154 ACADEMIC World Academic Union

ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 17, No. 2, 2022, pp.154-160

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

Li Nan[1]
School of Mathematics and Statistics, Nanjing University of information Science and Technology, Nanjing 210044, China.
(Received August 10, 2022, accepted October 11, 2022)

Abstract: We propose a new Kantorovich theorem for Newton's method on Lie groups for mappings and matrix low-rank optimization problems, which arises from many applications. Under the classical hypothesis of f, we establish the convergence criteria of Newton's method from Lie group to its Lie algebra with weakened conditions, which improves the corresponding results in [20].
keywords: Lie group; Newton's method; Trace function; Lipschitz condition.

1. Introduction

In recent years, more and more attentions have been focused on studying numerical algorithms on manifolds. Classical optimization problems on manifolds are given by symmetric eigenvalue problems, low-rank nearest correlation matrix estimation, invariant subspace computations, optimization problems with equality constraints (see [7][9][21]). In this paper we focus on optimization problems on Lie groups. Consider the following problem:

$$\min_{x\in M} \phi(x), \tag{1.1}$$

where M is a Riemannian manifold and ϕ is a real-valued function on M. We will explore the optimization problem when ϕ is matrix trace function. It is essentially a kind of constrained matrix optimization problem. Many scholars have studied the problem. In [20], $\phi: G \to \mathbb{R}$ in (1.1) be given by

$$\phi(x) = -\mathrm{tr}(x^{\mathrm{T}} C x Q) \text{ for each } x \in G, \tag{1.2}$$

where $G = \mathrm{SO}(n,\mathbb{R}) := \{x \in \mathbb{R}^{n\times n} | x^{\mathrm{T}} x = \mathrm{I}_n, \det x = 1\}$, C is a fixed symmetric matrix and $Q = \mathrm{diag}(0_{n-\varsigma,n-\varsigma}, Q_\varsigma)$ with $Q_\varsigma = \mathrm{diag}(q_1, \cdots, q_\varsigma), 0 < q_1 \le q_2 \le \cdots \le q_\varsigma$, solved a kind of matrix trace function optimization problem with orthogonal constraints. Xu solved a generalized singular value of a Grassmann matrix pair or a real matrix pair. If $Q_\varsigma = \mathrm{diag}(\mathrm{I}_\varsigma, 0_{n-\varsigma,n-\varsigma})$ for $1 \le \varsigma \le n$, Xu solved this case by Riemannian inexact Newton-CG method [21]. Different from method in [21], We consider Newton's method on Lie group to solve this problem.

Brockett studied the optimization problem when

$$\phi(x) = -\mathrm{tr}(x^{\mathrm{T}} Q x D) \text{ for each } x \in G \tag{1.3}$$

in (1.1), where Q is a fixed symmetric matrix and D with the following structure

$$D = \mathrm{diag}(1,2,\dots,n),$$

showed that the minimum $x^* \in G$ occurs when ${x^*}^{\mathrm{T}} Q x^*$ is a diagonal matrix with diagonal entries (eigenvalues

[1] Corresponding Author E-mail: linan3749@126.com.

of Q) in ascending order [3][4]. Sato & Iwai studied the maximum value of the following functions on Riemannian manifolds:

$$\mathrm{tr}(U^{\mathrm{T}}AVN),$$

where $U \in \mathbb{R}^{m\times p}, V \in \mathbb{R}^{n\times p}$ and $U^{\mathrm{T}}U + V^{\mathrm{T}}V = \mathrm{I}_p, N \in \mathbb{R}^{p\times p}$ is a diagonal matrix, $A \in \mathbb{R}^{m\times n}$. The global optimal solution of this problem provides a set of left and right singular vectors, and transforms the problem of matrix trace function into finding singular values and singular vectors of A [16]. Mahony developed the Newton method with a single-parameter subgroup on the Lie group, and proved the local convergence [12]; Xu designed the Newton-CG method for the Grassmannian manifold problem to solve the singular value of the matrix pair [21].

