

Kantorovich's theorems for Newton's method for mappings and optimization problems on Lie groups

JIN-HUA WANG[†]

Department of Mathematics, Zhejiang University of Technology, Hangzhou 310032,
People's Republic of China

AND

CHONG LI[‡]

Department of Mathematics, Zhejiang University, Hangzhou 310027,
People's Republic of China

[Received on 4 June 2008; revised on 1 April 2009]

With the classical assumptions on f , a convergence criterion of Newton's method (independent of affine connections) to find zeros of a mapping f from a Lie group to its Lie algebra is established, and estimates of the convergence domains of Newton's method are obtained, which improve the corresponding results in Owren & Welfert (2000, *BIT Numer. Math.*, **40**, 121–145) and Wang & Li (2006, *J. Zhejiang Univ. Sci. A*, **8**, 978–986). Applications to optimization are provided and the results due to Mahony (1996, *Linear Algebra Appl.*, **248**, 67–89) are extended and improved accordingly.

Keywords: Newton's method; Lie group; Lipschitz condition.

1. Introduction

Recently, there has been an increased interest in studying numerical algorithms on manifolds. Classical examples are given by eigenvalue problems, symmetric eigenvalue problems, invariant subspace computations, optimization problems with equality constraints, etc. (see, for example, Luenberger, 1972; Gabay, 1982; Smith, 1993, 1994; Mahony, 1994, 1996; Udriste, 1994; Edelman *et al.*, 1998; Owren & Welfert, 2000; Adler *et al.*, 2002; Absil *et al.*, 2007; Li *et al.*, 2009a). In particular, optimization problems on Lie groups or homogeneous spaces have been studied recently in the context of using continuous-time differential equations for solving problems in numerical linear algebra. For example, let $\phi: G \rightarrow \mathbb{R}$ be given by

$$\phi(x) = -\text{tr}(x^T Q x D) \quad \text{for each } x \in G, \quad (1.1)$$

where $G = \text{SO}(N, \mathbb{R}) := \{x \in \mathbb{R}^{N \times N} | x^T x = \mathbf{I}_N \text{ and } \det x = 1\}$, $D \in \mathbb{R}^{N \times N}$ is the diagonal matrix with diagonal entries $1, 2, \dots, N$ and Q is a fixed symmetric matrix. Brockett (1988, 1991) and Chu & Driessel (1990) considered the following optimization problem:

$$\min_{x \in G} \phi(x). \quad (1.2)$$

[†]Corresponding author. Email: wjh@zjut.edu.cn

[‡]Email: cli@zju.edu.cn

Brockett (1988, 1991) showed that the minimum $x^* \in G$ occurs when $x^{*\text{T}} Q x^*$ is a diagonal matrix with diagonal entries (eigenvalues of Q) in ascending order. Thus solving the nonlinear optimization problem is equivalent to solving the numerical linear algebra problem of computing the eigenvalues and eigenvectors of the matrix Q . There are many other examples where the linear algebra problems are formulated as optimization problems on Lie groups or homogeneous spaces (cf. Bayer & Lagarias, 1989; Helmke & Moore, 1990; Smith, 1991; Mahony *et al.*, 1993). For a general differentiable function ϕ on a Lie group some numerical optimization algorithms for solving (1.2) have been studied extensively (see, for example, Brockett, 1993; Mahony *et al.*, 1993; Shub & Smale, 1993; Smith, 1993; Mahony, 1994; Moore *et al.*, 1994). In particular, Mahony used one-parameter subgroups of a Lie group to develop a version of Newton's method on an arbitrary Lie group in Mahony (1996), where the approach for solving (1.2) via Newton's method was explored and the local convergence was analysed. In this manner the algorithm presented is independent of affine connections on the Lie group.

On the other hand, motivated by looking for approaches to solving ordinary differential equations on Lie groups, Owren & Welfert (2000) considered the implicit method for Lie groups, where they used the implicit Euler method as a generic example of an implicit integration method in a Lie group setting. Consider the initial value problem on G

$$\begin{cases} x' = x \cdot g(x), \\ x(0) = x^{(0)}, \end{cases} \tag{1.3}$$

where $g: G \rightarrow \mathcal{G}$ is a differentiable map and $x^{(0)}$ is a random starting point. The application of one step of the backward Euler method on (1.3) leads to the fixed-point problem

$$x = x^{(0)} \cdot \exp(lg(x)), \tag{1.4}$$

where l represents the size of the discretization step. Clearly, solving the problem (1.4) is equivalent to solving the equation

$$f(x) = 0, \tag{1.5}$$

where $f: G \rightarrow \mathcal{G}$ is the mapping defined by

$$f(x) = \exp^{-1}((x^{(0)})^{-1} \cdot x) - lg(x) \quad \text{for each } x \in G.$$

Owren & Welfert (2000) introduced Newton's method, independent of affine connections on the Lie group, for solving the equation (1.5) and showed that Newton's method for the map with its differential satisfying the classical Lipschitz condition is locally convergent quadratically. Recently, the authors of the present paper studied the problems of existence, uniqueness of solutions of (1.5) and estimates of convergence balls of Newton's method for (1.5) in Wang & Li (2007), where, however, all results except Theorem 2 of Wang & Li (2007) on the convergence of Newton's method were for Abelian groups. Extensions of Smale's point estimate theory for Newton's method on Lie groups were presented in Li *et al.* (2009b).

In a vector space framework, as is well known, one of the most important results on Newton's method is Kantorovich's theorem (cf. Kantorovich & Akilov, 1982). Under the mild condition that the second Fréchet derivative of F is bounded (or, more generally, 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, based on the knowledge of the first derivative around the initial point, ensuring the

existence, uniqueness of the solution of the equation and the quadratic convergence of Newton's method. Another important result on Newton's method is Smale's (1986) point estimate theory (i.e., α -theory and γ -theory), where the notion of approximate zeros was introduced and the rules for judging an initial point x_0 to be an approximate zero were established, depending on the information of the analytic non-linear operator at this initial point and at a solution x^* . There has been a lot of work on the weakness and/or the extension of the Lipschitz continuity made on the mappings (see, for example, Zabrejko & Nguen (1987), Gutiérrez & Hernández (2000), Wang (2000), Ezquerro & Hernández (2002a,b) and the references therein). In particular, Zabrejko & Nguen (1987) parameterized the classical Lipschitz continuity. Wang (2000) introduced the notion of Lipschitz conditions with an 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 Ferreira & Svaiter (2002) for Newton's method for vector fields on Riemannian manifolds, while extensions of the famous Smale's (1986) α -theory and γ -theory to analytic vector fields and analytic mappings on Riemannian manifolds were provided in Dedieu *et al.* (2003). In the recent paper Li & Wang (2006) the convergence criteria in Dedieu *et al.* (2003) 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 Wang & Li (2006), while the local behaviour of Newton's method on Riemannian manifolds was studied in Li & Wang (2005). Recently, inspired by the previous work of Zabrejko & Nguen (1987) on Kantorovich's majorant method, Alvarez *et al.* (2008) introduced a Lipschitz-type radial function for the covariant derivative of vector fields and mappings on Riemannian manifolds and established a unified convergence criterion for Newton's method on Riemannian manifolds.

In the spirit of the works mentioned above, a natural and interesting problem is whether an analogue of the well-known Kantorovich's theorem (independent of the connection) can be established for Newton's method (for solving (1.5) and/or the optimization problem (1.2)) on Lie groups. On a finite-dimensional completed and connected Riemannian manifold M , Newton's method is defined in terms of geodesics, and the property that there is at least one geodesic to connect any two points of M plays a key role in the study. Newton's method on a Lie group is defined in terms of one-parameter subgroups. However, there is no similar property for one-parameter subgroups in an arbitrary Lie group, which makes the study on a Lie group more complicated. In the recent paper Wang & Li (2007) we gave a kind of Kantorovich's theorem for (1.5) by using the following metric Lipschitz condition at x_0 :

$$\left\| df_{x_0}^{-1}(df_{x'} - df_x) \right\| \leq Ld(x', x) \quad \forall x', x \in G \text{ with } d(x_0, x) + d(x, x') < r_1, \quad (1.6)$$

where $d(\cdot, \cdot)$ is the Riemannian distance induced by the left-invariant Riemannian metric. Clearly, this kind of Lipschitz condition is still dependent on the metric on the underlying group and is generally very difficult to verify in a noncompact group (cf. Example 3.8 in Section 3).

The purpose of the present paper is to establish Kantorovich's theorem (independent of the connection) for Newton's method on a 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 is independent of the metric and weaker than the metric Lipschitz condition (1.6)), the convergence criterion of Newton's method for solving (1.5) is established in Section 3. As a consequence, estimates of the convergence domains are also obtained. The main feature of our results (Theorems 3.1 and 3.3) is that they are completely independent of the metrics on the groups. In particular, Theorem 3.1 improves and extends Theorem 2 of Wang & Li (2007), while Theorem 3.3 (and its corollaries) improves and extends the corresponding result in Owren & Welfert (2000). In Section 4 we will show that Newton's

method for solving the problem (1.2) is equivalent to the one for solving (1.5), where f is a map from a Lie group to its Lie algebra associated to ϕ , and, as applications, the convergence criterion and the estimates of the convergence domains of Newton's method for solving the problem (1.2) are provided. The results on the convergence domains improve the corresponding results due to Mahony (1996), while the result on the convergence criterion seems new for Newton's method for solving the problem (1.2). Examples are provided to show that our results in the present paper are applicable but not the results in Mahony (1996) and Owren & Welfert (2000).

2. Notions and preliminaries

Most of the notions and notation that are used in the present paper are standard (see, for example, Helgason, 1978; Varadarajan, 1984). A Lie group (G, \cdot) is a Hausdorff topological group with countable bases that 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 that is the tangent space $T_e G$ of G at e , equipped with the Lie bracket $[\cdot, \cdot]: \mathcal{G} \times \mathcal{G} \rightarrow \mathcal{G}$.

