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

<table>
<thead>
<tr>
<th>Journal:</th>
<th><em>IMA Journal of Numerical Analysis</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>IMAJNA-ES-2008-085.R2</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Original Manuscript</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>01-Apr-2009</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Li, Chong; Mathematics</td>
</tr>
<tr>
<td></td>
<td>Wang, Jin-Hua; Mathematics</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Newton's method, Lie group, Lipschitz condition</td>
</tr>
</tbody>
</table>
Kantorovich’s theorems of Newton’s method for mappings and optimization problems on Lie groups*

Jin-Hua Wang  
Department of Mathematics, Zhejiang University of Technology  
Hangzhou 310032, P R China  
Email: wjh@zjut.edu.cn

Chong Li  
Department of Mathematics, Zhejiang University  
Hangzhou 310027, P R China  
Email: cli@zju.edu.cn

Abstract: 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 and Welfert, The Newton iteration on Lie groups, BIT, Numer., 2000, 40(2), 121-145) and (Wang and Li, Kantorovich’s Theorem for Newton’s Method on Lie groups, J of Zhejiang University Science A, 2007, 8(6), 978-986). Applications to optimization are provided and the results due to (Mahony, The constrained Newton method on a Lie group and the symmetric eigenvalue problem, Linear Algebra Appl., 1996, 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 [1, 2, 11, 15, 26, 27, 28, 31, 34, 35, 37]. Especially, 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

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

where

*Supported in part by the National Natural Science Foundation of China (grant 10671175; 10731060).
NEWTON’S METHOD ON LIE GROUPS

where \( G = SO(N, \mathbb{R}) := \{ x \in \mathbb{R}^{N \times N} | x^T x = I_N \text{ and } \det x = 1 \} \), \( D \in \mathbb{R}^{N \times N} \) is the diagonal matrix with diagonal entries \( 1, 2, \ldots, N \), and \( Q \) is a fixed symmetric matrix. Brockett [5, 6] and Chu et al [8] considered the following optimization problem:

\[
\min_{x \in G} \phi(x). \tag{1.2}
\]

Brockett [5, 6] showed that the minimum \( x^* \in G \) occurs when \( x^{*T} Q x^* \) 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 matrix \( Q \). There are many other examples where the linear algebra problems are formulated as optimization problems on Lie groups or homogeneous spaces (cf. [4, 20, 29, 36]).

For a general differentiable function \( \phi \) on a Lie group, some numerical optimization algorithms for solving (1.2) have been studied extensively; see for example, [7, 28, 29, 30, 32, 35]. In particular, R. E. Mahony used one-parameter subgroups of a Lie group to develop a version of Newton’s method on an arbitrary Lie group in [27], where the approach for solving (1.2) via Newton’s method had been explored and the local convergence was analyzed. 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, B. Owren and B. Welfert considered in [31] the implicit method for Lie groups, where they used the implicit Euler method as a generic example of an implicit integration method in Lie group setting. Consider the initial value problem on \( G \)

\[
\begin{align*}
x' &= x \cdot g(x) \\
x(0) &= x(0),
\end{align*}
\]

where \( g : G \to G \) 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 \( f : G \to G \) is the mapping defined by

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

B. Owren and B. Welfert introduced in [31] 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 [40], where, however, all results except [40, Theorem 2] on convergence of Newton’s method are made on Abelian groups.

In a vector space framework, as is well-known, one of the most important results on Newton’s method is Kantorovich’s theorem (cf. [22, 23]). Under the mild condition that the second Fréchet
derivative of $F$ is bounded (or more general, the first derivative is Lipschitz continuous) on a
proper open metric ball of the initial point $x_0$. Kantorovich’s theorem provides a simple and clear
criterion, 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 point estimate theory (i.e.,
$\alpha$-theory and $\gamma$-theory) in [33], where the notions of approximate zeros were introduced and the
rules to judge an initial point $x_0$ to be an approximate zero were established, depending on the
information of the analytic nonlinear operator at this initial point and at a solution $x^*$, respectively.
There are a lot of works on the weakness and/or the extension of the Lipschitz continuity made
on the mappings; see for example, [12, 13, 16, 17, 41, 42] and references therein. In particular,
Zabrejko-Nguen parametrized in [42] the classical Lipschitz continuity. Wang introduced in [41]
the notion of Lipschitz conditions with $L$-average to unify both Kantorovich’s and Smale’s criteria.

In a Riemannian manifold framework, an analogue of the well-known Kantorovich’s theorem was
given in [14] for Newton’s method for vector fields on Riemannian manifolds while the extensions
of the famous Smale’s $\alpha$-theory and $\gamma$-theory in [33] to analytic vector fields and analytic mappings
on Riemannian manifolds were done in [9]. In the recent paper [25], the convergence criteria in
[9] were improved by using the notion of the $\gamma$-condition for the vector fields and mappings on
Riemannian manifolds. The radii of uniqueness balls of singular points of vector fields satisfying
the $\gamma$-conditions were estimated in [39], while the local behavior of Newton’s method on Riemannian
manifolds was studied in [24]. Recently, inspired by previous work of Zabrejko and Nguyen in [42] on
Kantorovich’s majorant method, Alvarez et al introduced in [3] a Lipschitz-type radial function for
the covariant derivative of vector fields and mappings on Riemannian manifolds, and established a
unified convergence criterion of Newton’s method on Riemannian manifolds.

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 estab-
lished 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 a recent paper
[40], we gave a kind of Kantorovich’s theorem for (1.5) by using the following metric Lipschitz
condition at $x_0$:

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

where $d(\cdot, \cdot)$ is the Riemannian distance induced by the left invariant Riemannian metric. Clearly,
this kind of Lipschitz conditions is still dependent on the metric on the underlying group and
generally very difficult to verify in a non-compact group (cf. Example 3.1 in Section 3).

The purpose of the present paper is to establish Kantorovich’s theorem (independent of the
connection) for Newton’s method on Lie group. More precisely, under the assumption that the
derivative of $f$ satisfies the Lipschitz condition around the initial point (which is in terms of
one-parameter semigroups and 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 consequences, estimates of the convergence domains are also obtained. The main feature of our results (Theorems 3.1 and 3.2), is that they are completely independent of the metrics on the groups. In particular, Theorem 3.1 improves and extends [40, Theorem 2]; while Theorem 3.2 (and its corollaries) improves and extends [31, Theorem 2]. 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 \( \phi \); 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 [27] while the result on 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 [31] and [27].

2 Notions and preliminaries

Most of the notions and notations which are used in the present paper are standard; see for example [19, 38]. A Lie group \((G, \cdot)\) is a Hausdorff topological group with countable bases which also has the structure of an analytic manifold such that the group product and the inversion are analytic operations in the differentiable structure given on the manifold. The dimension of a Lie group is that of the underlying manifold, and we shall always assume that it is \( m \)-dimensional. The symbol \( e \) designates the identity element of \( G \). Let \( G \) be the Lie algebra of the Lie group \( G \) which is the tangent space \( T_e G \) of \( G \) at \( e \), equipped with Lie bracket \([\cdot, \cdot] : G \times G \to G\).