Lie groups were originally used to solve differential equations. For solving ordinary differential equations on Lie groups, Owren and Welfert used the implicit Euler method for Lie groups [14]. Newton method is an effective method for solving approximate solutions of equations, and is widely used in large-scale optimal control problems, constrained smooth and non-smooth problems (see [13][15]). In Banach space, Kantorovich's theorem (see [10]) is an important result on Newton's method. It ensures the quadratic convergence of Newton's method, the existence and local uniqueness of the solution under very mild assumptions that the second Fréchet derivative of f is bounded on a proper open metric ball of the initial point x_0. Smith studied Newton's method in Riemannian manifolds [17][18], Ferreira and Svaiter generalized Kantorovich's theorem of Newton's method in Riemannian manifolds [5]. Li introduced the concept of the γ condition of the map f and established the γ condition of the Newton's method of the map f, extending and developing Smale's α-theory and γ-theory [11]. Wang established Kantorovich's theorem for Newton's method on Lie groups, under the classical assumption of the map f, they proved the convergence criterion of Newton's method to the zeros of the map f, and obtained the estimation of the convergence domains [20]. He established the unique ball of a zero of a map on Lie group and an estimation of the radius of convergence ball by Newton's method on a Lie group [6]. Argyros presented the local convergence analysis of Newton's method, obtained a larger convergence ball and a more precise distance error bound [1]. Argyros demonstrated semi-local convergence of Newton's method with sufficiently weak convergence criteria and tighter distance error bounds [2].

In this paper, we propose New Kantorovich's theorems for Newton method on Lie groups for mappings and matrix low-rank optimization problems. Under the classical assumption of f, we establish the convergence criterion of Newton's method from Lie group to its Lie algebra with weakened conditions. The rest of this paper is organized as follows. In Section 2 some useful notations, and lemma are given. In Section 3 we will give some theorems and an algorithm. Finally, in Section 4 concluding remarks are drawn.

1. Notions and preliminaries

Most of the notions and notation that are used in the present paper are standard. $\mathbb{R}$ and $\mathbb{R}^{n\times n}$ denote the sets of real numbers and $n \times n$ matrices with entries in $\mathbb{R}$. The symbols I_n and $0_{m\times n}$ represent the *n*-order identity matrix and the $m \times n$ zeros matrix, $\mathrm{tr}(\cdot)$ denote the trace function. A Lie group $(G,\cdot)$ is both a manifold and a topological group, and its group multiplication map and inverse map are both C^∞. We assume that the Lie group G is n-dimensional. The symbol e denotes the identity element of G. The tangent space T_eG of G at e is the Lie algebra of the Lie group G, and is also the set of all left-invariant vector fields of G , denoted as $\mathcal{G}$, equipped with the Lie bracket $[\cdot,\cdot]: \mathcal{G} \times \mathcal{G} \longrightarrow \mathcal{G}$. For any element x in the Lie group G, T_xG represents the tangent space of x.

Next, we will introduce some definitions that will be used. We define for each $y \in G$ the left translation $L_y: G \to G$ by

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

The differential of L_y at z is denoted by $(L_y{}')z$, which determines a linear isomorphism from the tangent space T_zG to $T_{(y\cdot z)}G$. The exponential map exp

$$\begin{aligned} \exp: \quad & \mathcal{G} \to G \\ & u \mapsto \exp(u) \end{aligned}$$

is a diffeomorphism on an open neighbourhood of $0 \in G$. When G is Abelian, exp is also a homomorphism from $\mathcal{G}$ to G, i.e.,

$$\exp(u+v) = \exp(u)\cdot\exp(v) = \exp(v)\cdot\exp(u) \quad \text{for all } u,v \in \mathcal{G}. \tag{2.2}$$