In what follows we will make use of the left translation of the Lie group G . We define for each $y \in G$ the left translation $L_y: G \rightarrow G$ 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 $\exp: \mathcal{G} \rightarrow G$ is certainly the most important construction associated to G and \mathcal{G} and is defined as follows. Given $u \in \mathcal{G}$ let $\sigma_u: \mathbb{R} \rightarrow G$ be the 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

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

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 neighbourhood 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 formula (cf. Varadarajan, 1984, p. 114), (2.5) must be replaced by

$$\exp(w) = \exp(u) \cdot \exp(v) \quad (2.6)$$

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

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

Let $f: G \rightarrow \mathcal{G}$ be a C^1 -map and let $x \in G$. We use f'_x to denote the differential of f at x . Then, by DoCarmo (1992, p. 9) (the proof given there for a smooth mapping still works for a C^1 -map), for each $\Delta_x \in T_x G$ and any nontrivial smooth curve $c: (-\varepsilon, \varepsilon) \rightarrow G$ with $c(0) = x$ and $c'(0) = \Delta_x$, one has that

$$f'_x \Delta_x = \left(\frac{d}{dt} (f \circ c)(t) \right)_{t=0}. \quad (2.8)$$

In particular,

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

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

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

Then, by (2.9), we have

$$df_x = f'_x \circ (L'_x)_e. \quad (2.11)$$

Also, by definition, we have for all $t \geq 0$ that

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

and

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

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 that plays a key role in the study. Let $x, y \in G$ and define

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

where we adopt the convention that $\inf \emptyset = +\infty$. It is easy to verify that $\varrho(\cdot, \cdot)$ is a distance on G and that 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,

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

Let $\mathcal{L}(\mathcal{G})$ denote the set of all linear operators on \mathcal{G} . Below we shall use the notion of the L -Lipschitz condition and a useful lemma.

DEFINITION 2.1 Let $r > 0$, let $x_0 \in G$ 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

$$\|T(x \cdot \exp u) - T(x)\| \leq L\|u\| \tag{2.15}$$

holds for any $u \in \mathcal{G}$ and $x \in C_r(x_0)$ such that $\|u\| + \varrho(x, x_0) < r$.

LEMMA 2.2 Let $0 < r \leq \frac{1}{L}$ and let $x_0 \in G$ be such that $df_{x_0}^{-1}$ exists. Suppose that $df_{x_0}^{-1}df$ satisfies the L -Lipschitz condition on $C_r(x_0)$. Let $x \in C_r(x_0)$ be such that there exist $k \geq 1$ and $u_0, \dots, u_k \in \mathcal{G}$ satisfying $x = x_0 \cdot \exp u_0 \cdots \exp u_k$ and $\sum_{i=0}^k \|u_i\| < r$. Then df_x^{-1} exists and

$$\|df_x^{-1}df_{x_0}\| \leq \frac{1}{1 - L(\sum_{i=0}^k \|u_i\|)}. \tag{2.16}$$

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

$$\|df_{x_0}^{-1}(df_{y_i \cdot \exp u_i} - df_{y_i})\| \leq L\|u_i\| \quad \text{for each } 0 \leq i \leq k. \tag{2.17}$$

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

$$\begin{aligned} \|df_{x_0}^{-1}df_x - \mathbf{I}_{\mathcal{G}}\| &= \|df_{x_0}^{-1}(df_{y_k \cdot \exp u_k} - df_{x_0})\| \\ &\leq \sum_{i=0}^k \|df_{x_0}^{-1}(df_{y_i \cdot \exp u_i} - df_{y_i})\| \\ &= L \left(\sum_{i=0}^k \|u_i\| \right) \\ &< 1. \end{aligned}$$

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

3. Convergence criteria

Following Owren & Welfert (2000), we define Newton's method with initial point x_0 for f on a Lie group as follows:

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

Let $\beta > 0$ and $L > 0$. The quadratic majorizing function h , which was used in Kantorovich & Akilov (1982) and Wang (2000), is defined by

$$h(t) = \frac{L}{2}t^2 - t + \beta \quad \text{for each } t \geq 0. \tag{3.2}$$

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

$$t_{n+1} = t_n - h'(t_n)^{-1}h(t_n) \quad \text{for each } n = 0, 1, \dots \tag{3.3}$$

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

$$r_1 = \frac{1 - \sqrt{1 - 2\lambda}}{L} \quad \text{and} \quad r_2 = \frac{1 + \sqrt{1 - 2\lambda}}{L}. \quad (3.4)$$

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

$$r_1 - t_n = \frac{\xi^{2^n - 1}}{\sum_{j=0}^{2^n - 1} \xi^j} r_1 \quad \text{for each } n = 0, 1, \dots, \quad (3.5)$$

where

$$\xi = \frac{1 - \sqrt{1 - 2\lambda}}{1 + \sqrt{1 - 2\lambda}}. \quad (3.6)$$

Recall that $f: G \rightarrow \mathcal{G}$ 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 $\beta := \|df_{x_0}^{-1} f(x_0)\|$.

THEOREM 3.1 Suppose that $df_{x_0}^{-1} df$ satisfies the L -Lipschitz condition on $C_{r_1}(x_0)$ and that

$$\lambda = L\beta \leq \frac{1}{2}. \quad (3.7)$$

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, for each $n = 0, 1, \dots$ the following assertions hold:

$$\varrho(x_{n+1}, x_n) \leq \|df_{x_n}^{-1} f(x_n)\| \leq t_{n+1} - t_n, \quad (3.8)$$

$$\varrho(x_n, x^*) \leq \frac{\xi^{2^n - 1}}{\sum_{j=0}^{2^n - 1} \xi^j} r_1. \quad (3.9)$$

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

$$\varrho(x_{n+1}, x_n) \leq \|v_n\| \leq t_{n+1} - t_n \quad (3.10)$$

holds for each $n = 0, 1, \dots$. Given 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),

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

Furthermore, assertions (3.8) and (3.9) 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 (3.10) is true for $n = 0$. We now proceed by mathematical induction on n . For this purpose we assume that v_n is well defined and (3.10) holds for each $n \leq k - 1$. Then

$$\sum_{i=0}^{k-1} \|v_i\| \leq t_k - t_0 = t_k < r_1 \quad \text{and} \quad x_k = x_0 \cdot \exp v_0 \cdots \exp v_{k-1}. \quad (3.11)$$

Thus we use Lemma 2.2 to conclude that $df_{x_k}^{-1}$ exists and

$$\|df_{x_k}^{-1}df_{x_0}\| \leq \frac{1}{1-Lt_k} = -h'(t_k)^{-1}. \tag{3.12}$$

Therefore v_k is well defined. Observe that

$$\begin{aligned} f(x_k) &= f(x_k) - f(x_{k-1}) - df_{x_{k-1}}v_{k-1} \\ &= \int_0^1 df_{x_{k-1} \cdot \exp(tv_{k-1})}v_{k-1} dt - df_{x_{k-1}}v_{k-1} \\ &= \int_0^1 [df_{x_{k-1} \cdot \exp(tv_{k-1})} - df_{x_{k-1}}]v_{k-1} dt, \end{aligned}$$

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

$$\begin{aligned} \|df_{x_0}^{-1}f(x_k)\| &\leq \int_0^1 \|df_{x_0}^{-1}[df_{x_{k-1} \cdot \exp(tv_{k-1})} - df_{x_{k-1}}]\| \|v_{k-1}\| dt \\ &\leq \int_0^1 L \|tv_{k-1}\| \|v_{k-1}\| dt \\ &\leq \frac{L}{2}(t_k - t_{k-1})^2 \\ &= h(t_{k-1}) + h'(t_{k-1})(t_k - t_{k-1}) + \frac{1}{2}h''(t_{k-1})(t_k - t_{k-1})^2 \\ &= h(t_k), \end{aligned} \tag{3.13}$$

where the first equality holds because $h(t_{k-1}) + h'(t_{k-1})(t_k - t_{k-1}) = 0$ by (3.3) and $h'' = L$. Combining this with (3.12) yields that

$$\begin{aligned} \|v_k\| &= \left\| -df_{x_k}^{-1}f(x_k) \right\| \\ &\leq \|df_{x_k}^{-1}df_{x_0}\| \|df_{x_0}^{-1}f(x_k)\| \\ &\leq -h'(t_k)^{-1}h(t_k) \\ &= t_{k+1} - t_k. \end{aligned} \tag{3.14}$$

Since $x_{k+1} = x_k \cdot \exp v_k$, we have $\varrho(x_{k+1}, x_k) \leq \|v_k\|$. This together with (3.14) gives that (3.10) holds for $n = k$, which completes the proof of the theorem. \square

The remainder of this section is devoted to an estimate of the convergence domain of Newton's method on G around a zero x^* of f . Below we shall always assume that $x^* \in G$ is such that $df_{x^*}^{-1}$ exists.