In the sequel we will make use of the left translation of the Lie group \( G \). We define for each \( y \in G \) the left translation \( L_y : G \to G \) by
\[
L_y(z) = y \cdot z \quad \text{for each } z \in G. \tag{2.1}
\]

The differential of \( L_y \) at \( z \) is denoted by \((L_y')_z\) which clearly determines a linear isomorphism form \( T_y G \) 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 form \( G \) to the tangent space \( T_e G \). The exponential map \( \exp : G \to G \) is certainly the most important construction associated to \( G \) and \( G \), and is defined as follows. Given \( u \in G \), let \( \sigma_u : \mathbb{R} \to G \) be the one-parameter subgroup of \( G \) determined by the left invariant vector field \( X_u : y \mapsto (L_y')_e(u) \); i.e., \( \sigma_u \) satisfies that
\[
\sigma_u(0) = e \quad \text{and} \quad \sigma_u'(t) = X_u(\sigma_u(t)) = (L_{\sigma_u(t)})_e(u) \quad \text{for each } t \in \mathbb{R}. \tag{2.2}
\]

The value of the exponential map \( \exp \) at \( u \) is then defined by
\[
\exp(u) = \sigma_u(1).
\]

Moreover, we have that
\[
\exp(tu) = \sigma_{tu}(1) = \sigma_u(t) \quad \text{for each } t \in \mathbb{R} \text{ and } u \in G \tag{2.3}
\]
and
\[
\exp(t + s)u = \exp(tu) \cdot \exp(su) \quad \text{for any } t, s \in \mathbb{R} \text{ and } u \in G. \tag{2.4}
\]
Note that the exponential map is not surjective in general. However, the exponential map is a

diffeomorphism on an open neighborhood of \(0 \in G\). In the case when \(G\) is Abelian, \(\exp\) is also a

homomorphism from \(G\) to \(G\), i.e.,

\[
\exp(u + v) = \exp(u) \cdot \exp(v) \quad \text{for all } u, v \in G.
\]  

(2.5)

In the non-abelian case, \(\exp\) is not a homomorphism and, by the Baker-Campbell-Hausdorff (BCH)

formula (cf. [38, p.114]), (2.5) must be replaced by

\[
\exp(w) = \exp(u) \cdot \exp(v)
\]  

(2.6)

for all \(u, v \in G\), where \(w\) is defined by

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

(2.7)

Let \(f : G \to G\) be a \(C^1\)-map and let \(x \in G\). We use \(f'_x\) to denote the differential of \(f\) at \(x\).

Then, by [10, 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 \(c : (-\varepsilon, \varepsilon) \to G\) with \(c(0) = x\) and \(c'(0) = \Delta_x\), one has that

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

(2.8)

In particular,

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

(2.9)

Define the linear map \(df_x : G \to G\) by

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

(2.10)

Then, by (2.9),

\[
df_x = f'_x \circ (L'_x)_{\varepsilon}.
\]  

(2.11)

Also, in view of definition, we have that, for all \(t \geq 0\)

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

(2.12)

and

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

(2.13)

For the remainder of the present paper, we always assume that \(\langle \cdot, \cdot \rangle\) is an inner product on \(G\)

and \(\| \cdot \|\) is the associated norm on \(G\). We now introduce the following distance on \(G\) which plays

a key role in the study. Let \(x, y \in G\) and define

\[
\rho(x, y) := \inf \left\{ \sum_{i=1}^k \| u_i \| \left| \begin{array}{c}
\text{there exist } k \geq 1 \text{ and } u_1, \ldots, u_k \in G \text{ such that } \\
y = x \cdot \exp u_1 \ldots \exp u_k
\end{array} \right. \right\},
\]  

(2.14)
where we adapt the convention that \( \inf \emptyset = +\infty \). It is easy to verify that \( \rho(\cdot, \cdot) \) is a distance on \( G \) and the topology induced by this distance is equivalent to the original one on \( G \).

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

\[
C_r(x) := \{ y \in G | \rho(x, y) < r \}.
\]

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

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

\[
\| T(x \exp u) - T(x) \| \leq L\| u \|
\]

holds for any \( u \in G \) and \( x \in C_r(x_0) \) such that \( \| u \| + \rho(x, x_0) < r \).

**Lemma 2.1.** Let \( 0 < r \leq \frac{1}{2} \) and let \( x_0 \in G \) be such that \( df^{-1}_{x_0} \) exists. Suppose that \( df^{-1}_{x_0}df \) 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 \( u_0, \ldots, u_k \in G \) satisfying \( x = x_0 \cdot \exp u_0 \cdot \ldots \cdot \exp u_k \) and \( \sum_{i=0}^{k} \| u_i \| < r \). Then \( df^{-1}_{x} \) exists and

\[
\| df^{-1}_{x_0} df \| \leq \frac{1}{1 - L(\sum_{i=0}^{k} \| u_i \|)}.
\]

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

\[
\| df^{-1}_{x_0} (df_{y_i} \exp u_i - df_{y_i}) \| \leq L \| u_i \| \quad \text{for each} \ 0 \leq i \leq k.
\]

Noting that \( y_{k+1} = x \), we have that

\[
\| df^{-1}_{x_0} df_x - I_G \| = \| df^{-1}_{x_0} (df_{y_{k+1}} \exp u_{k+1} - df_{x_0}) \|
\]

\[
\leq \sum_{i=0}^{k} \| df^{-1}_{x_0} (df_{y_i} \exp u_i - df_{y_i}) \|
\]

\[
= L \left( \sum_{i=0}^{k} \| u_i \| \right)
\]

\[
< 1.
\]

Thus the conclusion follows from the Banach lemma and the proof is complete. \( \square \)

### 3 Convergence criteria

Following [31], we define Newton’s method with initial point \( x_0 \) for \( f \) on a Lie group as follows.

\[
x_{n+1} = x_n \cdot \exp(-df^{-1}_{x_n}(f(x_n))) \quad \text{for each} \ n = 0, 1, \ldots.
\]

Let \( \beta > 0 \) and \( L > 0 \). The quadratic majorizing function \( h \), which was used in [23, 41], is defined by

\[
h(t) = \frac{L}{2} t^2 - t + \beta \quad \text{for each} \ t \geq 0.
\]
C. LI AND J. H. WANG

Let \( \{t_n\} \) denote the sequence generated by Newton’s method with initial value \( t_0 = 0 \) for \( h \), that is,
\[
t_{n+1} = t_n - h'(t_n)^{-1} h(t_n) \quad \text{for each } n = 0, 1, \ldots \tag{3.3}
\]
Assume that \( \lambda := L\beta \leq \frac{1}{2} \). Then \( h \) has two zeros \( r_1 \) and \( r_2 \):
\[
r_1 = \frac{1 - \sqrt{1 - 2\lambda}}{L} \quad \text{and} \quad r_2 = \frac{1 + \sqrt{1 - 2\lambda}}{L}; \tag{3.4}
\]
moreover, \( \{t_n\} \) is monotonic increasing and convergent to \( r_1 \), and satisfies that
\[
r_1 - t_n = \frac{\xi^{2^n - 1}}{\sum_{j=0}^{2^n-1} \xi^j} r_1 \quad \text{for each } n = 0, 1, \ldots \tag{3.5}
\]
where
\[
\xi = \frac{1 - \sqrt{1 - 2\lambda}}{1 + \sqrt{1 - 2\lambda}}. \tag{3.6}
\]
Recall that \( f : G \to G \) is a \( C^1 \)-mapping. In the remainder of this section, we always assume that \( x_0 \in G \) is such that \( df^{-1}_{x_0} \) exists and set \( \beta := \|df^{-1}_{x_0}f(x_0)\| \).