Then we will introduce the differential of f. Let $f : G \to \mathcal{G}$ be a C^1-map and let $x \in G$. We use f'_x to denote the differential of f at x. Define the linear map $\mathrm{d}f_x : \mathcal{G} \to \mathcal{G}$ by

$$\mathrm{d}f_x u = \left(\frac{\mathrm{d}}{\mathrm{d}t} f\big(x\cdot\exp(tu)\big)\right)_{t=0} \quad \text{for each } u \in \mathcal{G}. \tag{2.3}$$

Then, combining the above, we can conclude that

$$\mathrm{d}f_x = f'_x \circ (L'_x)_e. \tag{2.4}$$

In the remainder of this paper, we always assume that $\langle\cdot,\cdot\rangle$ is an inner product on $\mathcal{G}$ and $\|\cdot\|$ is the associated norm on $\mathcal{G}$. The distance $\rho(x,y)$ between two elements x,y on G, the corresponding ball $C_r(x)$ of radius r around x of G and the L-Lipschitz condition can be seen in [20].

LEMMA 2.1 Let $0 < r < \frac{1}{L}$ and let $x_0 \in G$ be such that the left inverse map of $\mathrm{d}f_{x_0}$ exists, denoted by $\mathrm{d}f_{x_0}^{\ddagger}$. Suppose that $\mathrm{d}f_{x_0}^{\ddagger}\mathrm{d}f$ 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 $u_0,\ldots,u_k \in \mathcal{G}_1 \subset \mathcal{G}$ satisfying $x = x_0\cdot\exp u_0 \cdots\cdots \exp u_k$ and $\sum_{i=0}^{k} \| u_i \| < r$. Then the left inverse map of $\mathrm{d}f_x$ exists and

$$\|\mathrm{d}f_x^{\ddagger}\mathrm{d}f_{x_0}\| \leq \frac{1}{1-L(\sum_{i=0}^{k}\|u_i\|)}. \tag{2.5}$$

Proof. We write $y_0 = x_0$ and $y_{i+1} = y_i\cdot\exp u_i$ for each $i = 0,\ldots,k$. According to the L-Lipschitz condition, and for any $z \in \mathcal{G}_1$, $\mathrm{d}f_{x_0}^{\ddagger}\mathrm{d}f_{x_0}z$ exists and is unique, we can get

$$\mathrm{d}f_{x_0}^{\ddagger}\left(\mathrm{d}f_{y_i\cdot\exp u_i} - \mathrm{d}f_{y_i}\right) \leq L \| u_i \| \quad \text{for each } 0 \leq i \leq k. \tag{2.6}$$

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

$$\begin{aligned}\|\mathrm{d}f_{x_0}^{\ddagger}\mathrm{d}f_x - I_{\mathcal{G}_1}\| &= \|\mathrm{d}f_{x_0}^{\ddagger}(\mathrm{d}f_{y_k\cdot\exp u_k} - \mathrm{d}f_{x_0})\| \\ &\leqslant \sum_{i=0}^{k} \|\mathrm{d}f_{x_0}^{\ddagger}(\mathrm{d}f_{y_i\cdot\exp u_i} - \mathrm{d}f_{y_i})\| \\ &= L\left(\sum_{i=0}^{k}\|u_i\|\right) \\ &< 1.\end{aligned}$$

Then, from Banach lemma, the proof is completed. □

We can see that different from optimization problem in [20] the left and right inverse maps of $\mathrm{d}f_{x_0}$ both exist, our problem only has the left inverse.

2. Convergence analysis and Algorithm

On Lie groups, according to Owren and Welfert [14], we define the iterative formula of Newton's method with initial value x_0 for f on the Lie group as follows:

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

We use quadratic majorizing function h, which was proposed in [10][19], is defined by

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

Let $\beta > 0, L > 0$ and assume that $\lambda := L\beta \leq \frac{1}{2}$. $\{t_n\}$ denote the iterative sequence generated by Newton's method with initial point $t_0 = 0$ for h, this is,

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

Then

$$r_1 = \frac{1-\sqrt{1-2\lambda}}{L} \tag{3.4}$$

is a zero of h. See [20].