LEMMA 3.2 Let $0 < r \leq \frac{1}{L}$ and let $x_0 \in C_r(x^*)$ be such that there exist $j \geq 1$ and $w_1, \dots, w_j \in \mathcal{G}$ satisfying

$$x_0 = x^* \cdot \exp w_1 \cdots \exp w_j \tag{3.15}$$

and $\sum_{i=1}^j \|w_i\| < r$. Suppose that $df_{x^*}^{-1}df$ satisfies the L -Lipschitz condition on $C_r(x^*)$. Then $df_{x_0}^{-1}$ exists and

$$\|df_{x_0}^{-1}f(x_0)\| \leq \frac{(2 + L \sum_{i=1}^j \|w_i\|) \sum_{i=1}^j \|w_i\|}{2(1 - L \sum_{i=1}^j \|w_i\|)}. \quad (3.16)$$

Proof. By Lemma 2.2, $df_{x_0}^{-1}$ exists and

$$\|df_{x_0}^{-1}df_{x^*}\| \leq \frac{1}{1 - L \sum_{i=1}^j \|w_i\|}. \quad (3.17)$$

We write $y_0 = x^*$ and $y_i = y_{i-1} \cdot \exp w_i$ for each $i = 1, \dots, j$. Thus, by (3.15), we have $y_j = x_0$. Fixing i , one has from (2.13) that

$$f(y_i) - f(y_{i-1}) = \int_0^1 df_{y_{i-1} \cdot \exp(\tau w_i)} w_i d\tau,$$

which implies that

$$df_{x^*}^{-1}(f(y_i) - f(y_{i-1})) = \int_0^1 df_{x^*}^{-1}(df_{y_{i-1} \cdot \exp(\tau w_i)} - df_{x^*}) w_i d\tau + w_i. \quad (3.18)$$

Since $df_{x^*}^{-1}df$ satisfies the L -Lipschitz condition on $C_r(x^*)$, it follows that

$$\|df_{x^*}^{-1}(df_{y_{\kappa-1} \cdot \exp w_\kappa} - df_{y_{\kappa-1}})\| \leq L \|w_\kappa\| \quad \text{for each } \kappa = 1, \dots, j.$$

This gives that

$$\begin{aligned} \|df_{x^*}^{-1}(df_{y_{i-1} \cdot \exp(\tau w_i)} - df_{x^*})\| &\leq \sum_{\kappa=1}^{i-1} \|df_{x^*}^{-1}(df_{y_{\kappa-1} \cdot \exp w_\kappa} - df_{y_{\kappa-1}})\| \\ &\quad + \|df_{x^*}^{-1}(df_{y_{i-1} \cdot \exp(\tau w_i)} - df_{y_{i-1}})\| \\ &\leq L \sum_{\kappa=1}^{i-1} \|w_\kappa\| + L\tau \|w_i\|. \end{aligned} \quad (3.19)$$

Noting that $f(x_0) = \sum_{i=1}^j (f(y_i) - f(y_{i-1}))$, we have from (3.18) and (3.19) that

$$\begin{aligned} \|df_{x^*}^{-1}f(x_0)\| &\leq \sum_{i=1}^j \left(\int_0^1 \|df_{x^*}^{-1}(df_{y_{i-1} \cdot \exp(\tau w_i)} - df_{x^*})\| \|w_i\| d\tau + \|w_i\| \right) \\ &\leq \sum_{i=1}^j \left(\int_0^1 L \left(\sum_{\kappa=1}^{i-1} \|w_\kappa\| + \tau \|w_i\| \right) \|w_i\| d\tau + \|w_i\| \right) \\ &= \left(\frac{L}{2} \sum_{i=1}^j \|w_i\| + 1 \right) \sum_{i=1}^j \|w_i\|. \end{aligned}$$

Combing this with (3.17) yields that

$$\|df_{x_0}^{-1} f(x_0)\| \leq \|df_{x_0}^{-1} df_{x^*}\| \|df_{x^*}^{-1} f(x_0)\| \leq \frac{(2 + L \sum_{i=1}^j \|w_i\|) \sum_{i=1}^j \|w_i\|}{2(1 - L \sum_{i=1}^j \|w_i\|)},$$

which completes the proof of the lemma. □

THEOREM 3.3 Let $0 < r \leq \frac{1}{4L}$. Suppose that $f(x^*) = 0$ and that $df_{x^*}^{-1} df$ satisfies the L -Lipschitz condition on $C_{3r/(1-Lr)}(x^*)$. Let $x_0 \in C_r(x^*)$. Then the sequence $\{x_n\}$ generated by Newton's method (3.1) with initial point x_0 is well defined and converges quadratically to a zero y^* of f and $\varrho(x^*, y^*) < \frac{3r}{1-Lr}$.

Proof. Since $x_0 \in C_r(x^*)$, there exist $j \geq 1$ and $w_1, \dots, w_j \in \mathcal{G}$ satisfying

$$x_0 = x^* \cdot \exp w_1 \cdots \exp w_j$$

and $\sum_{i=1}^j \|w_i\| < r$. By Lemma 3.2, $df_{x_0}^{-1}$ exists and

$$\beta := \|df_{x_0}^{-1} f(x_0)\| \leq \frac{(2 + L \sum_{i=1}^j \|w_i\|) \sum_{i=1}^j \|w_i\|}{2(1 - L \sum_{i=1}^j \|w_i\|)}. \tag{3.20}$$

Set $\bar{L} = \frac{L}{1 - L \sum_{i=1}^j \|w_i\|}$ and $\bar{r} = \frac{(2+Lr)r}{1-Lr}$. Then $df_{x_0}^{-1} df$ satisfies the \bar{L} -Lipschitz condition on $C_{\bar{r}}(x_0)$.

To see this let $x \in C_{\bar{r}}(x_0)$ and $u \in \mathcal{G}$ be such that $\|u\| + \varrho(x_0, x) < \bar{r}$. Thus

$$\|u\| + \varrho(x, x^*) \leq \|u\| + \varrho(x, x_0) + \varrho(x_0, x^*) < \bar{r} + r \leq \frac{3r}{1 - Lr}.$$

Since $df_{x^*}^{-1} df$ satisfies the L -Lipschitz condition on $C_{3r/(1-Lr)}(x^*)$, it follows that

$$\|df_{x^*}^{-1}(df_{x \cdot \exp u} - df_x)\| \leq L\|u\|. \tag{3.21}$$

Consequently, by Lemma 2.2, one has that

$$\begin{aligned} \|df_{x_0}^{-1}(df_{x \cdot \exp u} - df_x)\| &\leq \|df_{x_0}^{-1} df_{x^*}\| \|df_{x^*}^{-1}(df_{x \cdot \exp u} - df_x)\| \\ &\leq \frac{L}{1 - L \sum_{i=1}^j \|w_i\|} \|u\| \\ &= \bar{L}\|u\|. \end{aligned}$$

This shows that $df_{x_0}^{-1} df$ satisfies the \bar{L} -Lipschitz condition on $C_{\bar{r}}(x_0)$. Let

$$\bar{\lambda} := \bar{L}\beta \quad \text{and} \quad \bar{r}_1 = \frac{1 - \sqrt{1 - 2\bar{\lambda}}}{\bar{L}} = \frac{2\beta}{1 + \sqrt{1 - 2\bar{\lambda}}}. \tag{3.22}$$

Then, by (3.20), we have

$$\bar{r}_1 \leq 2\beta \leq \frac{(2 + L \sum_{i=1}^j \|w_i\|) \sum_{i=1}^j \|w_i\|}{1 - L \sum_{i=1}^j \|w_i\|} \leq \bar{r}. \tag{3.23}$$

Moreover, since $L \sum_{i=1}^j \|w_i\| \leq Lr \leq \frac{1}{4}$, it follows that

$$\bar{\lambda} \leq \frac{(2 + L\|v\|)L\|v\|}{2(1 - L\|v\|)^2} \leq \frac{(2 + \frac{1}{4})\frac{1}{4}}{2(1 - \frac{1}{4})^2} = \frac{1}{2}. \quad (3.24)$$

Thus Theorem 3.1 is applicable and the sequence $\{x_n\}$ generated by Newton's method (3.1) with initial point x_0 converges to a zero, say y^* , of f and $\varrho(x_0, y^*) \leq \bar{r}_1$. Furthermore,

$$\varrho(x^*, y^*) \leq \varrho(x^*, x_0) + \varrho(x_0, y^*) < r + \bar{r}_1 \leq r + \bar{r} \leq \frac{3r}{1 - Lr}.$$

The proof of the theorem is complete. \square

In particular, taking $r = \frac{1}{4L}$ in Theorem 3.3 the following corollary is obvious.

COROLLARY 3.4 Suppose that $f(x^*) = 0$ and that $df_{x^*}^{-1}df$ satisfies the L -Lipschitz condition on $C_{1/L}(x^*)$. Let $x_0 \in C_{1/4L}(x^*)$. Then the sequence $\{x_n\}$ generated by Newton's method (3.1) with initial point x_0 is well defined and converges quadratically to a zero y^* of f with $\varrho(x^*, y^*) < \frac{1}{L}$.

Theorem 3.3 as well as Corollary 3.4 gives an estimate of the convergence domain for Newton's method. However, we do not know whether the limit y^* of the sequence generated by Newton's method with initial point x_0 from this domain is equal to the zero x^* . The following corollary provides the convergence domain from which the sequence generated by Newton's method with initial point x_0 converges to the zero x^* . Let $r > 0$. We use $\mathbf{B}(0, r)$ to denote the open ball at 0 with radius r on \mathcal{G} , that is,

$$\mathbf{B}(0, r) := \{v \in \mathcal{G} \mid \|v\| < r\}.$$

COROLLARY 3.5 Suppose that $f(x^*) = 0$ and that $df_{x^*}^{-1}df$ satisfies the L -Lipschitz condition on $C_{1/L}(x^*)$. Let $\rho > 0$ be the largest number such that $C_\rho(e) \subseteq \exp(\mathbf{B}(0, \frac{1}{L}))$ and let $r = \min\{\frac{\rho}{3+L\rho}, \frac{1}{4L}\}$. Let us write $N(x^*, r) := x^* \cdot \exp(\mathbf{B}(0, r))$. Then, for each $x_0 \in N(x^*, r)$, the sequence $\{x_n\}$ generated by Newton's method (3.1) with initial point x_0 is well defined and converges quadratically to x^* .