**Theorem 3.1.** Suppose that \( df^{-1}_{x_0}f \) satisfies the \( L \)-Lipschitz condition on \( C_{r_1}(x_0) \) and that
\[
\lambda = L\beta \leq \frac{1}{2}. \tag{3.7}
\]
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, the following assertions hold for each \( n = 0, 1, \ldots \):
\[
\varrho(x_{n+1}, x_n) \leq \|df^{-1}_{x_0}f(x_n)\| \leq t_{n+1} - t_n; \tag{3.8}
\]
\[
\varrho(x_n, x^*) \leq \frac{\xi^{2^n - 1}}{\sum_{j=0}^{2^n-1} \xi^j} r_1. \tag{3.9}
\]
**Proof.** Write \( v_n = -df^{-1}_{x_n}f(x_n) \) for each \( n = 0, 1, \ldots \). Below we shall show that each \( v_n \) is well-defined and
\[
\varrho(x_{n+1}, x_n) \leq \|v_n\| \leq t_{n+1} - t_n \tag{3.10}
\]
holds for each \( n = 0, 1, \ldots \). Granting this, one sees that the sequence \( \{x_n\} \) generated by Newton’s method (3.1) with initial point \( x_0 \) is well-defined and converges to a zero \( x^* \) of \( f \), because, by (3.1),
\[
x_{n+1} = x_n \cdot \exp v_n \quad \text{for each } n = 0, 1, \ldots \tag{3.11}
\]
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
\[
\|v_0\| = \| -df^{-1}_{x_0}(f(x_0)) \| = \beta = t_1 - t_0 \tag{3.12}
\]
it follows that (3.10) is true for \( n = 0 \). We now proceed by mathematical induction on \( n \). For this purpose, assume that \( v_n \) is well-defined and (3.10) holds for each \( n \leq k - 1 \). Then
\[
\sum_{i=0}^{k-1} \|v_i\| \leq t_k - t_0 < r_1 \quad \text{and} \quad x_k = x_0 \cdot \exp v_0 \cdot \ldots \cdot \exp v_{k-1}. \tag{3.11}
\]
Thus, we use Lemma 2.1 to conclude that $df^{-1}_{x_k}$ exists and

$$
\|df^{-1}_{x_k} df_x\| \leq \frac{1}{1-Lt_k} = -h'(t_k)^{-1}.
$$

(3.12)

Therefore, $v_k$ is well-defined. Observe that

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

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

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

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

$$
\|df^{-1}_{x_0} f(x_k)\| \leq \int_0^1 \|df^{-1}_{x_0} [df_{x_{k-1}} \exp(tv_{k-1}) - df_{x_{k-1}}]\|v_{k-1}\| dt
$$

$$
\leq \int_0^1 L \|tv_{k-1}\| \|v_{k-1}\| dt
$$

$$
\leq \frac{L}{2} (t_k - t_{k-1})^2
$$

$$
= h(t_k) + h'(t_k)(t_k - t_{k-1}) + \frac{1}{2} h''(t_k)(t_k - t_{k-1})^2
$$

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

$$
\|v_k\| = \|df^{-1}_{x_0} f(x_k)\|
$$

$$
\leq \|df^{-1}_{x_0} df_x\|\|df^{-1}_{x_0} f(x_k)\|
$$

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

$$
= t_{k+1} - t_k.
$$

(3.14)

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

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 $df^{-1}_{x^*}$ exists.

**Lemma 3.1.** Let $0 < r \leq \frac{1}{L}$, and let $x_0 \in C_r(x^*)$ be such that there exist $j \geq 1$ and $w_1, \ldots, w_j \in G$ satisfying

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

(3.15)

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

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

(3.16)
Proof. By Lemma 2.1, \( df_{x_0}^{-1} \) exists and

\[
\|df_{x_0}^{-1} df_x\| \leq \frac{1}{1 - L \sum_{i=1}^{j} \|w_i\|}.
\]

(3.17)

Write \( y_0 = x^* \) and \( y_i = y_{i-1} \cdot \exp w_i \) for each \( i = 1, \ldots, j \). Thus, by (3.15), we have \( y_j = x_0 \). Fix \( i \), 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

\[
df_{x_*}^{-1} (f(y_i) - f(y_{i-1})) = \int_0^1 df_{x_*}^{-1} (df_{y_{i-1} \cdot \exp(\tau w_i)} - df_x) w_i d\tau + w_i.
\]

(3.18)

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

\[
\|df_{x_*}^{-1} (df_{y_{i-1} \cdot \exp w_i} - df_{y_{i-1}})\| \leq L \|w_i\| \quad \text{for each} \quad \kappa = 1, \ldots, j.
\]

This gives that

\[
\|df_{x_*}^{-1} (df_{y_{i-1} \cdot \exp(\tau w_i)} - df_x)\| \\
\leq \sum_{\kappa=1}^{j} \|df_{x_*}^{-1} (df_{y_{i-1} \cdot \exp w_i} - df_{y_{i-1}})\| + \|df_{x_*}^{-1} (df_{y_{i-1} \cdot \exp(\tau w_i)} - df_{y_{i-1}})\|
\]

(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

\[
\|df_{x_0}^{-1} f(x_0)\| \leq \sum_{i=1}^{j} \left( \int_0^1 \|df_{x_*}^{-1} (df_{y_{i-1} \cdot \exp(\tau w_i)} - df_x)\| w_i \|d\tau + \|w_i\| \right)
\]

\[
\leq \sum_{i=1}^{j} \left( \int_0^1 L \left( \sum_{\kappa=1}^{j} \|w_\kappa\| + \tau \|w_i\| \right) \|w_i\| d\tau \right)
\]

\[
= \left( \frac{L}{2} \sum_{i=1}^{j} \|w_i\| + 1 \right) \sum_{i=1}^{j} \|w_i\|.
\]

Combining 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.

