A Geodesic-Based Riemannian Gradient Approach to Averaging on the Lorentz Group

Jing Wang [1], Huafei Sun [1,2,*] and Didong Li [3]

[1] School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, China; wangjingzzumath@gmail.com
[2] Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing 100081, China
[3] Department of Mathematics, Duke University, Durham, NC 27708, USA; didongli@math.duke.edu
[*] Correspondence: huafeisun@bit.edu.cn; Tel.: +86-136-4129-2882

Abstract: In this paper, we propose an efficient algorithm to solve the averaging problem on the Lorentz group $O(n, k)$. Firstly, we introduce the geometric structures of $O(n, k)$ endowed with a Riemannian metric where geodesic could be written in closed form. Then, the algorithm is presented based on the Riemannian-steepest-descent approach. Finally, we compare the above algorithm with the Euclidean gradient algorithm and the extended Hamiltonian algorithm. Numerical experiments show that the geodesic-based Riemannian-steepest-descent algorithm performs the best in terms of the convergence rate.

Keywords: Lorentz group; geodesic; average; Riemannian-steepest-descent algorithm; extended Hamiltonian algorithm

1. Introduction

Averaging measured data on matrix manifolds is one of the most frequently arising problems in many fields. For example, one may need to determine an average eye of a set of optical systems [1]. In [2], the authors concern the statistics of covariance matrices for biomedical engineering applications. The current research mainly focuses on the compact Lie groups, such as the special orthogonal group $SO(n)$ [3] and the unitary group $U(n)$ [4], or other matrix manifolds, such as the special Euclidean group $SE(n, \mathbb{R})$ [5], Grassmann manifold $Gr(n, k)$ [6,7], the Stiefel manifold $St(n, k)$ [8,9] and the symmetry positive-definite matrix manifold $SPD(n)$ [10,11]. However, there is little research on averaging over the Lorentz group $O(n, k)$.

In [12], Clifford algebra, as a generalization of quaternions, is applied to approximate the average on the special Lorentz group $SO(1, 2)$. In engineering, the Lorentz group plays a fundamental role, e.g., in the analysis of motion of charged particles in electromagnetism [13]. In physics, the Lorentz group forms a basis for the transformation of parabolic catadioptric image features [14]. According to [15], the Lorentz group $O(3, 1)$ is equal to the linear transformations of the projective plane and used to characterize the catadioptric fundamental matrices that are applied to study the catadioptric cameras.

This paper aims at investigating the averaging on the Lorentz group with a left-invariant metric that plays the role of Riemannian metric so that the geodesic is given in closed form. The considered averaging optimization problem could be solved numerically via a geodesic-based Riemannian-steepest-descent algorithm (RSDA). Furthermore, the devised method RSDA is compared with the line-search algorithm, Euclidean gradient algorithm (EGA) and the second-order learning algorithm, extended Hamiltonian algorithm (EHA), proposed in [16,17].

The remainder of the paper is organized as follows. In Section 2, we summarize the geometric structures of the Lorentz group. The averaging optimization problem, the geodesic-based Riemannian-steepest-descent algorithm and the extended Hamiltonian algorithm on the Lorentz group

are presented in Section 3. In Section 4, we show the numerical results of the RSDA and compare the convergence rate of the RSDA with the other two algorithms. Section 5 presents the conclusions of the paper.

2. Geometry of the Lorentz Group

In this section, we review the foundations of Lorentz group and describe a Riemannian metrization in which the geodesic could be written in closed forms.

Let $M(n, \mathbb{R})$ be the set of $n \times n$ real matrices and $GL(n, \mathbb{R})$ be its subset containing only nonsingular matrices. It is well-known that $GL(n, \mathbb{R})$ is a Lie group, i.e., a group which is also a differentiable manifold and for which the operations of group multiplication and inverse are smooth.

Definition 1. *The Lorentz group $O(n, k)$ as a Lie subgroup of $GL(n + k, \mathbb{R})$ is defined by*

$$O(n, k) := \{A \in GL(n + k, \mathbb{R}) \mid A^\top I_{n,k} A = I_{n,k}\}, \quad I_{n,k} := \begin{pmatrix} I_n & \mathbf{0}_{n \times k} \\ \mathbf{0}_{k \times n} & -I_k \end{pmatrix},$$

where I is identity matrix, and $\mathbf{0}$ is $n \times k$ zero matrix.