Proof. Let $x_0 \in N(x^*, r)$. Then there exists $v \in \mathcal{G}$ such that $x_0 = x^* \cdot \exp v$ and

$$\|v\| < r \leq \min\left\{\frac{\rho}{3 + L\rho}, \frac{1}{4L}\right\}.$$

This implies that $x_0 \in C_r(x^*)$ and

$$\frac{3r}{1 - Lr} \leq \min\left\{\frac{1}{L}, \rho\right\}. \quad (3.25)$$

Hence Theorem 3.3 can be applied to conclude 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, say y^* , of f and $\varrho(x^*, y^*) < \frac{3r}{1 - Lr}$. This together with (3.25) implies that $\varrho(y^*, x^*) < \rho$. Hence there exists $u \in \mathcal{G}$ such that $\|u\| < \frac{1}{L}$ and

$y^* = x^* \cdot \exp u$. Since

$$\begin{aligned} \left\| df_{x^*}^{-1} \int_0^1 df_{x^* \cdot \exp(\tau u)} d\tau - \mathbf{I} \right\| &= \left\| \int_0^1 df_{x^*}^{-1} (df_{x^* \cdot \exp(\tau u)} - df_{x^*}) d\tau \right\| \\ &\leq \int_0^1 L\tau \|u\| d\tau \\ &< 1, \end{aligned}$$

it follows from the Banach lemma that $df_{x^*}^{-1} \int_0^1 df_{x^* \cdot \exp(\tau u)} d\tau$ is invertible and so is $\int_0^1 df_{x^* \cdot \exp(\tau u)} d\tau$. Note that

$$\int_0^1 df_{x^* \cdot \exp(\tau u)} d\tau u = f(y^*) - f(x^*) = 0.$$

We obtain that $u = 0$ and $y^* = x^*$, which completes the proof of the corollary. □

Recall that, in the special case when G is a compact connected Lie group, G has a bi-invariant Riemannian metric (cf. DoCarmo, 1992, p. 46). Below we assume that G is a compact connected Lie group that is endowed with a bi-invariant Riemannian metric. Therefore an estimate of the convergence domain with the same property as in Corollary 3.5 is described in the following corollary.

COROLLARY 3.6 Let G be a compact connected Lie group that is endowed with a bi-invariant Riemannian metric. Let $0 < r \leq \frac{1}{4L}$. Suppose that $f(x^*) = 0$ and that $df_{x^*}^{-1} df$ satisfies the L -Lipschitz condition on $C_{3r/(1-Lr)}(x^*)$. Let $x_0 \in C_r(x^*)$. Then the sequence $\{x_n\}$ generated by Newton's method (3.1) with initial point x_0 is well defined and converges quadratically to x^* .

Proof. By Theorem 3.3, the sequence $\{x_n\}$ generated by Newton's method (3.1) with initial point x_0 is well defined and converges to a zero, say y^* , of f with $\varrho(x^*, y^*) < \frac{3r}{1-Lr}$. Clearly, there is a minimizing geodesic c connecting $x^{*-1} \cdot y^*$ and e . Since G is a compact connected Lie group that is endowed with a bi-invariant Riemannian metric, it follows from Helgason (1978, p. 224) that c is also a one-parameter subgroup of G . Consequently, there exists $u \in \mathcal{G}$ such that $y^* = x^* \cdot \exp u$ and

$$\|u\| = \varrho(x^*, y^*) < \frac{3r}{1-Lr}.$$

Since

$$\begin{aligned} \left\| df_{x^*}^{-1} \int_0^1 df_{x^* \cdot \exp(\tau u)} d\tau - \mathbf{I} \right\| &= \left\| \int_0^1 df_{x^*}^{-1} (df_{x^* \cdot \exp(\tau u)} - df_{x^*}) d\tau \right\| \\ &\leq \int_0^1 L\tau \|u\| d\tau \\ &< 1, \end{aligned}$$

it follows from the Banach lemma that $df_{x^*}^{-1} \int_0^1 df_{x^* \cdot \exp(\tau u)} d\tau$ is invertible and so is $\int_0^1 df_{x^* \cdot \exp(\tau u)} d\tau$. As

$$\int_0^1 df_{x^* \cdot \exp(\tau u)} d\tau u = f(y^*) - f(x^*) = 0,$$

we have that $u = 0$, and so $y^* = x^*$. This completes the proof of the corollary. □

In particular, taking $r = \frac{1}{4L}$ in Corollary 3.6, one has the following result.

COROLLARY 3.7 Let G be a compact connected Lie group that is endowed with a bi-invariant Riemannian metric. Suppose that $f(x^*) = 0$ and that $df_{x^*}^{-1}df$ satisfies the L -Lipschitz condition on $C_{1/L}(x^*)$. Let $x_0 \in C_{1/4L}(x^*)$. Then the sequence $\{x_n\}$ generated by Newton's method (3.1) with initial point x_0 is well defined and converges quadratically to x^* .

Clearly, Corollaries 3.5 and 3.7 improve the corresponding local convergence result in Owren & Welfert (2000, Theorem 4.6). The following example is devoted to an application of our results in this section to the initial value problem on the special linear group $SL(N, \mathbb{R})$ and the special orthogonal group $SO(N, \mathbb{R})$ that was considered by Owren & Welfert (2000). This, in particular, presents an example for which our results in the present paper are applicable but not the results in Mahony (1996) and Owren & Welfert (2000).

For the following example we define the second differential $d^2\theta_x: \mathcal{G}^2 \rightarrow \mathcal{G}$ for a C^2 -map $\theta: G \rightarrow \mathcal{G}$ as follows:

$$d^2\theta_x u_1 u_2 = \left(\frac{\partial^2}{\partial t_2 \partial t_1} \theta(x \cdot \exp t_2 u_2 \cdot \exp t_1 u_1) \right)_{t_2=t_1=0}. \quad (3.26)$$

Then, by (2.13), we have that

$$d\theta_{x \cdot \exp(tu)} - d\theta_x = \int_0^t d^2\theta_{x \cdot \exp(su)} u \, ds \quad \text{for each } u \in \mathcal{G} \text{ and } t \in \mathbb{R}. \quad (3.27)$$

EXAMPLE 3.8 Let N be a positive integer and let \mathbf{I}_N be the $N \times N$ identity matrix. Let G be the special linear group under standard matrix multiplication (cf. Varadarajan, 1984), that is,

$$G = SL(N, \mathbb{R}) := \{x \in \mathbb{R}^{N \times N} \mid \det x = 1\}.$$

Then G is a connected Lie group and its Lie algebra is

$$\mathcal{G} = T_e G = \mathfrak{sl}(N, \mathbb{R}) := \{v \in \mathbb{R}^{N \times N} \mid \text{tr}(v) = 0\},$$

where $e = \mathbf{I}_N$. We endow \mathcal{G} with the standard inner product

$$\langle u, v \rangle_e = \text{tr}(u^T v) \quad \text{for any } u, v \in \mathcal{G}. \quad (3.28)$$

Hence the corresponding norm $\|\cdot\|$ is the Frobenius norm. Moreover, the exponential map $\exp: \mathcal{G} \rightarrow G$ is given by

$$\exp(v) = \sum_{n \geq 0} \frac{v^n}{n!} \quad \text{for each } v \in \mathcal{G}$$

and its inverse is the logarithm (cf. Hall, 2004, p. 34)

$$\exp^{-1}(z) = \sum_{k \geq 1} (-1)^{k-1} \frac{(z - \mathbf{I}_N)^k}{k} \quad \text{for each } z \in G \text{ with } \|z - \mathbf{I}_N\| < 1. \quad (3.29)$$

Let $g: G \times \mathbb{R} \rightarrow \mathcal{G}$ be a differentiable map and let $x^{(0)}$ be a random starting point. Consider the following initial value problem on G studied in [Owren & Welfert \(2000\)](#):

$$\begin{cases} x' = x \cdot g(x), \\ x(0) = x^{(0)}. \end{cases} \tag{3.30}$$

The application of one step of the backward Euler method on (3.30) leads to the fixed-point problem

$$x = x^{(0)} \cdot \exp(lg(x)), \tag{3.31}$$

where l represents the size of the discretization step. Let $f: G \rightarrow \mathcal{G}$ be defined by

$$f(x) = \exp^{-1}((x^{(0)})^{-1} \cdot x) - lg(x) \quad \text{for each } x \in G.$$

Thus solving the equation (3.31) is equivalent to finding a zero of f .

Let g be the function considered in [Owren & Welfert \(2000\)](#) that is defined by

$$g(x) = (\sin x)(2x - 5x^2) - ((\sin x)(2x - 5x^2))^T \quad \text{for each } x \in G,$$

where

$$\sin x = \sum_{j \geq 1} (-1)^{j-1} \frac{(x)^{2j-1}}{(2j-1)!} \quad \text{for each } x \in G.$$

Consider the special case when $l = 1$ and $x^{(0)} = \exp v_0$ with $v_0 \in \mathcal{G}$ satisfying $\|v_0\| \leq \frac{1}{11}$. To apply our results we have to estimate the norm of df_x^{-1} . To do this we write $w(\cdot) = \exp^{-1}((x^{(0)})^{-1} \cdot (\cdot))$. Let $x := \exp v_1 \exp v_2 \cdots \exp v_k$ for some $v_1, v_2, \dots, v_k \in \mathcal{G}$ with $\sum_{i=0}^k \|v_i\| < \frac{1}{4}$. Since

$$e^t - 1 \leq \frac{5}{4}t \quad \text{for each } t \in \left[0, \frac{1}{4}\right], \tag{3.32}$$

one can use mathematical induction to prove that

$$\|x - \mathbf{I}_N\| \leq e^{\sum_{i=1}^k \|v_i\|} - 1 \leq \frac{5}{4} \sum_{i=1}^k \|v_i\|. \tag{3.33}$$

Consequently, we have

$$\|(x^{(0)})^{-1}x - \mathbf{I}_N\| \leq \frac{5}{4} \sum_{i=0}^k \|v_i\| < \frac{5}{16} < 1. \tag{3.34}$$

Thus, by definition and using (3.29), one has for each $u \in \mathcal{G}$ that

$$dw_x(u) = \sum_{j \geq 1} (-1)^{j-1} \frac{\sum_{i=0}^{j-1} ((x^{(0)})^{-1}x - \mathbf{I}_N)^i (x^{(0)})^{-1}xu((x^{(0)})^{-1}x - \mathbf{I}_N)^{j-1-i}}{j}. \tag{3.35}$$