\( \square \)

Theorem 3.2. Let \( 0 < r \leq \frac{1}{4L} \). Suppose that \( f(x^*) = 0 \) and that \( df_{x_*}^{-1} df \) satisfies the L-Lipschitz condition on \( C_{\frac{1}{2L}}(x^*) \). 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 \( g(x^*, y^*) < \frac{3r}{1-Lr} \).
NEWTON’S METHOD ON LIE GROUPS

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

\[
x_0 = x^* \cdot \exp w_1 \ldots \exp w_j
\]

and \( \sum_{i=1}^{j} \|w_i\| < r \). By Lemma 3.1, \( df_{x_0}^{-1} \) exists and

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

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

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

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

Since \( df_{x_0}^{-1} df \) satisfies the \( L \)-Lipschitz condition on \( C_r \frac{x^*}{1 + L r} \), it follows that

\[
\|df_{x_0}^{-1}(df_{x_0} \exp u - df_x)\| \leq L \|u\|.
\] (3.21)

Consequently, by Lemma 2.1, one has that

\[
\|df_{x_0}^{-1}(df_{x_0} \exp u - df_x)\|
\leq \| df_{x_0}^{-1} df_x \| \|df_x^{-1}(df_{x_0} \exp u - df_x)\|
\leq \frac{L}{1 - L \sum_{i=1}^{j} \|w_i\|} \|u\|
= \tilde{L} \|u\|.
\]

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

\[
\tilde{\lambda} := \frac{\beta}{\bar{L}} \quad \text{and} \quad \bar{r}_1 = 1 - \frac{\sqrt{1 - 2\lambda}}{L} = \frac{2 \beta}{1 + \sqrt{1 - 2\lambda}}.
\] (3.22)

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,

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

The proof of the theorem is complete. \( \square \)

In particular, taking \( r = \frac{1}{4L} \) in Theorem 3.2, the following corollary is obvious.
Corollary 3.1. Suppose that $f(x^*) = 0$ and that $df^{-1}x \xi f$ satisfies the $L$-Lipschitz condition on $C^1_T(x^*)$. Let $x_0 \in C^1_T(x^*)$. 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 3.2 as well as Corollary 3.1 gives an estimate of the convergence domain for Newton’s method. However, we don’t 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 $B(0, r)$ to denote the open ball at 0 with radius $r$ on $G$, that is,

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

**Corollary 3.2.** Suppose that $f(x^*) = 0$ and that $df^{-1}x \xi f$ satisfies the $L$-Lipschitz condition on $C^1_T(x^*)$. Let $\rho > 0$ be the biggest number such that $C_\rho(e) \subseteq \exp(B(0, \frac{1}{L}))$ and let $r = \min \left\{ \frac{\rho}{3 + L\rho}, \frac{1}{4L} \right\}$. Write $N(x^*, r) := x^* \cdot \exp(B(0, r))$. 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 $v \in G$ such that $x_0 = x^* \cdot \exp v$ and

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

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

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

Hence Theorem 3.2 is applicable to concluding 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 $u \in G$ such that $\|u\| < \frac{1}{L}$ and $y^* = x^* \cdot \exp u$. Since

$$\left\| df_x^{-1} \int_0^1 df_x \cdot \exp(\tau u) d\tau - I \right\| = \left\| \int_0^1 df_x^{-1}(df_x \cdot \exp(\tau u) - df_x) d\tau \right\|$$

$$\leq \int_0^1 L\tau \|u\| d\tau$$

$$< 1,$$

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

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

We get that $u = 0$ and $y^* = x^*$, which completes the proof of the corollary. \qed
Recall that in the special case when $G$ is a compact connected Lie group, $G$ has a bi-invariant Riemannian metric (cf. [10, p.46]). Below, we assume that $G$ is a compact connected Lie group and endowed with a bi-invariant Riemannian metric. Therefore, an estimate of the convergence domain with the same property as in Corollary 3.2 is described in the following corollary.

**Corollary 3.3.** Let $G$ be a compact connected Lie group and endowed with a bi-invariant Riemannian metric. Let $0 < r < \frac{1}{1L}$. Suppose that $f(x^*) = 0$ and that $df_{x^*}^{-1}df$ satisfies the $L$-Lipschitz condition on $C_{\frac{1}{1L}} f(x^*)$. 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 3.2, 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 and endowed with a bi-invariant Riemannian metric, it follows from [19, p.224] that $c$ is also a one-parameter subgroup of $G$. Consequently, there exits $u \in G$ such that $y^* = x^* \cdot \exp(u)$. Therefore, due to the Banach lemma, $y^*$ is well-defined and converges quadratically to $x^*$.

As $\int_0^1 df_{x^*}^{-1}(df_{x^*} \cdot \exp(\tau u))d\tau$ is invertible and so is $\int_0^1 df_{x^*} \cdot \exp(\tau u)d\tau$. We then have that $u = 0$ and so $y^* = x^*$. This completes the proof of the corollary. \hfill \Box

**Corollary 3.4.** Let $G$ be a compact connected Lie group and endowed with a bi-invariant Riemannian metric. Suppose that $f(x^*) = 0$ and that $df_{x^*}^{-1}df$ satisfies the $L$-Lipschitz condition on $C_{\frac{1}{1L}} f(x^*)$. Let $x_0 \in C_{\frac{1}{1L}} f(x^*)$. 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 3.2 and 3.4 improve the corresponding local convergence result in [31, 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 $SO(N, \mathbb{R})$ which has been considered by Owren and Welfert in [31]. This, in particular, presents an example for which our results in the present paper are applicable but not each of the results in [31] and [27].
For the following example, we define the second differential \( d^2 \theta : G^2 \to G \) for a \( C^2 \)-map \( \theta : G \to G \):

\[
d^2 \theta_{x_1 u_2} = \left( \frac{\partial^2}{\partial t_2 \partial t_1} \theta(x \cdot \exp t_2 u_2 \cdot \exp t_1 u_1) \right)_{t_2 = t_1 = 0}
\]  

(3.26)

Then, by (2.13), we have that

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

(3.27)

**Example 3.1.** Let \( N \) be a positive integer and let \( I_N \) be the \( N \times N \) identity matrix. Let \( G \) be the special linear group under standard matrix multiplication (cf. [38]), that is,

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

Then \( G \) is a connected Lie group and its Lie algebra is

\[
G = T_e G = \mathfrak{sl}(N, \mathbb{R}) := \{ v \in \mathbb{R}^{N \times N} | \text{tr}(v) = 0 \},
\]

where \( e = I_N \). We endow \( G \) with the standard inner product

\[
\langle u, v \rangle_e = \text{tr}(u^T v) \quad \text{for any } u, v \in G;
\]  

(3.28)

hence the corresponding norm \( \| \cdot \| \) is the Frobenius norm. Moreover, the exponential map \( \exp : G \to G \) is given by

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

and its inverse is the logarithm (cf [18, p.34]):

\[
\exp^{-1}(z) = \sum_{k \geq 1} (-1)^{k-1} \frac{(z - I_N)^k}{k} \quad \text{for each } z \in G \text{ with } \| z - I_N \| < 1.
\]  

(3.29)

Let \( g : G \times \mathbb{R} \to G \) be a differentiable map and \( x^{(0)} \) a random starting point. Consider the following initial value problem on \( G \) studied in [31]:

\[
\begin{cases}
x' &= x \cdot g(x) \\
x(0) &= x^{(0)}
\end{cases}
\]  

(3.30)

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

\[
x = x^{(0)} \cdot \exp(l g(x)),
\]  

(3.31)

where \( l \) represents the size of the discretization step. Let \( f : G \to G \) be defined by

\[
f(x) = \exp^{-1}((x^{(0)})^{-1} \cdot x) - l g(x) \quad \text{for each } x \in G.
\]

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

Let \( g \) be the function considered in [31], which 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 \( v_0 \in G \) satisfying \( \|v_0\| \leq \frac{1}{10} \). To apply our results, we have to estimate the norm of \( df_x^{-1} \). To do this, write \( w(\cdot) = \exp^{-1}((x(0))^{-1} \cdot (\cdot)) \).