Recall that $f : G \to \mathcal{G}$ is a C^1-mapping. From now on, we always assume that $x_0 \in G$ is such that the left inverse map of $\mathrm{d}f_{x_0}$ i.e. $\mathrm{d}f_{x_0}^{\ddagger}$ exists and set $\beta := \| \mathrm{d}f_{x_0}^{\ddagger} f(x_0) \|$.

THEOREM 3.1 Suppose that $\mathrm{d}f_{x_0}^{\ddagger}\mathrm{d}f$ satisfies the L-Lipschitz condition on $C_{r_1}(x_0)$ and

$$\lambda = L\beta \leq \frac{1}{2}. \tag{3.5}$$

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, we write $v_n = -\mathrm{d}f_{x_n}^{\ddagger} f(x_n)$ for each $n = 0,1, \ldots$, then the following relations hold:

$$\rho(x_{n+1}, x_n) \leq \| v_n \| \leq t_{n+1} - t_n, \tag{3.6}$$

Proof. Note that v_0 is well defined by assumption and $x_1 = x_0 \cdot \exp v_0$, then $\rho(x_1, x_0) \leq \| v_0 \|$. Since $\| v_0 \| = \| -\mathrm{d}f_{x_0}^{\ddagger} f(x_0) \| = \beta = t_1 - t_0$, then (3.6) is true for $n = 0$. We now use mathematical induction. We assume that v_n is well defined and (3.6) holds for each $n \leq k-1$. Then

$$\sum_{i=0}^{k-1} \|v_i\| \leq t_k - t_0 = t_k < r_1 \quad and \quad x_k = x_0 \cdot \exp v_0 \cdots\cdots \exp v_{k-1}. \tag{3.7}$$

From Lemma 2.1 we can conclude that $\mathrm{d}f_{x_k}^{\ddagger}$ exists, for any $z \in \mathcal{G}_1 \subset \mathcal{G}$, $(\mathrm{d}f_{x_k}^{-1})\mathrm{d}f_{x_k}z$ exists and unique, and

$$\| \mathrm{d}f_{x_k}^{\ddagger}\mathrm{d}f_{x_0} \| \leq \frac{1}{1-Lt_k} = -h'(t_k)^{-1}. \tag{3.8}$$

Therefore v_k is well defined. Combined with the proof of Wang [20], we can obtain the following inequality

$$\| \mathrm{d}f_{x_0}^{\ddagger} f(x_k) \| \leq h(t_k), \tag{3.9}$$

Combining this with (3.8) yields that

$$\begin{aligned} \|v_k\| &= \left\|-\mathrm{d}f_{x_k}^{\ddagger} f(x_k)\right\| \\ &\leqslant \left\|\mathrm{d}f_{x_k}^{\ddagger}\mathrm{d}f_{x_0}\right\|\left\|\mathrm{d}f_{x_0}^{\ddagger} f(x_k)\right\| \\ &\leqslant -h'(t_k)^{-1}h(t_k) \\ &= t_{k+1} - t_k. \end{aligned} \tag{3.10}$$

Since $x_{k+1} = x_k \cdot \exp v_k$, we have $\rho(x_{k+1}, x_k) \leq \| v_k \|$. Then we can get that (3.6) holds for $n = k$, which completes the proof of the theorem. □

Wang [20] gave the relevant proof when $\mathrm{d}f_{x(\cdot)}^{-1}$ exists, with the proof method of Wang [20], when only the left inverse exists, the relevant Lemma and Theorem are also true.