Since

$$\|(x^{(0)})^{-1}xu\| = \|(x^{(0)})^{-1}xu - u + u\| \leq (\|(x^{(0)})^{-1}x - \mathbf{I}_N\| + 1)\|u\|,$$

it follows from (3.35) that

$$\begin{aligned} \|dw_x(u) - u\| &\leq \left(\sum_{j \geq 2} \frac{j(\|(x^{(0)})^{-1}x - \mathbf{I}_N\| + 1)\|(x^{(0)})^{-1}x - \mathbf{I}_N\|^{j-1}}{j} + \|(x^{(0)})^{-1}x - \mathbf{I}_N\| \right) \|u\| \\ &= \frac{2\|(x^{(0)})^{-1}x - \mathbf{I}_N\|}{1 - \|(x^{(0)})^{-1}x - \mathbf{I}_N\|} \|u\|. \end{aligned}$$

Hence, by (3.34), we have

$$\|dw_x - \mathbb{I}_{\mathcal{G}}\| \leq \frac{2\|(x^{(0)})^{-1}x - \mathbf{I}_N\|}{1 - \|(x^{(0)})^{-1}x - \mathbf{I}_N\|} \leq \frac{10 \sum_{i=0}^k \|v_i\|}{4 - 5 \sum_{i=0}^k \|v_i\|}, \quad (3.36)$$

where $\mathbb{I}_{\mathcal{G}}$ is the identity on \mathcal{G} . Similarly, one has

$$\begin{aligned} \|d^2w_x\| &\leq \sum_{j \geq 1} [(\|(x^{(0)})^{-1}x - \mathbf{I}_N\| + 1)\|(x^{(0)})^{-1}x - \mathbf{I}_N\|^{j-1} \\ &\quad + (j-1)(\|(x^{(0)})^{-1}x - \mathbf{I}_N\| + 1)^2\|(x^{(0)})^{-1}x - \mathbf{I}_N\|^{j-2}] \\ &= \frac{3 + \|(x^{(0)})^{-1}x - \mathbf{I}_N\|}{(1 - \|(x^{(0)})^{-1}x - \mathbf{I}_N\|)^2}. \end{aligned}$$

Combining this and (3.34) gives that

$$\|d^2w_x\| \leq \frac{3 + \frac{5}{4}(\sum_{i=0}^k \|v_i\|)}{(1 - \frac{5}{4}(\sum_{i=0}^k \|v_i\|))^2}. \quad (3.37)$$

Let $x_0 = \mathbf{I}_N$ and $L = 51$. Below we verify that $df_{x_0}^{-1}df$ satisfies the L -Lipschitz condition at x_0 on $C_{1/50}(x_0)$. For this purpose we let $v \in \mathcal{G}$ and $x \in C_{1/50}(x_0)$. Then there exist $k \geq 1$ and $v_1, \dots, v_k \in \mathcal{G}$ such that $x = x_0 \cdot \exp v_1 \cdots \exp v_k$ and $\|v\| + \sum_{i=1}^k \|v_i\| < \frac{1}{50}$. Let $s \in [0, 1]$ and write $y := x \cdot \exp sv = \exp v_1 \cdots \exp v_k \cdot \exp sv$. Note that, by (3.36), we have that

$$\|dw_{x_0} - \mathbb{I}_{\mathcal{G}}\| \leq \frac{10\|v_0\|}{4 - 5\|v_0\|} \leq \frac{2}{7} \quad (3.38)$$

because $\|v_0\| \leq \frac{1}{11} < \frac{1}{10}$. Since $\sum_{i=0}^k \|v_i\| + \|v\| \leq \frac{1}{10} + \sum_{i=1}^k \|v_i\| + \|v\| < \frac{3}{25}$, it follows from (3.37) that

$$\|d^2w_y\| \leq \frac{3 + \frac{3}{20}}{(1 - \frac{5}{4}(\frac{1}{10} + \sum_{i=1}^k \|v_i\| + s\|v\|))^2} \leq \frac{144}{35(1 - \frac{10}{7}(\sum_{i=1}^k \|v_i\| + s\|v\|))^2}. \quad (3.39)$$

On the other hand, note that

$$g(x) = 2 \sum_{j \geq 1} (-1)^{j-1} \frac{x^{2j} - (x^{2j})^T}{(2j-1)!} - 5 \sum_{j \geq 1} (-1)^{j-1} \frac{x^{2j+1} - (x^{2j+1})^T}{(2j-1)!}. \quad (3.40)$$

Then, by definition, $dg_{x_0} = (-6 \cos 1 - 16 \sin 1)\mathbb{I}_G$. Hence

$$dg_{x_0}^{-1} = \frac{1}{-6 \cos 1 - 16 \sin 1} \mathbb{I}_G \quad \text{and} \quad (\mathbb{I}_G - dg_{x_0})^{-1} = \frac{1}{1 + 6 \cos 1 + 16 \sin 1} \mathbb{I}_G. \tag{3.41}$$

It follows from (3.38) that

$$\begin{aligned} \|(\mathbb{I}_G - dg_{x_0})^{-1}\| \|df_{x_0} - (\mathbb{I}_G - dg_{x_0})\| &= \|(\mathbb{I}_G - dg_{x_0})^{-1}\| \|dw_{x_0} - \mathbb{I}_G\| \\ &\leq \frac{1}{1 + 6 \cos 1 + 16 \sin 1} \cdot \frac{2}{7} < 1. \end{aligned}$$

Thus the Banach lemma is applied to conclude that

$$\|df_{x_0}^{-1}\| \leq \frac{1/1 + 6 \cos 1 + 16 \sin 1}{1 - (1/1 + 6 \cos 1 + 16 \sin 1) \cdot \frac{2}{7}} \leq \frac{1}{10}. \tag{3.42}$$

By definition, we have

$$\|d^2g_y\| \leq 2 \sum_{j \geq 1} 2 \frac{(2j)^2 \|y^{2j}\|}{(2j-1)!} + 5 \sum_{j \geq 1} 2 \frac{(2j+1)^2 \|y^{2j+1}\|}{(2j-1)!}. \tag{3.43}$$

Using (3.33), we have that $\|y^i\| \leq \frac{5}{4}i \left(\sum_{i=1}^k \|v_i\| + s\|v\|\right) + 1$ for each $i \geq 1$ and it follows from (3.43) that

$$\|d^2g_y\| \leq \frac{5}{4} \left(\sum_{i=1}^k \|v_i\| + s\|v\| \right) \sum_{j \geq 1} \frac{4(2j)^3 + 10(2j+1)^3}{(2m-1)!} + \sum_{j \geq 1} \frac{4(2j)^2 + 10(2j+1)^2}{(2j-1)!}. \tag{3.44}$$

Note that for each $j \geq 1$ we have

$$\frac{2(2j)^i}{(2j-1)!} \leq \frac{(2j-1+i) \cdots (2j+1)2j}{(2j-1)!} + \frac{(2j-2+i) \cdots 2j(2j-1)}{(2j-2)!} \quad \text{for } i = 2, 3.$$

Then, by elementary calculations, we have that

$$2 \sum_{j \geq 1} \frac{(2j)^i}{(2j-1)!} \leq \sum_{l \geq 0} \frac{(l+i) \cdots (l+1)}{l!} = i!2^i e \quad \text{for } i = 2, 3.$$

Similarly,

$$2 \sum_{j \geq 1} \frac{(2j+1)^i}{(2j-1)!} \leq (i+1)!2^i e \quad \text{for } i = 2, 3.$$

Combining (3.44) with the above two inequalities gives the following estimate:

$$\|d^2g_y\| \leq 1320e \left(\sum_{i=1}^k \|v_i\| + s\|v\| \right) + 136e. \tag{3.45}$$

This together with (3.39) and (3.42) yields that

$$\begin{aligned} \left\| \mathrm{d}f_{x_0}^{-1} \mathrm{d}^2 f_y \right\| &\leq \left\| \mathrm{d}f_{x_0}^{-1} \right\| \left\| \mathrm{d}^2 w_y \right\| + \left\| \mathrm{d}f_{x_0}^{-1} \mathrm{d}^2 g_y \right\| \\ &\leq \frac{72}{175 \left(1 - \frac{10}{7} \left(\sum_{i=1}^k \|v_i\| + s \|v\| \right) \right)^2} + \frac{e}{10} \left(1320 \left(\sum_{i=1}^k \|v_i\| + s \|v\| \right) + 136 \right). \end{aligned} \quad (3.46)$$

Then it follows from (3.27) and the fact that $y = x \cdot \exp sv$ that

$$\begin{aligned} &\left\| \mathrm{d}f_{x_0}^{-1} (\mathrm{d}f_{x \cdot \exp v} - \mathrm{d}f_x) \right\| \\ &\leq \int_0^1 \left\| \mathrm{d}f_{x_0}^{-1} \mathrm{d}^2 f_{x \cdot \exp sv} \right\| \|v\| \mathrm{d}s \\ &\leq \int_0^1 \left(\frac{72}{175 \left(1 - \frac{10}{7} \left(\sum_{i=1}^k \|v_i\| + s \|v\| \right) \right)^2} + \frac{e}{10} \left(1320 \left(\sum_{i=1}^k \|v_i\| + s \|v\| \right) + 136 \right) \right) \|v\| \mathrm{d}s \\ &\leq 51 \|v\|. \end{aligned} \quad (3.47)$$

Thus the claim stands. Moreover, $r_1 < \frac{1}{L} < \frac{1}{50}$, and so that $\mathrm{d}f_{x_0}^{-1} \mathrm{d}f$ satisfies the L -Lipschitz condition at x_0 on $C_{r_1}(x_0)$. Noting that $f(x_0) = -v_0$, we have by (3.42) that

$$\lambda = L\beta = L \left\| \mathrm{d}f_{x_0}^{-1} f(x_0) \right\| \leq L \left\| \mathrm{d}f_{x_0}^{-1} \right\| \|v_0\| \leq 51 \cdot \frac{1}{10} \cdot \|v_0\| < \frac{1}{2}.$$

Thus Theorem 3.1 is applied to conclude that the sequence generated by (3.1) with initial point $x_0 = \mathbf{I}_N$ converges to a zero x^* of f .