For any $A \in O(n, k)$, the following identities hold:

$$\det(A) = \pm 1, \quad I_{n,k}^2 = I_{n+k}, \quad I_{n,k}^\top = I_{n,k}^{-1} = I_{n,k}, \quad A^\top = I_{n,k} A^{-1} I_{n,k}, \quad A^{-\top} = I_{n,k} A I_{n,k}.$$

The tangent space $T_A O(n, k)$, the Lie algebra $\mathfrak{o}(n, k)$ and the normal space $N_A O(n, k)$ associated with the Lorentz group $O(n, k)$ can be characterized as follows:

$$\begin{cases} T_A O(n, k) = \{A I_{n,k} \Omega \mid \Omega \in \mathbb{R}^{(n+k) \times (n+k)}, \Omega^\top = -\Omega\}, \\ \mathfrak{o}(n, k) = \{I_{n,k} \Omega \mid \Omega \in \mathbb{R}^{(n+k) \times (n+k)}, \Omega^\top = -\Omega\}, \\ N_A O(n, k) = \{I_{n,k} A S \mid S \in \mathbb{R}^{(n+k) \times (n+k)}, S^\top = S\}. \end{cases}$$

We employ the left-invariant metric as the Riemannian metric on $T_A O(n, k)$ at $A \in O(n, k)$

$$\langle X, Y \rangle_A := \langle A^{-1} X, A^{-1} Y \rangle_{I_{n+k}} = \text{tr}((A^{-1} X)^\top (A^{-1} Y)), \quad X, Y \in T_A O(n, k). \tag{1}$$

Let $\gamma : [0, 1] \to O(n, k)$ be a smooth curve in $O(n, k)$, and its length associated with the Riemannian metric (1) is defined as

$$L(\gamma) = \int_0^1 \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}} \, dt. \tag{2}$$

The geodesic distance between two points A_1 and A_2 in $O(n, k)$ is the infimum of lengths of curves connecting them, namely,

$$d(A_1, A_2) := \inf\{L(\gamma) \mid \gamma : [0, 1] \to O(n, k), \gamma(0) = A_1, \gamma(1) = A_2\}.$$

Applying the variational method from Equation (2), we can find that the geodesic equation satisfies

$$\ddot{\gamma} + \Gamma_\gamma(\dot{\gamma}, \dot{\gamma}) = 0,$$

where Γ denotes the Christoffel symbol, the over-dot and the double over-dot denote first-order and second-order derivation with respect to parameter t, respectively. When the values $\gamma(0) = A \in O(n, k)$ and $\dot{\gamma}(0) = X \in T_A O(n, k)$ are specified, we shall denote the solution as $\gamma_{A,X}(t)$, which is associated with the Riemannian exponential map $\exp_A : T_A O(n, k) \to O(n, k)$ defined as

$$\exp_A(X) = \gamma_{A,X}(1).$$

Let $f : O(n,k) \to \mathbb{R}$ be a differentiable function and $\nabla_A f \in T_A O(n,k)$ denotes the Riemannian gradient of function f with respect to the metric (1), which is defined by the compatibility condition

$$\langle \nabla_A f, X \rangle_A = \langle \partial_A f, X \rangle^E, \quad \forall X \in T_A O(n,k), \tag{3}$$

where $\partial_A f$ denotes the Euclidean gradient of function f and $\langle \cdot, \cdot \rangle^E$ denotes the Euclidean metric.

The exponential of a matrix A in $M(n, \mathbb{R})$ is defined by

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k.$$

We remark that $e^{A+B} = e^A e^B$ when $AB = BA$ for A and B in $M(n, \mathbb{R})$.

For the Frobenius norm $\| \cdot \|$, when $\|A - I_n\| < 1$, the logarithm of A is given by

$$\log A = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(A - I_n)^k}{k}.$$

In general, $\log(AB) \neq \log A + \log B$. We here recall the important fact that

$$\log(B^{-1}AB) = B^{-1} \log(A)B, \quad B \in GL(n, \mathbb{R}).$$

Lemma 1. *Suppose that $f(A)$ is a scalar function of $n \times n$ matrix A in $M(n, \mathbb{R})$ [18]. If*

$$d f(A) = \mathrm{tr}(W \, d\, A)$$

holds, then

$$\partial_A f = W^\top,$$

where $W^\top$ denotes the derivative of $f(A)$.