Let \( x := \exp v_1 \exp v_2 \ldots \exp v_k \) for some \( v_1, v_2, \ldots, v_k \in G \) with \( \sum_{i=0}^{k} \|v_i\| < \frac{1}{4} \). Since
\[
e^t - 1 \leq \frac{5}{4} t \quad \text{for each } t \in [0, \frac{1}{4}],
\]
one can use mathematical induction to prove that
\[
\|x - I_N\| \leq e^{\sum_{i=1}^{k} \|v_i\|} - 1 \leq \frac{5}{4} \sum_{i=1}^{k} \|v_i\|.
\]

Consequently,
\[
\|(x(0))^{-1} x - I_N\| \leq \frac{5}{4} \sum_{i=0}^{k} \|v_i\| < \frac{5}{16} < 1.
\]

Thus, by definition and using (3.29), one has that, for each \( u \in G \),
\[
dw_x(u) = \sum_{j \geq 1} (-1)^{j-1} \frac{(x(0))^{-1} x - I_N)^j}{j^2} \sum_{i=0}^{j} ((x(0))^{-1} x - I_N)^{i} (x(0))^{-1} xu ((x(0))^{-1} x - I_N)^{j-i}.
\]

Since
\[
\|(x(0))^{-1} xu\| = \|(x(0))^{-1} xu - u + u\| \leq (\|(x(0))^{-1} x - I_N\| + 1)\|u\|,
\]
it follows from (3.35) that
\[
\|dw_x(u) - u\| \leq \left( \sum_{j \geq 2} \frac{j \|(x(0))^{-1} x - I_N\|^{j+1} \|x(0))^{-1} x - I_N\|^{j-1}}{j^2} + \|(x(0))^{-1} x - I_N\| \right) \|u\|.
\]

Hence, by (3.34), we have
\[
\|dw_x - \mathbb{I}_G\| \leq \frac{2 \|(x(0))^{-1} x - I_N\|}{1 - \|(x(0))^{-1} x - I_N\|} \leq \frac{10 \sum_{i=0}^{k} \|v_i\|}{4 - \frac{5}{4} \sum_{i=0}^{k} \|v_i\|},
\]
where \( \mathbb{I}_G \) is the identity on \( G \). Similarly, one has
\[
\|d^2 w_x\| \leq \left( \sum_{j \geq 1} \frac{j \|(x(0))^{-1} x - I_N\|^{j+1} \|x(0))^{-1} x - I_N\|^{j-1} \|x(0))^{-1} x - I_N\|^{j+1} + j(j+1)(\|(x(0))^{-1} x - I_N\|^{j-1} \|x(0))^{-1} x - I_N\|^{j+2}}{j^2} \right) \frac{3 + \frac{5}{4} \sum_{i=0}^{k} \|v_i\|}{1 - \|(x(0))^{-1} x - I_N\|}.
\]

Combining this and (3.34) gives that
\[
\|d^2 w_x\| \leq \frac{3 + \frac{5}{4} \sum_{i=0}^{k} \|v_i\|}{(1 - \frac{5}{4} \sum_{i=0}^{k} \|v_i\|)^2}.
\]
Let \( x_0 = I_N \) and \( L = 51 \). Below, we verify that \( d f_{x_0}^{-1} d f \) satisfies the \( L \)-Lipschitz condition at \( x_0 \) on \( C^{\infty}_{\mathbb{R}}(x_0) \). To this purpose, let \( v \in \mathcal{G} \) and \( x \in C^{\infty}_{\mathbb{R}}(x_0) \). Then there exist \( k \geq 1 \) and \( v_1, \ldots, v_k \in \mathcal{G} \) such that \( x = x_0 \cdot \exp v_1 \cdot \ldots \cdot \exp v_k \) and \( \|v\| + \sum_{i=1}^{k} \|v_i\| < \frac{1}{50} \). Let \( s \in [0, 1] \) and write \( y := x \cdot \exp sv = \exp v_1 \ldots \exp v_k \cdot \exp sv \). Note that by (3.36), we have that

\[
\|d w_{x_0} - I_{\mathcal{G}}\| \leq \frac{10}{4 - 5\|v_0\|} \leq \frac{2}{7} \tag{3.38}
\]

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

\[
\|d^2 y_{|G}||d f_{x_0} - (I_{\mathcal{G}} - d g_{x_0})^{-1}\| = \|(I_{\mathcal{G}} - d g_{x_0})^{-1}\||d w_{x_0} - I_{\mathcal{G}}\| \leq \frac{1}{1 + 6 \cos 1 + 16 \sin 1} \frac{2}{7} \leq 1.
\]

Thus the Banach lemma is applied to conclude that

\[
\|d f_{x_0}^{-1}\| \leq \frac{1}{1 + 6 \cos 1 + 16 \sin 1} \frac{2}{9} \leq \frac{1}{10}. \tag{3.42}
\]

By definition,

\[
\|d^2 g_{|G}\| \leq 2 \sum_{j \geq 1} 2(\frac{2j}{j})^2 \|y_{|G}^j\|^2 + 5 \sum_{j \geq 1} 2 (\frac{2j + 1}{j})^2 \|y_{|G}^{j+1}\| \tag{3.43}
\]

Using (3.33), we have that \( \|y_{|G}^i\| \leq \frac{2}{3} \left( \sum_{i=1}^{k} \|v_i\| + \|v\| \right) + 1 \) for each \( i \geq 1 \) and it follows from (3.43) that

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

Note that, for each \( j \geq 1 \),

\[
\frac{2(2j)^i}{(2j - 1)!} \leq \frac{(2j - 1 + i) \ldots (2j + 1)2j}{(2j - 1)!} + \frac{(2j - 2 + i) \ldots 2j(2j - 1)}{(2j - 2)!} \text{ for } i = 2, 3.
\]
Then, by elementary calculations, we have that
\[ 2 \sum_{j \geq 1} \frac{(2j)^i}{(2j-1)!} \leq \sum_{l \geq 0} \frac{(l+i) \ldots (l+1)}{l!} = i! 2^i e \quad \text{for } i = 2, 3. \]