To illustrate an application of Corollary 3.4 let us take $x^{(0)} = \mathbf{I}_N$, that is, $f: G \rightarrow \mathcal{G}$ is defined by

$$f(x) = \exp^{-1}(x) - (\sin x)(2x - 10x^2) + ((\sin x)(2x - 10x^2))^T \quad \text{for each } x \in G.$$

Then $x^* := \mathbf{I}_N$ is a zero of f . Furthermore, by (3.35), we have

$$\mathrm{d}w_{x^*} = \mathbb{I}_{\mathcal{G}} \quad \text{and} \quad \mathrm{d}f_{x^*} = \mathrm{d}w_{x^*} - \mathrm{d}g_{x^*} = (1 + 6 \cos 1 + 16 \sin 1) \mathbb{I}_{\mathcal{G}}. \quad (3.48)$$

As before, let $v \in \mathcal{G}$ and $x \in C_{1/50}(x^*)$. Then there exist $k \geq 1$ and $v_1, v_2, \dots, v_k \in \mathcal{G}$ such that $x = \exp v_1 \exp v_2 \cdots \exp v_k$ and $\sum_{i=1}^k \|v_i\| + \|v\| < \frac{1}{50}$. Similarly, let $s \in [0, 1]$ and write $y := x \exp sv = \exp v_1 \exp v_2 \cdots \exp v_k \exp sv$. Note that, by (3.37) (as $v_0 = 0$), one has that

$$\left\| \mathrm{d}^2 w_y \right\| \leq \frac{3 + \frac{1}{40}}{\left(1 - \frac{5}{4} \left(\sum_{i=1}^k \|v_i\| + s \|v\| \right) \right)^2}. \quad (3.49)$$

Note that $x^* = x_0 = \mathbf{I}_N$. Then, using (3.43)–(3.45) and (3.49), one can verify (with almost the same arguments as we did for (3.46) and (3.47)) that

$$\left\| \mathrm{d}f_{x^*}^{-1} \mathrm{d}^2 f_y \right\| \leq \frac{121}{480 \left(1 - \frac{5}{4} \left(\sum_{i=1}^k \|v_i\| + s \|v\| \right) \right)^2} + \frac{e}{12} \left(1320 \left(\sum_{i=1}^k \|v_i\| + s \|v\| \right) + 136 \right)$$

and

$$\|df_{x^*}^{-1}(df_{x \cdot \exp v} - df_x)\| \leq 51\|v\|,$$

that is, $df_{x^*}^{-1}df$ satisfies the L -Lipschitz condition at x^* on $C_{1/50}(x^*)$ and so on $C_{1/L}(x^*)$ because $\frac{1}{L} < \frac{1}{50}$. Take $x_0 = x^* \cdot \exp v$ with $v \in \mathcal{G}$ and $\|v\| < \frac{1}{204}$. Corollary 3.4 is applied to conclude that the sequence generated by (3.1) with initial point x_0 is well defined and converges quadratically to a zero y^* of f in $C_{1/L}(x^*)$.

Furthermore, if we take G to be the special orthogonal group under standard matrix multiplication, that is,

$$G = \text{SO}(N, \mathbb{R}) := \{x \in \mathbb{R}^{N \times N} \mid x^T x = \mathbf{I}_N \text{ and } \det x = 1\}. \tag{3.50}$$

Then G is a compact connected Lie group and its Lie algebra is the set of all $N \times N$ skew-symmetric matrices, that is,

$$\mathcal{G} = \mathfrak{so}(N, \mathbb{R}) := \{v \in \mathbb{R}^{N \times N} \mid v^T + v = 0\}. \tag{3.51}$$

Note that $[u, v] = uv - vu$ and $\langle [u, v], w \rangle_e = -\langle u, [w, v] \rangle_e$ for any $u, v, w \in \mathcal{G}$. One can easily verify (cf. DoCarmo, 1992, p. 41) that the left-invariant Riemannian metric induced by the inner product in (3.28) is a bi-invariant metric on G . Then Corollary 3.7 is applicable and the sequence generated by (3.1) with initial point x_0 is well defined and converges quadratically to x^* .

We end this section with a remark.

REMARK 3.9 It would be helpful to make some comparisons of Theorem 3.1 with the corresponding result given by theorem 2 in Wang & Li (2007), where the convergence criterion (3.7) was provided under the following metric L -Lipschitz condition at x_0 :

$$\left\| df_{x_0}^{-1}(df_{x'} - df_x) \right\| \leq Ld(x', x) \quad \forall x', x \in G \text{ with } d(x_0, x) + d(x, x') < r_1, \tag{3.52}$$

where $d(\cdot, \cdot)$ is the Riemannian distance induced by the left-invariant Riemannian metric. Clearly, this kind of metric L -Lipschitz condition is dependent on the metric on G and is much stronger than the L -Lipschitz condition on $C_{r_1}(x_0)$ given in Definition 2.1:

$$\left\| df_{x_0}^{-1}(df_{x \cdot \exp u} - df_x) \right\| \leq L\|u\| \quad \forall x \in C_{r_1}(x_0) \text{ and } u \in \mathcal{G} \text{ with } \varrho(x, x_0) + \|u\| < r_1. \tag{3.53}$$

Usually, in a noncompact Lie group, (3.52) is difficult to verify, in particular, for the points x and x' that no one-parameter semigroups connect because $df_{(\cdot)}$ contains no information about the distance between the two points in G . Hence it is not convenient to apply Theorem 2 of Wang & Li (2007). For example, consider the group $G = \text{SL}(N, \mathbb{R})$ in Example 3.8. It would be very difficult to verify the metric L -Lipschitz condition (3.52) (in fact, we do not know how to do that).

4. Applications to optimization problems

Let $\phi: G \rightarrow \mathbb{R}$ 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 Mahony (1996), where a local quadratic convergence result was established for a smooth function ϕ .

Let $X \in \mathcal{G}$. Following Mahony (1996), we use \tilde{X} to denote the left-invariant vector field associated with X defined by

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

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

$$(\tilde{X}\phi)(x) = \left. \frac{d}{dt} \right|_{t=0} \phi(x \cdot \exp tX). \quad (4.2)$$

Let $\{X_1, \dots, X_n\}$ be an orthonormal basis of \mathcal{G} . According to Helmke & Moore (1994, p. 356) (see also Mahony, 1996), $\text{grad } \phi$ is a vector field on G defined by

$$\text{grad } \phi(x) = (\tilde{X}_1, \dots, \tilde{X}_n)(\tilde{X}_1\phi(x), \dots, \tilde{X}_n\phi(x))^T = \sum_{j=1}^n \tilde{X}_j\phi(x)\tilde{X}_j \quad \text{for each } x \in G. \quad (4.3)$$

Then Newton's method with initial point $x_0 \in G$ that was considered in Mahony (1996) can be written in a coordinate-free form as follows.

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

$$\text{grad } \phi(x_k) + \text{grad } (\tilde{X}^k\phi)(x_k) = 0.$$

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

Set $k \leftarrow k + 1$ and repeat.

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

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

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

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

Then $H_{(\cdot)}\phi$ defines a mapping from G to $\mathcal{L}(\mathcal{G})$. The following proposition gives the equivalence between df_x and $H_x\phi$.

PROPOSITION 4.2 Let $f(\cdot)$ and $H_{(\cdot)}\phi$ be defined by (4.4) and (4.5), respectively. Then

$$df_x = H_x\phi \quad \text{for each } x \in G. \quad (4.6)$$

Proof. Let $x \in G$ and let $\{X_1, \dots, X_n\}$ be an orthonormal basis of \mathcal{G} . In view of (4.3) and (4.4), we have that

$$f(x) = (X_1, \dots, X_n)((\tilde{X}_1\phi)(x), \dots, (\tilde{X}_n\phi)(x))^T. \quad (4.7)$$

Since ϕ is a C^2 -mapping, it is easy to see by definition that

$$\tilde{X}_j(\tilde{X}\phi)(x) = \tilde{X}(\tilde{X}_j\phi)(x) \quad \text{for each } X \in \mathcal{G} \text{ and } j = 1, 2, \dots, n. \quad (4.8)$$

Therefore, by (4.7), we have that

$$\begin{aligned} df_x(X) &= \left. \frac{d}{dt} \right|_{t=0} f(x \cdot \exp tX) \\ &= (X_1, \dots, X_n) \left(\left. \frac{d}{dt} \right|_{t=0} (\tilde{X}_1\phi)(x \cdot \exp tX), \dots, \left. \frac{d}{dt} \right|_{t=0} (\tilde{X}_n\phi)(x \cdot \exp tX) \right)^T \\ &= (X_1, \dots, X_n) (\tilde{X}(\tilde{X}_1\phi)(x), \dots, \tilde{X}(\tilde{X}_n\phi)(x))^T \\ &= (X_1, \dots, X_n) (\tilde{X}_1(\tilde{X}\phi)(x), \dots, \tilde{X}_n(\tilde{X}\phi)(x))^T \\ &= (L'_x)^{-1} (\tilde{X}_1, \dots, \tilde{X}_n) (\tilde{X}_1(\tilde{X}\phi)(x), \dots, \tilde{X}_n(\tilde{X}\phi)(x))^T, \end{aligned}$$

where the fourth equality holds because of (4.8). This means that (4.6) holds and the proof is complete. \square

REMARK 4.3 One can easily see from Proposition 4.2 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 $\beta_\phi := \|(H_{x_0}\phi)^{-1}(L'_{x_0})^{-1} \text{grad } \phi(x_0)\|$. Recall that r_1 is defined by (3.4). Then the main theorem of this section is as follows.

THEOREM 4.4 Suppose that

$$\lambda := L\beta_\phi \leq \frac{1}{2} \tag{4.9}$$

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 at which $\text{grad } \phi(x^*) = 0$.