Next, we will introduce the Lie group and its Lie algebra to be used in our optimization problem. Take the Lie group G to be the special orthogonal group under standard matrix multiplication, let

$$G = \mathrm{SO}(n, \mathbb{R}) := \{x \in \mathbb{R}^{n\times n} \big| x^{\mathrm{T}}x = \mathrm{I}_n, \det x = 1\}. \tag{3.11}$$

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

$$\mathcal{G} = \mathrm{so}(n, \mathbb{R}) := \{v \in \mathbb{R}^{n\times n} \big| v^{\mathrm{T}} + v = 0\}. \tag{3.12}$$

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

$$\min \phi(x) := -\mathrm{tr}\left(x^{\mathrm{T}}CxQ_\varsigma\right) \quad \text{for each} \quad x \in G, \tag{3.13}$$

where C is a fixed symmetric matrix and Q_ς with the following structure

$$Q_\varsigma = \mathrm{diag}\left(\mathrm{I}_\varsigma, \mathrm{O}_{(n-\varsigma)\times(n-\varsigma)}\right) \quad \text{for} \quad 1 \leq \varsigma \leq n.$$

Let $X \in \mathcal{G}$. Following Mahony [12], we use $\tilde{X}$ to denote the left-invariant vector field associated with X defined by

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

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

$$(\tilde{X}\phi)(x) = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \phi(x \cdot \exp\ tX). \tag{3.14}$$

Let $\{X_1, \ldots, X_n\}$ be an orthonormal basis of $\mathcal{G}$. According to Helmke [8], grad ϕ is a vector field on G defined by

$$\mathrm{grad}\,\phi(x) = \left(\tilde{X}_1, \ldots, \tilde{X}_n\right)\left(\tilde{X}_1\phi(x), \ldots, \tilde{X}_n\phi(x)\right)^{\mathrm{T}} = \sum_{j=1}^{n} \tilde{X}_j\phi(x)\tilde{X}_j \text{ for each } x \in G \tag{3.15}$$

Then is known (see [12][17][18]) that

$$\mathrm{grad}\,\phi(x) = -x[x^{\mathrm{T}}Cx, Q_\varsigma], \tag{3.16}$$

$$\mathrm{grad}\,(\tilde{X}\phi)(x) = -x[x^{\mathrm{T}}Cx, [Q_\varsigma, X^{\mathrm{T}}]]. \tag{3.17}$$

Then Newton method with initial point $x_0 \in G$ can be written in a coordinate-free form as follows.

Algorithm 3.2 (for $1 \leq \varsigma \leq \frac{n}{2}$)

Step 0. $x_0 \in G$;

Step 1. Find $X_k \in \mathcal{G}_1 = \left\{\begin{pmatrix} 0_{\varsigma\times\varsigma} & X_{12} \\ -X_{12}^{\mathrm{T}} & 0_{(n-\varsigma)\times(n-\varsigma)} \end{pmatrix} \mid X_{12} \in \mathbb{R}^{\varsigma\times(n-\varsigma)}\right\} \subset \mathcal{G}$ such that

$$\mathrm{grad}\ \phi(x_k) + \mathrm{grad}\left(\tilde{X}^k\phi\right)(x_k) = 0,$$

then

$$\left[[Q_\varsigma, X_k^{\mathrm{T}}], x_k^{\mathrm{T}}Cx\right] = \left[x_k^{\mathrm{T}}Cx_k, Q_\varsigma\right];$$

Step 2. $x_{k+1} = \exp X^k \cdot x_k$, where exp X^k is the matrix exponential of X_k;

Step 3. See $k := k + 1$ and go to Step 1.

We can see that due to the special structure of Q_ς, converting X_k into a block matrix, the diagonal block is zero matrices, which reduces computation and storage space.

Since X_k is anti-symmetric, we let $X_k = U^{\mathrm{T}}\Lambda U$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_j (j = 1, \ldots, n)$ being pure imaginary numbers. Then it is easy to know $\exp X^k$ is orthogonal. Since x_k is orthogonal, then $x_{k+1} = \exp X^k \cdot x_k$ is also orthogonal.