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

This, together with (3.39) and (3.42) yields that
\[ ||d^2f^{-1}d^2f_y|| \leq ||d^2f^{-1}||d^2w_u|| + ||d^2f^{-1}d^2g_y|| \]
\[ \leq \frac{72}{175(1-\frac{10}{7}(\sum_{i=1}^{k} ||v_i|| + s||v||))^2 + \frac{e}{10}(1320(\sum_{i=1}^{k} ||v_i|| + s||v||) + 136)}. \] (3.46)

Then it follows from (3.27) and the fact \( y = x \cdot \exp sv \) that
\[ ||d^2f^{-1}(d^2f_{x,y} - df_x)|| \]
\[ \leq \int_0^1 ||d^2f^{-1}d^2f_{x,y}sv|| ||v|| ds \]
\[ \leq \int_0^1 \left( \frac{72}{175(1-\frac{10}{7}(\sum_{i=1}^{k} ||v_i|| + s||v||))^2 + \frac{e}{10}(1320(\sum_{i=1}^{k} ||v_i|| + s||v||) + 136)} ||v|| ds \right) (3.47) \]
\[ \leq 51||v||. \]

Thus, the claim stands. Moreover, \( r_1 < \frac{1}{\ell} < \frac{1}{\ell_0} \) and so that \( d^2f^{-1}df \) 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)
\[ \lambda = L \beta = L||d^2f^{-1}f(x_0)|| \leq L||d^2f^{-1}||v_0|| \leq 51 \cdot \frac{1}{10} \cdot ||v_0|| < \frac{1}{2}. \]

Thus, Theorem 3.1 is applicable to concluding that the sequence generated by (3.1) with initial point \( x_0 = I_N \) converges to a zero \( x^* \) of \( f \).

To illustrate the application of Corollary 3.1, take \( x(0) = I_N \), that is, \( f : G \to G \) 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 \( x^* := I_N \) is a zero of \( f \). Furthermore by (3.35),
\[ dw_{x^*} = I_G \quad \text{and} \quad df_{x^*} = dw_{x^*} - dg_{x^*} = (1 + 6 \cos 1 + 16 \sin 1)I_G. \] (3.48)

As before, let \( v \in G \) and \( x \in C_{\frac{1}{30}}(x^*) \). Then there exist \( k \geq 1 \) and \( v_1, v_2, \ldots, v_k \in G \) such that \( x = \exp v_1 \exp v_2 \ldots \exp v_k \) and \( \sum_{i=1}^{k} ||v_i|| + ||v|| < \frac{1}{30} \). Similar, let \( s \in [0, 1] \) and write \( y := x \cdot \exp sv = \exp v_1 \exp v_2 \ldots \exp v_k \exp sv \). Note that by (3.37) (as \( v_0 = 0 \), one has that
\[ ||d^2w_y|| \leq \frac{3 + \frac{1}{10}}{(1 - \frac{5}{4}(\sum_{i=1}^{k} ||v_i|| + s||v||))^2}. \] (3.49)
Note that $x^* = x_0 = I_N$. 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
\[ \|d_{x_0}^{-1}d^2_{x_0}f\| \leq \frac{121}{480(1 - \frac{5}{4}(\sum_{i=1}^{k} \|v_i\| + s\|v\|))^2} + \frac{e}{12} \left( \sum_{i=1}^{k} \|v_i\| + s\|v\| \right) + 136 \]
and
\[ \|d_{x_0}^{-1}(d_{f_{x_0}} - d_{fx})\| \leq 51\|v\|. \]
That is, $d_{x_0}^{-1}d_f$ satisfies the L-Lipschitz condition at $x^*$ on $C_{\frac{1}{2}}(x^*)$ and on $C_{\frac{3}{2}}(x^*)$ because $\frac{1}{L} < \frac{1}{25}$. Take $x_0 = x^* \cdot \exp v$ with $v \in G$ and $\|v\| < \frac{1}{25}$. Corollary 3.1 is applicable to concluding 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 $C_{\frac{1}{2}}(x^*)$.

Furthermore, if we take $G$ to be the special orthogonal group under standard matrix multiplication, that is,
\begin{equation}
G = SO(N, \mathbb{R}) := \{ x \in \mathbb{R}^{N \times N} | x^T x = I_N \text{ and } \det x = 1 \}. \tag{3.50}
\end{equation}
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,
\begin{equation}
\mathcal{G} = \text{so}(N, \mathbb{R}) := \{ v \in \mathbb{R}^{N \times N} | v^T + v = 0 \}. \tag{3.51}
\end{equation}
Note that $[u, v] = uv - vu$ and $\langle [u, v], w \rangle = -\langle u, [w, v] \rangle$ for any $u, v, w \in \mathcal{G}$. One can easily verify (cf. [10, 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 3.4 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 3.1.** It would be helpful to make some comparisons of Theorem 3.1 with the corresponding result [40, Theorem 2], where the convergence criterion (3.7) was provided under the following metric L-Lipschitz condition at $x_0$:
\begin{equation}
\|d_{f_{x_0}}^{-1}(d_{f_{x'}} - d_{fx})\| \leq Ld(x', x), \forall x', x \in G \text{ with } d(x_0, x) + d(x, x') < r_1, \tag{3.52}
\end{equation}
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 much stronger than the L-Lipschitz condition on $C_{r_1}(x_0)$ given in Definition 2.1:
\begin{equation}
\|d_{f_{x_0}}^{-1}(d_{f_{x_0}^u} - d_{fx})\| \leq L\|u\|, \forall x \in C_{r_1}(x_0) \text{ and } u \in \mathcal{G} \text{ with } g(x, x_0) + \|u\| < r_1. \tag{3.53}
\end{equation}
Usually, in a non-compact Lie group, (3.52) is difficult to verify, in particular for points $x, x'$ that no one-parameter semigroups connect, because $d_{f(\cdot)}$ contains no information about the distance between the two points in $G$. Hence [40, Theorem 2] is not convenient to apply. For example, consider the group $G = SL(N, \mathbb{R})$ in Example 3.1. It would be very difficult to verify the metric L-Lipschitz condition (3.52) (in fact, we don’t know how to do that).
4 Applications to optimization problems

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

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

Newton’s method for solving (4.1) was presented in [27], where local quadratical convergence result was established for smooth function $\phi$.