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

$$\|(H_{x_0}\phi)^{-1}\| \|H_{x \cdot \exp u}\phi - H_x\phi\| \leq L\|u\| \quad \text{for } x \in G \text{ and } u \in \mathcal{G} \text{ with } \varrho(x_0, x) + \|u\| < r_1, \tag{4.10}$$

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

Proof. Recall that f is defined by (4.4). Then, by Proposition 4.2, $df_x = H_x\phi$ for each $x \in G$. Hence, by the assumptions, $df_{x_0}^{-1}df$ satisfies the L -Lipschitz condition on $C_{r_1}(x_0)$ and condition (3.7) is satisfied because $L\beta = L\beta_\phi \leq \frac{1}{2}$. 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 4.3, 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 (4.10) 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 we 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 4.3,

$$x_{n+1} = x_n \cdot \exp(-v_n) \quad \text{for each } n = 0, 1, \dots, \tag{4.11}$$

and, by Theorem 3.1, we have

$$\|v_n\| \leq t_{n+1} - t_n \quad \text{for each } n = 0, 1, \dots \tag{4.12}$$

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

$$\begin{aligned} \|H_{x_0}\phi^{-1}\| \| (H_{x_{n+1}}\phi - H_{x_0}\phi) \| &= \|H_{x_0}\phi^{-1}\| \| (H_{x_n \cdot \exp(-v_n)}\phi - H_{x_0}\phi) \| \\ &= \sum_{j=0}^n \|H_{x_0}\phi^{-1}\| \| H_{x_j \cdot \exp(-v_n)}\phi - H_{x_j}\phi \| \\ &\leq \sum_{j=0}^n L \|v_n\| \\ &\leq \sum_{j=0}^n L(t_{n+1} - t_n) \\ &\leq Lr_1 \end{aligned} \tag{4.13}$$

due to (4.10)–(4.12). Since

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

it follows that

$$\left| \frac{\lambda^*}{\lambda^0} - 1 \right| \leq \lim_{n \rightarrow \infty} \|H_{x_0}\phi^{-1}\| \|H_{x_{n+1}}\phi - H_{x_0}\phi\| \leq Lr_1 < 1$$

due to (4.13). This implies that $\lambda^* > 0$ and completes the proof. □

Similar to the proof of Theorem 4.4, we can verify the following corollaries.

COROLLARY 4.5 Let x^* be a local optimal solution of (4.1) such that $(H_{x^*}\phi)^{-1}$ exists. Suppose that $(H_{x^*}\phi)^{-1}(H_{(\cdot)}\phi)$ satisfies the L -Lipschitz condition on $C_{1/L}(x_0)$. Let $x_0 \in C_{1/4L}(x^*)$. Then the sequence generated by Algorithm 4.1 with initial point x_0 is well defined and converges to a critical point, say y^* , of ϕ and $\varrho(x^*, y^*) < \frac{1}{L}$.

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

$$\|(H_{x^*}\phi)^{-1}\| \|H_{x \cdot \exp u}\phi - H_x\phi\| \leq L\|u\| \quad \text{for } x \in G \text{ and } u \in \mathcal{G} \text{ with } \varrho(x^*, x) + \|u\| < \frac{1}{L}, \tag{4.14}$$

then y^* is also a local solution of (4.1).

COROLLARY 4.6 Let x^* be a local optimal solution of (4.1) such that $(H_{x^*}\phi)^{-1}$ exists. Suppose that $(H_{x^*}\phi)^{-1}(H_{(\cdot)}\phi)$ satisfies the L -Lipschitz condition on $C_{1/L}(x_0)$. Let $\rho > 0$ be the largest number such that $C_\rho(e) \subseteq \exp(\mathbf{B}(0, \frac{1}{L}))$ and let $r = \min\{\frac{\rho}{3+L\rho}, \frac{1}{4L}\}$. Write $N(x^*, r) := x^* \cdot \exp(\mathbf{B}(0, r))$. Then, for each $x_0 \in N(x^*, r)$, the sequence generated by Algorithm 4.1 with initial point x_0 is well defined and converges quadratically to x^* .

COROLLARY 4.7 Let G be a compact connected Lie group that is endowed with a bi-invariant Riemannian metric. Let x^* be a local optimal solution of (4.1) such that $(H_{x^*}\phi)^{-1}$ exists. Suppose that $(H_{x^*}\phi)^{-1}(H_{(\cdot)}\phi)$ satisfies the L -Lipschitz condition on $C_{1/L}(x_0)$. Let $x_0 \in C_{1/4L}(x^*)$. Then the sequence generated by Algorithm 4.1 with initial point x_0 is well defined and converges quadratically to x^* .

We end this paper with two examples with ϕ defined by (1.1) for which our results in the present paper are applicable but not the results in Mahony (1996).

EXAMPLE 4.8 Let G and \mathcal{G} be given by (3.50) and (3.51), respectively, with $N = 3$. Consider the optimization problem with $\phi: G \rightarrow \mathbb{R}$ given by

$$\phi(x) = -\text{tr}(x^T Qx D) \quad \text{for each } x \in G,$$

where D is the diagonal matrix with diagonal entries 1, 2 and 3 and Q is a fixed symmetric matrix. Then it is known (see, for example, Smith, 1993, 1994; Mahony, 1996) that

$$\text{grad } \phi(x) = -(L'_x)_e[x^T Qx, D] = -x[x^T Qx, D] \tag{4.15}$$

and

$$\tilde{X}\phi(x) = -\text{tr}(x^T Qx [D, X^T]). \tag{4.16}$$

Therefore, in view of (4.15), we have

$$\text{grad } (\tilde{X}\phi)(x) = -(L'_x)_e[x^T Qx, [D, X^T]] = -x[x^T Qx, [D, X^T]].$$

This implies that

$$(H_x\phi)X = (L'_x)_e^{-1} \text{grad } (\tilde{X}\phi)(x) = -[x^T Qx, [D, X^T]]. \tag{4.17}$$

Fix $X \in \mathcal{G}$ and define the map $g: G \rightarrow \mathcal{G}$ by

$$g(x) := (H_x\phi)X = -[x^T Qx, [D, X^T]] \quad \text{for each } x \in G. \tag{4.18}$$

Then, by definition (cf. (2.10)), for each $s \in [0, 1]$ and $u \in \mathcal{G}$ we have

$$\begin{aligned} \text{d}g_{x \cdot \exp su} u &= \left. \frac{d}{dt} \right|_{t=0} - [(x e^{su} e^{tu})^T Qx e^{su} e^{tu}, [D, X^T]] \\ &= - \left[\left. \frac{d}{dt} \right|_{t=0} e^{-tu} (x e^{su})^T Qx e^{su} e^{tu}, [D, X^T] \right] \\ &= - \left[\left. \frac{d}{dt} \right|_{t=0} \text{Ad}_{e^{-tu}} ((x e^{su})^T Qx e^{su}), [D, X^T] \right] \\ &= -[[-u, (x e^{su})^T Qx e^{su}], [D, X^T]], \end{aligned}$$

where

$$\text{Ad}_{e^{Xt}}(Y) = Y + t[X, Y] + \frac{t^2}{2!}[X, [X, Y]] + \dots \quad \text{for each } X, Y \in \mathcal{G}. \tag{4.19}$$

Remember that the norm $\|\cdot\|$ is the Frobenius norm. Consequently, for each $u \in \mathcal{G}$ we have

$$\|dg_{x \cdot \exp su}u\| \leq 8\|u\|\|Q\|\|D\|\|X\| = 8\sqrt{14}\|u\|\|Q\|\|X\|.$$

Hence, applying (2.13), we have that

$$\|g(x \cdot \exp u) - g(x)\| \leq \int_0^1 \|dg_{x \cdot \exp(su)}u\| ds \leq 8\sqrt{14}\|u\|\|Q\|\|X\|.$$

This together with (4.18) implies that

$$\|H_{x \cdot \exp u}\phi - H_x\phi\| \leq 8\sqrt{14}\|u\|\|Q\| \quad \text{for each } u \in \mathcal{G}. \quad (4.20)$$

In particular, taking

$$Q = \begin{pmatrix} 1 & 0.003 & 0 \\ 0.003 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

we have

$$\|H_{x \cdot \exp u}\phi - H_x\phi\| \leq 16\sqrt{21}\|u\| \quad \text{for each } u \in \mathcal{G}. \quad (4.21)$$

Let

$$b_1 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad b_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad b_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}. \quad (4.22)$$

Then $\{b_1, b_2, b_3\}$ is a basis of $\mathfrak{so}(3, \mathbb{R})$. Thus we can endow the 2-norm $\|\cdot\|_2$ on $\mathfrak{so}(3, \mathbb{R})$ defined by

$$\|u\|_2 = \sqrt{u_1^2 + u_2^2 + u_3^2} \quad \text{for each } u = u_1b_1 + u_2b_2 + u_3b_3 \in \mathfrak{so}(3, \mathbb{R}).$$

It is routine to check that $\|u\| = \sqrt{2}\|u\|_2$ for each $u \in \mathfrak{so}(3, \mathbb{R})$. Let $x_0 = \mathbf{I}_3$ and let

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & -0.003 \\ 0 & -0.006 & -2 \end{pmatrix}.$$

Then, by (4.17), we have that

$$(H_{x_0}\phi)(b_1, b_2, b_3) = (b_1, b_2, b_3)C,$$

and so

$$(H_{x_0}\phi)^{-1}(b_1, b_2, b_3) = (b_1, b_2, b_3)C^{-1}. \quad (4.23)$$

Therefore for any $u = u_1b_1 + u_2b_2 + u_3b_3 \in \mathfrak{so}(3, \mathbb{R})$ we have

$$\begin{aligned} \|(H_{x_0}\phi)^{-1}u\| &= \|(b_1, b_2, b_3)C^{-1}(u_1, u_2, u_3)^T\| \\ &= \sqrt{2}\|C^{-1}(u_1, u_2, u_3)^T\|_2 \\ &\leq \sqrt{2}\|C^{-1}\| \|u\|_2 \\ &= \|C^{-1}\| \|u\|. \end{aligned}$$

Consequently,

$$\|(H_{x_0}\phi)^{-1}\| \leq \|C^{-1}\| \leq \sqrt{2}. \tag{4.24}$$

Combining this with (4.21) yields that $(H_{x_0}\phi)^{-1}(H_{(\cdot)}\phi)$ satisfies the L -Lipschitz condition with $L = 16\sqrt{42}$. Furthermore, by (4.15) and (4.23), we have

$$\beta_\phi = \|(H_{x_0}\phi)^{-1}(L'_{x_0})^{-1} \text{grad } \phi(x_0)\| = \|(H_{x_0}\phi)^{-1}[Q, D]\| = 0.003\sqrt{2}.$$

Hence $\lambda = L\beta_\phi < \frac{1}{2}$. Thus Theorem 4.4 is applicable and the sequence generated by Algorithm 4.1 with initial point $x_0 = \mathbf{I}_3$ is well defined and convergent to a critical point x^* of ϕ .