The structure of the Riemannian gradient associated with the Riemannian metric (1) is given by the following result.

Theorem 1. *The Riemannian gradient of a sufficiently regular function $f : O(n,k) \to \mathbb{R}$ associated with the Riemannian metric (1) satisfies*

$$\nabla_A f = \frac{1}{2} A I_{n,k} (I_{n,k} A^\top \partial_A f - \partial_A^\top f A I_{n,k}). \tag{4}$$

Proof of Theorem 1. According to equality (3), the gradient $\nabla_A f$ is computed as

$$\langle \partial_A f - A^{-\top} A^{-1} \nabla_A f, X \rangle^E = 0, \quad \forall X \in T_A O(n,k),$$

as X is arbitrary, the condition above implies that $\partial_A f - A^{-\top} A^{-1} \nabla_A f \in N_A O(n,k)$, hence $\partial_A f - A^{-\top} A^{-1} \nabla_A f = I_{n,k} A S$, with $S^\top = S$, and thus we have

$$\nabla_A f = A A^\top \partial_A f - A I_{n,k} S.$$

Since $\nabla_A f \in T_A O(n,k)$, we have

$$(\nabla_A f)^\top I_{n,k} A + A^\top I_{n,k} \nabla_A f = 0.$$

Inserting the expression $\nabla_A f$ into the equation above, we have

$$S = \frac{1}{2}(I_{n,k} A^\top \partial_A f + \partial_A^\top f A I_{n,k}).$$

Substituting S into the expression of $\nabla_A f$, we finish the proof. □

In the text, we would like to give credit to [19] for the explicit expression of geodesic over the general linear group endowed with the natural left-invariant Riemannian metric (1). Under the same Riemannian metric, it is clear that the geodesic on the Lorentz group can be written in closed form (cf. Theorem 2.14 in [19]).

Proposition 1. *The geodesic* $\gamma_{A,X} : [0,1] \to O(n,k)$*, with* $A \in O(n,k)$*,* $X \in T_A O(n,k)$ *corresponding to the Riemannian metric* (1) *has expression*

$$\gamma_{A,X}(t) = A\,e^{t(A^{-1}X)^\top}\,e^{t[A^{-1}X - (A^{-1}X)^\top]}. \tag{5}$$

Proof of Proposition 1. In fact, by the variational formula $\delta \int_0^1 \langle \dot\gamma, \dot\gamma \rangle_\gamma \, dt = 0$, the geodesic equation in Christoffel form reads

$$\ddot\gamma + \dot\gamma \dot\gamma^\top \gamma^{-\top} - \gamma \dot\gamma^\top \gamma^{-\top} \gamma^{-1} \dot\gamma - \dot\gamma \gamma^{-1} \dot\gamma = 0.$$

By the definition that $H(t) := \gamma^{-1}(t)\dot\gamma(t)$, we see that the geodesic satisfies the Euler–Lagrange equation

$$\dot H(t) = H^\top(t)H(t) - H(t)H^\top(t).$$

We can verify that the curve $\gamma(t) = A\,e^{t(A^{-1}X)^\top}\,e^{t[A^{-1}X - (A^{-1}X)^\top]}$ emanating from A with a velocity X is the solution of the Euler–Lagrange equation above. □

Remark 1. *From the proof of Proposition* 1 *and the fact* $A^{-\top} = I_{n,k} A I_{n,k}$*,* $\forall X \in T_A O(n,k)$*, we conclude that*

$$\begin{aligned}
\Gamma_A(X,X) &= XX^\top A^{-\top} - AX^\top A^{-\top} A^{-1} X - XA^{-1}X \\
&= XX^\top I_{n,k} A I_{n,k} - AX^\top I_{n,k} A I_{n,k} A^{-1} X - XA^{-1}X.
\end{aligned}$$

3. Optimization on the Lorentz Group

This section aims at investigating the averaging optimization problem on the Lorentz group, and finding the average value out of a set of Lorentz matrices. Sections 3.1 and 3.2 recall the geodesic-based Riemannian-steepest-descent algorithm and the extended Hamiltonian algorithm on the Lorentz group, respectively.