Let $X \in G$. Following [27], we use $\tilde{X}$ to denote the left invariant vector field associated with $X$ defined by

$$\tilde{X}(x) = (L_x^t)^{-1} e^{X} \quad \text{for each } x \in G,$$

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

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

Let $\{X_1, \ldots, X_n\}$ be an orthonormal basis of $G$. According to [21, p. 356] (see also [27]), $\text{grad}\phi$ is a vector field on $G$ defined by

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

Then, Newton’s method with initial point $x_0 \in G$ considered in [27] can be written in a coordinate free form as follows:

**Algorithm 4.1** Find $X^k \in G$ such that $\tilde{X}^k = (L_x^t)^{-1} e X^k$ 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 $f : G \to G$ be a mapping defined by

$$f(x) = (L_x^t)^{-1} \text{grad}\phi(x) \quad \text{for each } x \in G. \quad (4.4)$$

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

$$(H_x\phi)X = (L_x^t)^{-1} \text{grad}(\tilde{X}\phi)(x) \quad \text{for each } X \in G. \quad (4.5)$$

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

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

$$df_x = H_x\phi \quad \text{for each } x \in G. \quad (4.6)$$
Proof. Let \( x \in G \) and let \( \{X_1, \ldots, X_n\} \) be an orthonormal basis of \( G \). In view of (4.3) and (4.4), we have that
\[
f(x) = (X_1, \ldots, X_n)((\bar{X}_1 \phi)(x), \ldots, (\bar{X}_n \phi)(x))^T.
\] (4.7)
Since \( \phi \) is a \( C^2 \)-mapping, it is easy to see by definition that
\[
\bar{X}_j(\bar{X}_\phi)(x) = \bar{X}(\bar{X}_j \phi)(x) \quad \text{for each} \quad X \in G \quad \text{and} \quad j = 1, 2, \ldots, n.
\] (4.8)
Therefore, by (4.7), we have that
\[
df_x(X) = \frac{d}{dt} \big|_{t=0} f(x \cdot \exp tX)
= (X_1, \ldots, X_n)(\frac{d}{dt} \big|_{t=0} (\bar{X}_1 \phi)(x \cdot \exp tX), \ldots, \frac{d}{dt} \big|_{t=0} (\bar{X}_n \phi)(x \cdot \exp tX))^T
= (X_1, \ldots, X_n)(\bar{X}_1(\bar{X}_1 \phi)(x), \ldots, \bar{X}_n(\bar{X}_n \phi)(x))^T
= (L_\phi^{-1}(\bar{X}_1, \ldots, \bar{X}_n)(\bar{X}_1(\bar{X}_1 \phi)(x), \ldots, \bar{X}_n(\bar{X}_n \phi)(x))^T
\]
where the fourth equality holds because of (4.8). This means that (4.6) holds and the proof is complete.

Remark 4.1. One can easily see from Proposition 4.1 that, with the same initial point, the sequence generated by Algorithm 4.1 for \( \phi \) 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 \( \beta_\phi := \| (H_{x_0} \phi)^{-1}(L_{x_0}') \phi \| \). Recall that \( r_1 \) is defined by (3.4). Then the main theorem of this section is as follows.

Theorem 4.1. Suppose that
\[
\lambda := L \beta_\phi \leq \frac{1}{2},
\] (4.9)
and that \((H_{x_0} \phi)^{-1}(L_{x_0}') \phi \) 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 \( \phi \): \( \text{grad}(\phi(x^*)) = 0 \).

Furthermore, if \( H_{x_0} \phi \) is additionally positive definite and the following Lipschitz condition is satisfied:
\[
\| (H_{x_0} \phi)^{-1} \| H_{x_0} \phi - H_x \phi \| \leq L \| u \| \quad \text{for} \quad x \in G \quad \text{and} \quad u \in G \quad \text{with} \quad g(x_0, x) + \| u \| < r_1.
\] (4.10)
Then \( x^* \) is a local solution of (4.1).

Proof. Recall that \( f \) is defined by (4.4). Then by Proposition 4.1, \( df_x = H_x \phi \) for each \( x \in G \). Hence, by assumptions, \( df_x \) 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 4.1, one sees that the first assertion holds.

To prove the second assertion, we assume that \( H_{x_0} \phi \) is additionally positive definite and the Lipschitz condition (4.10) is satisfied. It’s sufficient to prove that \( H_{x^*} \phi \) is positive definite. Let \( \lambda^* \) and \( \lambda^0 \) be the minimum eigenvalues of \( H_{x^*} \phi \) and \( H_{x_0} \phi \), respectively. Then \( \lambda^0 > 0 \). We have
NEWTON’S METHOD ON LIE GROUPS

to show that $\lambda^* > 0$. To do this, let $\{x_n\}$ be the sequence generated by Algorithm 4.1 and write $v_n = df_{x_n}^{-1}f(x_n)$ for each $n = 0, 1, \ldots$. Then, by Remark 4.1,

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

and by Theorem 3.1,

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

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

$$\|H_x \phi^{-1}\| (H_{x_{n+1}} \phi - H_x \phi) = \|H_x \phi^{-1}\| (H_{x_n} \exp(-v_n) \phi - H_x \phi) = \sum_{j=0}^n \|H_x \phi^{-1}\| (H_{x_j} \exp(-v_n) \phi - H_{x_j} \phi) \leq \sum_{j=0}^n \|v_n\| \leq \sum_{j=0}^n L(t_{n+1} - t_n) \leq Lr_1$$

thanks to (4.10)-(4.12). Since

$$\left| \frac{\lambda^*}{\lambda_0} - 1 \right| = \left| \frac{1}{\lambda_0} \min_{v \in \mathcal{G}, \|v\| = 1} \langle (H_x \phi) v, v \rangle - \min_{v \in \mathcal{G}, \|v\| = 1} \langle (H_x \phi) v, v \rangle \right| \leq \|H_x \phi^{-1}\| \|H_x \phi - H_{x_0} \phi\|,$$

it follows that

$$\left| \frac{\lambda^*}{\lambda_0} - 1 \right| \leq \lim_{n \to \infty} \|H_x \phi^{-1}\| \|H_{x_{n+1}} \phi - H_{x_0} \phi\| \leq Lr_1 < 1$$

thanks to (4.13). This implies that $\lambda^* > 0$ and completes the proof. \hfill $\square$

Similar to the proof of Theorem 4.1, we can verify the following corollaries.

**Corollary 4.1.** Let $x^*$ be a local optimal solution of (4.1) such that $(H_{x^*} \phi)^{-1}$ exists. Suppose that $(H_{x^*} \phi)^{-1}(H_{x^*} \phi)$ satisfies the $L$-Lipschitz condition on $C^1_\mathcal{G}(x_0)$. Let $x_0 \in C^1_\mathcal{G}(x^*)$. 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 $\phi$ and $\rho(x^*, y^*) < \frac{1}{L}$.

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

$$\|H_{x^*} \phi^{-1}\| \|H_{x_0} \exp u \phi - H_x \phi\| \leq L \|u\| \quad \text{for } x \in \mathcal{G} \text{ and } u \in \mathcal{G} \text{ with } \rho(x^*, x) + \|u\| < \frac{1}{L}. \quad (4.14)$$

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

**Corollary 4.2.** Let $x^*$ be a local optimal solution of (4.1) such that $(H_{x^*} \phi)^{-1}$ exists. Suppose that $(H_{x^*} \phi)^{-1}(H_{x^*} \phi)$ satisfies the $L$-Lipschitz condition on $C^1_\mathcal{G}(x_0)$. Let $\rho > 0$ be the biggest number such that $C_\rho(e) \subseteq \exp(B(0, \frac{1}{L}))$ and let $r = \min \left\{ \frac{\rho}{3 + L \rho}, \frac{1}{4L} \right\}$. Write $N(x^*, r) := x^* \cdot \exp(B(0, r))$. 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 4.3. Let $G$ be a compact connected Lie group and 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 $(H_{x^*}\phi)^{-1}(H_{x^*}\phi)$ satisfies the $L$-Lipschitz condition on $C^1_T(x_0)$. Let $x_0 \in C^1_T(x^*)$. 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 $\phi$ defined by (1.1), for which our results in the present paper are applicable but not the results in [27].