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

$$f(x) = (L'_x)_e^{-1}\mathrm{grad}\,\phi(x) \quad \text{for each } x \in G. \tag{3.18}$$

Define the linear operator $H_x\phi: \mathcal{G}_1 \to \mathcal{G}$ for each $x \in G$ by

$$(H_x\phi)X = (L'_x)_e^{-1}\mathrm{grad}\left(\tilde{X}\phi\right)(x) \quad \text{for each } X \in \mathcal{G}_1. \tag{3.19}$$

Then $H_{(\cdot)}\phi$ defines a mapping from G to $\mathcal{L}(\mathcal{G})$. We can get that

$$(H_x\phi)X = \left[x^{\mathrm{T}}Cx, \left[Q_\varsigma, X^{\mathrm{T}}\right]\right]. \tag{3.20}$$

Fix $X \in \mathcal{G}_1$ and define the map $g: G \to \mathcal{G}$ by

$$g(x) = (H_x\phi)X = \left[x^{\mathrm{T}}Cx, \left[Q_\varsigma, X^{\mathrm{T}}\right]\right]. \tag{3.21}$$

for each $x \in G$.

From Wang [20] we know that $\mathrm{d}f_x = H_x\phi$ for each $x \in G$, $f(\cdot)$ and $H_{(\cdot)}\phi$ be defined by (3.18) and (3.19), and with the same initial point, the sequence generated by Algorithm 3.2 for ϕ is consistent with the one generated by Newton's method (3.1) for f defined by (3.18).

We can have that $(H_x\phi)(\cdot)$ is not a surjective map from $\mathcal{G}_1$ to $\mathcal{G}$ but injective map while x_0 is well defined. Let$x_0 \in G$be such that $(H_{x_0}\phi)^{\ddagger}$ exists, and let $\beta_\phi := \| (H_{x_0}\phi)^{\ddagger}\left(L'_{x_0}\right)_e^{-1}\mathrm{grad}\,\phi(x_0) \|$. $(H_{x_0}\phi)^{\ddagger}$ denote the left inverse of $H_{x_0}\phi$. Wang [20] gave the relevant proof when $(H_{x_0}\phi)^{-1}$ exists, with the proof method of them, when only the left inverse exists, the following Theorem is also true. r_1 is defined by (3.4).

THEOREM 3.3

Suppose that

$$\lambda := L\beta_\phi \leq \frac{1}{2} \tag{3.22}$$

and $(H_{x_0}\phi)^{\ddagger}(H_{(\cdot)}\phi)$ satisfies the L-Lipschitz condition on $C_{r_1}(x_0)$. Then the sequence generated by Algorithm 3.2 with initial point x_0 is well defined, and it converges to a critical point x^* of ϕ at which grad $\phi(x^*) = 0$.

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

$$\left\|\left(H_{x_0}\phi\right)^{\ddagger}\right\| H_{x\cdot\exp u}\phi - H_x\phi \parallel \leqslant L \parallel u \parallel \text{ for } x \in G \text{ and } u \in \mathcal{G} \text{ with } \rho(x_0, x) + \parallel u \parallel < r_1, \tag{3.23}$$

then x^* is a local solution of (3.13).

Now we have that Algorithm 3.2 is feasible. We can use Algorithm 3.2 to solve a class of matrix trace function optimization problems like (3.13).

Compared with problem in Wang [20], if Q is a rank-deficient matrix and the structure makes the Lie algebra become a subset of the orthogonal matrix Lie algebra, which yields new optimization problem. Then in Algorithm 3.2, $H_{x_0}\phi$ is not a reversible mapping as in Wang [20], and can only satisfy the left inverse mapping. At this point, it is necessary to prove the convergence analysis when only the left inverse exists. Compared with problem (1.2) the conditions are weakened. When $\varsigma = n$ our problem reduces to model in Wang [20].

