10.1093/imanum/23.3.395]

Edelman, A., Arias, T.A., Smith, T., 1998. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, **20**:303-353. [doi:10.1137/S0895479895290954]

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

Gabay, D., 1982. Minimizing a differentiable function over a differential manifold. *J. Optim. Theory Appl.*, **37**:177-219. [doi:10.1007/BF00934767]

Gutierrez, J.M., Hernandez, M.A., 2000. Newton's method under weak Kantorovich conditions. *IMA J. Numer. Anal.*, **20**:521-532. [doi:10.1093/imanum/20.4.521]

Kantorovich, L.V., 1948. On Newton method for functional equations. *Dokl. Acad. Nauk*, **59**(7):1237-1240.

Kantorovich, L.V., Akilov, G.P., 1982. Functional Analysis. Pergamon, Oxford.

Li, C., Wang, J.H., 2005. Convergence of Newton's method and uniqueness of zeros of vector fields on Riemannian manifolds. *Sci. in China (Ser. A)*, **48**(11):1465-1478. [doi:10.1360/04ys0147]

Li, C., Wang, J.H., 2006. Newton's method on Riemannian manifolds: Smale's point estimate theory under the γ-condition. *IMA J. Numer. Anal.*, **26**(2):228-251. [doi:10.1093/imanum/dri039]

Mahony, R.E., 1996. The constrained Newton method on a Lie group and the symmetric eigenvalue problem. *Linear Algebra Appl.*, **248**:67-89. [doi:10.1016/0024-3795(95)00171-9]

Owren, B., Welfert, B., 2000. The Newton iteration on Lie groups. *BIT Numer. Math.*, **40**:121-145. [doi:10.1023/A:1022322503301]

Smale, S., 1986. 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. Springer, New York, p.185-196.

Smith, S.T., 1993. Geometric Optimization Method for Adaptive Filtering. Ph.D Thesis, Harvard University, Cambridge, MA.

Smith, S.T., 1994. Optimization Techniques on Riemannian Manifolds. Fields Institute Communications. American Mathematical Society, Providence, RI, **3**:113-146.

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

Varadarajan, V.S., 1984. Lie Groups, Lie Algebras and Their Representations. GTM No. 102. Springer-Verlag, New York.

Wang, X.H., 1997. Convergence on the iteration of Halley family in weak conditions. *Chin. Sci. Bull.*, **42**:552-555. (in Chinese).

Wang, X.H., Han, D.F., 1997. Criterion α and Newton's method in weak condition. *Chin. J. Numer. Appl. Math.*, **19**:96-105.

Wang, X.H., 1999. Convergence of Newton's method and inverse function theorem in Banach space. *Math. Comput.*, **68**:169-186. [doi:10.1090/S0025-5718-99-00999-0]

Wang, J.H., Li, C., 2006. Uniqueness of the singular point of vector field on Riemannian manifold under the γ-condition. *J. Complexity*, **22**:533-548. [doi:10.1016/j.jco.2005.11.004]

Warner, F.W., 1983. Foundations of Differentiable Manifolds and Lie Groups. GTM No. 94. Springer-Verlag, New York.