Example 4.1. Let $G$ and $G$ be given respectively by (3.50) and (3.51) with $N = 3$. Consider the optimization problem with $\phi : G \to \mathbb{R}$ given by

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

where $D$ is the diagonal matrix with diagonal entries $1, 2, 3$ and $Q$ is a fixed symmetric matrix. Then, it is known (see for example [27, 35, 34]) that

$$
\text{grad}\phi(x) = -(L'_{x^*}x)[x^TQx, D] = -x[x^TQx, D] \tag{4.15}
$$

and

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

Therefore, in view of (4.15),

$$
\text{grad}(\tilde{X}\phi)(x) = -(L'_{x^*}x)[x^TQx, [D, X^T]] = -x[x^TQx, [D, X^T]].
$$

This implies that

$$(H_{x^*}\phi)X = (L'_{x^*}x)^{-1}\text{grad}(\tilde{X}\phi)(x) = -[x^TQx, [D, X^T]]. \tag{4.17}
$$

Fix $X \in G$ and define the map $g : G \to G$ by

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

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

$$
d_{g_{x^*}\exp su} = \begin{bmatrix} \frac{d}{dt} |_{t=0} -[(xe^{t\nu}e^{t\nu})^TQxe^{t\nu}e^{t\nu}, [D, X^T]] \\
\frac{d}{dt} |_{t=0} e^{-tu}(xe^{tu})^TQxe^{tu}, [D, X^T] \\
\frac{d}{dt} |_{t=0} Ad_{e^{-ts}}((xe^{su})^TQxe^{su}), [D, X^T] \\
\frac{d}{dt} |_{t=0} (-u, (xe^{su})^TQxe^{su}), [D, X^T] \\
\end{bmatrix},
$$

where

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

Remember that the norm $\| \cdot \|$ is the Frobenius norm. Consequently, for each $u \in G$,

$$
\|d_{g_{x^*}\exp su}\| \leq 8\|u\|\|Q\|\|D\|\|X\| = 8\sqrt{14}\|u\|\|Q\|\|X\|.
$$

Hence, applying (2.13), we have that

$$
\|g(x \cdot exp u) - g(x)\| \leq \int_0^1 \|d_{g_{x^*}\exp (su)}u\|ds \leq 8\sqrt{14}\|u\|\|Q\|\|X\|.
$$
This together with (4.18) implies that
\[ \| H_{x,\exp} u \phi - H_x \phi \| \leq 8 \sqrt{14} \| u \| \| Q \| \quad \text{for each } u \in \mathcal{G}. \]  
(4.20)

In particular, taking
\[ Q = \begin{pmatrix} 1 & 0.003 & 0 \\ 0.003 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \]
we have
\[ \| H_{x,\exp} u \phi - H_x \phi \| \leq 16 \sqrt{21} \| u \| \quad \text{for each } u \in \mathcal{G}. \]  
(4.21)

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

Then \( \{b_1, b_2, b_3\} \) is a basis of \( \mathfrak{so}(3, \mathbb{R}) \). Thus, we can endow 2-norm \( \| \cdot \|_2 \) on \( \mathfrak{so}(3, \mathbb{R}) \) defined by
\[ \| u \|_2 = \sqrt{u_1^2 + u_2^2 + u_3^2} \quad \text{for each } u = u_1 b_1 + u_2 b_2 + u_3 b_3 \in \mathfrak{so}(3, \mathbb{R}). \]

It is routine to check that \( \| u \| = \sqrt{2} \| u \|_2 \) for each \( u \in \mathfrak{so}(3, \mathbb{R}) \). Let \( x_0 = I_3 \) 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
\[ (H_{x_0} \phi)^{-1}(b_1, b_2, b_3) = (b_1, b_2, b_3)C^{-1}. \]  
(4.23)

Therefore, for any \( u = u_1 b_1 + u_2 b_2 + u_3 b_3 \in \mathfrak{so}(3, \mathbb{R}), \)
\[ \| (H_{x_0} \phi)^{-1} u \| = \| (b_1, b_2, b_3)C^{-1}(u_1, u_2, u_3)^T \| 
= \sqrt{2} \| C^{-1}(u_1, u_2, u_3)^T \|_2 
\leq \sqrt{2} \| C^{-1} \| \| u \|_2 
= \| C^{-1} \| \| u \|. \]

Consequently,
\[ \| (H_{x_0} \phi)^{-1} \| \leq \| C^{-1} \| \leq \sqrt{2}. \]  
(4.24)

Combining this with (4.21) yields that \( (H_{x_0} \phi)^{-1}(H_{\phi} \phi) \) satisfies the \( L \)-Lipschitz condition with \( L = 16 \sqrt{42} \). Furthermore, by (4.15) and (4.23),
\[ \beta_\phi = \| (H_{x_0} \phi)^{-1}(L_{x_0} \phi)^{-1} \text{grad}(\phi(x_0)) \| = \| (H_{x_0} \phi)^{-1}[Q, D] \| = 0.003 \sqrt{2}. \]

Hence, \( \lambda = \beta_\phi < \frac{1}{2} \). Thus, Theorem 4.1 is applicable and the sequence generated by Algorithm 4.1 with initial point \( x_0 = I_3 \) is well-defined and convergent to a critical point \( x^* \) of \( \phi \).
Example 4.2. Consider the same problem as in Example 4.1 but with 

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

Let 

\[ x^* = \frac{1}{3} \begin{pmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 2 & 2 & 1 \end{pmatrix}. \]  \hspace{1cm} (4.25)

Since \( x^T Q x^* = \text{diag}(-2, 1, 4) \), it follows that \( x^* \) is a local optimal solution of \( \phi \) (cf. [5, 6, 27]). Let \( x \in G \) and \( u \in G \). Then it follows from (4.20) that

\[ \|H_{x, \exp u \phi} - H_x \phi\| \leq 8\sqrt{14}\|u\|\|Q\| = 56\sqrt{6}\|u\|. \]  \hspace{1cm} (4.26)

Let \( b_1, b_2, b_3 \) 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 did in Example 4.1, we can show that

\[ (H_x \phi)(b_1, b_2, b_3) = (b_1, b_2, b_3)C \]

and

\[ \|(H_x \phi)^{-1}\| \leq \|C^{-1}\| = \frac{\sqrt{33}}{12}. \]  \hspace{1cm} (4.27)

Hence, \( H_x \phi \) is positive definite and \( (H_x \phi)^{-1}(H_x \phi) \) satisfies the \( L \)-Lipschitz condition with \( L = 14\sqrt{22} \) thanks 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 4.3 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 \( x^* \).

References


NEWTON’S METHOD ON LIE GROUPS


NEWTON’S METHOD ON LIE GROUPS