EXAMPLE 4.9 Consider the same problem as in Example 4.8 but with Q defined by

$$Q = \begin{pmatrix} 2 & -2 & 0 \\ -2 & 1 & -2 \\ 0 & -2 & 0 \end{pmatrix}.$$

Let

$$x^* = \frac{1}{3} \begin{pmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 2 & 2 & 1 \end{pmatrix}. \tag{4.25}$$

Since $x^{*T}Qx^* = \text{diag}(-2, 1, 4)$, it follows that x^* is a local optimal solution of ϕ (cf. Brockett, 1988, 1991; Mahony, 1996). Let $x \in G$ and $u \in \mathcal{G}$. Then it follows from (4.20) that

$$\|H_{x \cdot \exp u}\phi - H_x\phi\| \leq 8\sqrt{14}\|u\|\|Q\| = 56\sqrt{6}\|u\|. \tag{4.26}$$

Let b_1, b_2 and b_3 be given by (4.22) and let

$$C = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 12 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

With the same arguments as we had in Example 4.8, we can show that

$$(H_{x^*}\phi)(b_1, b_2, b_3) = (b_1, b_2, b_3)C$$

and

$$\|(H_{x^*}\phi)^{-1}\| \leq \|C^{-1}\| = \frac{\sqrt{33}}{12}. \tag{4.27}$$

Hence $H_{x^*}\phi$ is positive definite and $(H_{x^*}\phi)^{-1}(H_{(\cdot)}\phi)$ satisfies the L -Lipschitz condition with $L = 14\sqrt{22}$ due to (4.26) and (4.27). Now let $x_0 = x^* \cdot \exp v$ with x^* given by (4.25) and $v = 0.00269b_1$. Then $\|v\| < \frac{1}{4L}$. Note that G is compact connected and endowed with a bi-invariant Riemannian metric. Hence Corollary 4.7 is applicable and the sequence generated by Algorithm 4.1 with initial point $x_0 = x^* \cdot \exp v$ is well defined and converges quadratically to x^* .

Funding

This research was partially supported by the National Natural Science Foundation of China (grants 10671175; 10731060).

REFERENCES

- ABSIL, P. A., BAKER, C. G. & GALLIVAN, K. A. (2007) Trust-region methods on Riemannian manifolds. *Found. Comput. Math.*, **7**, 303–330.
- ADLER, R., DEDIEU, J. P., MARGULIES, J., MARTENS, M. & SHUB, M. (2002) Newton method on Riemannian manifolds and a geometric model for human spine. *IMA J. Numer. Anal.*, **22**, 1–32.
- ALVAREZ, F., BOLTE, J. & MUNIER, J. (2008) A unifying local convergence result for Newton's method in Riemannian manifolds. *Found. Comput. Math.*, **8**, 197–226.
- BAYER, D. A. & LAGARIAS, J. C. (1989) The nonlinear geometry of linear programming I, affine and projective scaling trajectories; II, Legendre transform coordinates and central trajectories. *Trans. Am. Math. Soc.*, **314**, 499–581.
- BROCKETT, R. W. (1988) Dynamical systems that sort lists, diagonalise matrices and solve linear programming problems. *Proceedings of 27th IEEE Conference on Decision and Control*. Austin, TX, pp. 799–803.
- BROCKETT, R. W. (1991) Dynamical systems that sort lists, diagonalise matrices and solve linear programming problems. *Linear Algebra Appl.*, **146**, 79–91.
- BROCKETT, R. W. (1993) Differential geometry and the design of gradient algorithms. *Proc. Sympos. Pure Math.*, **54**, 69–92.
- CHU, M. T. & DRIESSEL, K. R. (1990) The projected gradient method for least squares matrix approximation with spectral constraints. *SIAM J. Numer. Anal.*, **27**, 1050–1060.
- DEDIEU, J. P., PRIOURET, P. & MALAJOVICH, G. (2003) Newton's method on Riemannian manifolds: covariant alpha theory. *IMA J. Numer. Anal.*, **23**, 395–419.
- DOCARMO, M. P. (1992) *Riemannian Geometry*. Boston, MA: Birkhäuser.
- EDELMAN, A., ARIAS, T. A. & SMITH, T. (1998) The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, **20**, 303–353.
- EZQUERRO, J. A. & HERNÁNDEZ, M. A. (2002a) Generalized differentiability conditions for Newton's method. *IMA J. Numer. Anal.*, **22**, 187–205.
- EZQUERRO, J. A. & HERNÁNDEZ, M. A. (2002b) On an application of Newton's method to nonlinear operators with w -conditioned second derivative. *BIT*, **42**, 519–530.
- FERREIRA, O. P. & SVAITER, B. F. (2002) Kantorovich's theorem on Newton's method in Riemannian manifolds. *J. Complex.*, **18**, 304–329.
- GABAY, D. (1982) Minimizing a differentiable function over a differential manifold. *J. Optim. Theory Appl.*, **37**, 177–219.
- GUTIÉRREZ, J. M. & HERNÁNDEZ, M. A. (2000) Newton's method under weak Kantorovich conditions. *IMA J. Numer. Anal.*, **20**, 521–532.
- HALL, B. C. (2004) *Lie Groups, Lie Algebras, and Representations*. Graduate Text in Mathematics, vol. 222. New York: Springer.
- HELGASON, S. (1978) *Differential Geometry, Lie Groups, and Symmetric Spaces*. New York: Academic Press.

- HELMKE, U. & MOORE, J. B. (1990) Singular value decomposition via gradient flows. *Syst. Control Lett.*, **14**, 369–377.
- HELMKE, U. & MOORE, J. B. (1994) *Optimization and Dynamical Systems*. Communications and Control Engineering. London: Springer.
- KANTOROVICH, L. V. & AKILOV, G. P. (1982) *Functional Analysis*. Oxford: Pergamon.
- LI, C., LÓPEZ, G. & VICTORIA, M. M. (2009a) Monotone vector fields and the proximal point algorithm on Hadamard manifolds. *J. Lond. Math. Soc.*, **79**, 663–683.
- LI, C. & WANG, J. H. (2005) Convergence of Newton's method and uniqueness of zeros of vector fields on Riemannian manifolds. *Sci. China Ser. A*, **48**, 1465–1478.
- 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**, 228–251.
- LI, C., WANG, J. H. & DEDIEU, J. P. (2009b) Smale's point estimate theory for Newton's method on Lie groups. *J. Complex.*, **25**, 128–151.
- LUENBERGER, D. G. (1972) The gradient projection method along geodesics. *Manage. Sci.*, **18**, 620–631.
- MAHONY, R. E. (1994) Optimization algorithms on homogeneous spaces: with applications in linear systems theory. *Ph.D. Thesis*, Department of Systems Engineering, University of Canberra, Australia.
- MAHONY, R. E. (1996) The constrained Newton method on a Lie group and the symmetric eigenvalue problem. *Linear Algebra Appl.*, **248**, 67–89.
- MAHONY, R. E., HELMKE, U. & MOORE, J. B. (1993) Pole placement algorithms for symmetric realisations. *Proceedings of 32nd IEEE Conference on Decision and Control*, San Antonio, TX, pp. 1355–1358.
- MOORE, J. B., MAHONY, R. E. & HELMKE, U. (1994) Numerical gradient algorithms for eigenvalue and singular value calculations. *SIAM J. Matrix Anal. Appl.*, **15**, 881–902.
- OWREN, B. & WELFERT, B. (2000) The Newton iteration on Lie groups. *BIT Numer. Math.*, **40**, 121–145.
- SHUB, M. & SMALE, S. (1993) Complexity of Bezout's theorem, I: geometric aspects. *J. Am. Math. Soc.*, **6**, 459–501.
- SMALE, S. (1986) Newton's method estimates from data at one point. *The Merging of Disciplines: New Directions in Pure, Applied and Computational Mathematics* (R. Ewing, K. Gross & C. Martin eds). New York: Springer, pp. 185–196.
- SMITH, S. T. (1991) Dynamical systems that perform the singular value decomposition. *Syst. Control Lett.*, **16**, 319–327.
- 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* (A. Bloch ed), vol. 3. Providence, RI: American Mathematical Society, pp. 113–146.
- UDRISTE, C. (1994) *Convex Functions and Optimization Methods on Riemannian Manifolds*. Mathematics and Its Applications, Vol. 297. Dordrecht: Kluwer Academic Publishers.
- VARADARAJAN, V. S. (1984) *Lie Groups, Lie Algebras and Their Representations*. Graduate Text in Mathematics, vol. 102. New York: Springer.
- WANG, J. H. & LI, C. (2006) Uniqueness of the singular points of vector fields on Riemannian manifolds under the γ -condition. *J. Complex.*, **22**, 533–548.
- WANG, J. H. & LI, C. (2007) Kantorovich's theorem for Newton's method on Lie groups. *J. Zhejiang Univ. Sci. A*, **8**, 978–986.
- WANG, X. H. (2000) Convergence of Newton's method and uniqueness of the solution of equations in Banach space. *IMA J. Numer. Anal.*, **20**, 123–134.
- ZABREJKO, P. P. & NGUEN, D. F. (1987) The majorant method in the theory of Newton–Kantorovich approximates and the Ptak error estimates. *Numer. Funct. Anal. Optim.*, **9**, 671–674.