3. Concluding remarks

In this paper, we provide New Kantorovich's theorems for Newton method on Lie groups for mappings and matrix low-rank optimization problems, which arises from many applications and can solve many problems in various fields. Under the classical assumption of f, we establish the convergence criterion of Newton's method from Lie group to its Lie algebra with weakened conditions, which improves the corresponding results in Wang [20]. In addition, a new algorithm is proposed to solve this optimization problem, and the feasibility of the algorithm is proved.

References:

1. ARGYROS, I. K. & HILOUT, S. (2015) Enlarging the convergence ball of Newton's method on Lie groups. Journal of Numerical Analysis and Approximation Theory., 44, 11-24.
2. ARGYROS, I. K. (2022) Convergence of Newton' s Method on Lie Groups. The Theory and Applications of Iteration Methods., 349-360.
3. 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, 799-803.
4. BROCKETT, R. W. (1991) Dynamical systems that sort lists, diagonalise matrices and solve linear programming problems. Linear Algebra Appl., 146, 79-91.
5. FERREIRA, O. P. & SVAITER, B. F. (2002) Kantorovich's theorem on Newton's method in Riemannian manifolds. J. Complex., 18, 304-329.
6. HE, J. S. & WANG, J. H. & YAO, J. C. (2013) Local Convergence of Newtons Method on Lie Groups and Uniqueness Balls. Abstract and Applied Analysis., 2013, 1-9.
7. HE, P. & XU, X. H. & DING, J. & FAN, B. C. (2020) Low-rank nonnegative matrix factorization on Stiefel manifold. Information Sciences: an International Journal., 514, 131-148.
8. HELMKE, U. & MOORE, J. B. (1994) Optimization and Dynamical Systems. Communications and Control

Engineering. London: Springer.

9. HU, J. & LIU, X. & WEN, Z. W. & YUAN, Y. X. (2020) A brief introduction to manifold optimization. J. Oper. Res. Soc. China, 8, 199-248.
10. KANTOROVICH, L. V. & AKILOV, G. P. (1982) Functional Analysis. Oxford: Pergamon.
11. LI, C. & WANG, J. H. & DEDIEU, J. P. (2009) Smale's point estimate theory for Newton's method on Lie groups. J. Complex., 25, 128-151.
12. MAHONY, R. E. (1996) The constrained Newton method on a Lie group and the symmetric eigenvalue problem. Linear Algebra Appl., 248, 67-89.
13. OLIVEIRA, F. R. & FERREIRA, O. P. (2020) Inexact Newton method with feasible inexact projections for solving constrained smooth and nonsmooth equations. Applied Numerical Mathematics., 156, 63-76.
14. OWREN, B. & WELFERT, B. (2000) The Newton iteration on Lie groups. BIT Numer. Math., 40, 121-145.
15. PABLOS, B. & GERDTS, M. (2021) Substructure exploitation of a nonsmooth Newton method for large-scale optimal control problems with full discretization. Mathematics and Computers in Simulation., 190, 641-658.
16. SATO, H. & IWAI, T. (2013) A Riemannian Optimization Approach to the Matrix Singular Value Decomposition. SIAM Journal on Optimization., 23, 188-212.
17. SMITH, S. T. (1993) Geometric optimization method for adaptive filtering. Ph.D. Thesis, Harvard University, Cambridge, MA.
18. SMITH, S. T. (1994) Optimization techniques on Riemannian manifolds. Fields Institute Communications, (A. Bloch ed), vol. 3. Providence, RI: American Mathematical Society, 113-146.
19. 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.
20. WANG, J. H. & LI, C. (2011) Kantorovich's theorems for Newton's method for mappings and optimization problems on Lie groups. IMA Journal of Numerical Analysis., 31, 322-347.
21. XU, W. W. , MICHAEL K. N. & BAI, Z. J. (2022) Geometric Inexact Newton Method for Generalized Singular Values of Grassmann Matrix Pair. SIAM Journal on Matrix Analysis and Applications., 43, 535-560.