3.1. Riemannian-Steepest-Descent Algorithm on the Lorentz Group

According to the expression of geodesic (5) with $A = \gamma(0)$, $B = \gamma(1) \in O(n,k)$ and $X = \dot\gamma(0) \in T_A O(n,k)$, we have

$$A^{-1}B = e^{(A^{-1}X)^\top}\,e^{A^{-1}X - (A^{-1}X)^\top}, \tag{6}$$

and, furthermore, we have

$$\log(A^{-1}B) = \log[e^{(A^{-1}X)^\top}\,e^{A^{-1}X - (A^{-1}X)^\top}]. \tag{7}$$

Note that the geodesic distance $d(A, B) = \sqrt{\langle X, X \rangle_A} = \|A^{-1}X\|$. From Equation (6), it is difficult to give an explicit expression of X in A and B. In [20], the authors show that the geodesic distance connecting A and B could be obtained by solving the following problem

$$\min_{U \in \mathfrak{o}(n,k)} \| e^{U^\top} e^{U - U^\top} - A^{-1}B \|^2. \tag{8}$$

Then, the geodesic distance between A and B, $d(A, B) = \|U^*\|$, where U^* is a minimizer of Equation (8). In order to compute U^*, a descent procedure can be applied so we need to find the gradient first. The differential with respect to U of the function $J(U) = \| e^{U^\top} e^{U - U^\top} - A^{-1}B \|^2$ is given by

$$\begin{aligned}
\mathrm{d}\,J(U) &= 2\,\mathrm{d}(\mathrm{tr}((e^{U^\top} e^{U - U^\top} - A^{-1}B)^\top (e^{U^\top} e^{U - U^\top} - A^{-1}B))) \\
&= 2\,\mathrm{tr}((e^{U^\top} e^{U - U^\top} - A^{-1}B)^\top e^{U^\top}\,\mathrm{d}\,U\,e^{U - U^\top}) \\
&= 2\,\mathrm{tr}(e^{U - U^\top}(e^{U^\top} e^{U - U^\top} - A^{-1}B)^\top e^{U^\top}\,\mathrm{d}\,U),
\end{aligned}$$

and thus, according to Lemma 1, we can get the Euclidean gradient of $J(U)$

$$\partial_U J = 2\,e^U (e^{U^\top} e^{U - U^\top} - A^{-1}B)\,e^{U^\top - U}. \tag{9}$$

On the other hand, recall the Goldberg exponential formula in [21]

$$\log(e^X e^Y) = X + Y + \sum_{n=2}^{\infty} \sum_{|\omega|=n} g_\omega \omega, \tag{10}$$

where ω denotes a word in the letters X and Y, and the inner sum is over all words ω with length $|\omega| = n$. The symbol g_ω denotes Goldberg coefficient on the word ω, a rational number. If the word ω begins with X, say

$$\omega = \omega_{X,Y} = X^{r_1} Y^{r_2} X^{r_3} \cdots (X \bigvee Y)^{r_m},$$

where $r_1, \cdots, r_m$ are positive integers satisfying $r_1 + \cdots + r_m = n$, and $(X \bigvee Y)^{r_m}$ denotes X^{r_m} if m is odd and Y^{r_m} if m is even. $\omega_{Y,X}$ is similarly defined.

The Goldberg coefficient g_ω is

$$g_\omega = \int_0^1 t^{m'}(t-1)^{m''} G_{r_1}(t) \cdots G_{r_m}(t)\,\mathrm{d}\,t,$$

in which $m' = [m/2]$ and $m'' = [(m-1)/2]$, $[\,]$ denotes the greatest integer of the enclosed number, and the polynomials $G_r(t)$ are defined recursively by $G_1(t) = 1$ and $G_r(t) = r^{-1}(\mathrm{d}\,/\,\mathrm{d}\,t)\{r(r-1)G_{r-1}(t)\}$ for $r = 2, 3, \ldots$.

According to Theorem 1 in [22], we can obtain the following result.

Proposition 2. *If $\|U\| < \frac{1}{2}$, then*

$$\| \log(A^{-1}B) - U \| \le \frac{4\|U\|^2}{1 - 2\|U\|}. \tag{11}$$

Proof of Proposition 2. According to the Goldberg exponential formula (10), Equation (7) can be recast as

$$\log(A^{-1}B) = \log(e^{U^\top} e^{U - U^\top}) = U + \sum_{n=2}^{\infty} \sum_{|\omega|=n} g_\omega \omega, \tag{12}$$

in which ω denotes a word in the letters $U^\top$ and $U - U^\top$.

By using properties of the Goldberg coefficient g_ω (see [23]), we get

$$|g_\omega| \leq 2^{-(n-m)}(mC_{m-1}^{m'}),$$

where $C_n^m = n!/m!(n-m)!$ is the usual binomial coefficient. Furthermore, the number of words $\omega_{U^\top, U-U^\top}$ of length $|\omega| = n$ is just C_{n-1}^{m-1}. Thus, the sum of $|g_\omega|$ extended over all words ω of length n and involved m parts is

$$2 \times C_{n-1}^{m-1} \times 2^{-(n-m)}(mC_{m-1}^{m'})^{-1}.$$

Therefore, due to $\|U - U^\top\| \leq 2\|U\|$, and for $n \geq 2$, $\|\omega\| \leq 2^{n-1}\|U\|^n$, we have

$$
\begin{aligned}
\|\log(A^{-1}B) - U\| &\leq \sum_{n=2}^{\infty} \sum_{|\omega|=n} |g_\omega|\|\omega\| \\
&\leq \sum_{n=2}^{\infty} \sum_{m=1}^{n} 2 \times C_{n-1}^{m-1} \times 2^{-(n-m)}(mC_{m-1}^{m'})^{-1} \times 2^{n-1}\|U\|^n \\
&\leq \sum_{n=2}^{\infty} \sum_{m=1}^{n} 2^m C_{n-1}^{m-1}(2^{m-1})^{-1}\|U\|^n \\
&= 2 \sum_{n=2}^{\infty} \sum_{m=1}^{n} C_{n-1}^{m-1}\|U\|^n \\
&= \sum_{n=2}^{\infty} 2^n \|U\|^n \\
&= \frac{4\|U\|^2}{1 - 2\|U\|}.
\end{aligned}
$$

Here, in the last step, the equality holds if and only if $\|U\| < \frac{1}{2}$. □

Remark 2. *If the condition $\|U - U^\top\| \leq \epsilon\|U\|$, $0 < \epsilon \leq 1$ is added, then, when $\|U\| < 1$,*

$$\|\log(A^{-1}B) - U\| \leq \frac{2\epsilon\|U\|^2}{1 - \|U\|}.$$

Remark 3. *Proposition 2 implies that we can replace the geodesic distance $d(A, B) = \|U\|$ by $\|\log(A^{-1}B)\|$ where the latter one is much easier to handle.*

In order to compare different Lorentz matrices, the following discrepancy $D : O(n,k) \times O(n,k) \to \mathbb{R}$ is defined

$$D(A, B) := \|\log(A^{-1}B)\| = \sqrt{\mathrm{tr}\left(\log^\top(A^{-1}B)\log(A^{-1}B)\right)}, \quad A, B \in O(n,k). \tag{13}$$

For the Lorentz group, the criterion function $f : O(n,k) \to \mathbb{R}$ to measure a collection $\{B_1, B_2, \ldots, B_N\}$ of samples randomly generated is defined by

$$f(A) := \frac{1}{N} \sum_q \|\log(B_q^{-1}A)\|^2, \quad A, B_q \in O(n,k), \quad q = 1, 2, \cdots, N, \tag{14}$$

which appears as a specialization, to the Lorentz group, of the Karcher criterion [24].

Definition 2. *The average value of $\{B_1, B_2, \ldots, B_N\}$ is defined as*

$$\bar{A} := \arg \min_{A \in O(n,k)} f(A). \tag{15}$$

Based on the case in the Euclidean space of the line search method with the negative Euclidean gradient, in order to achieve the average value, a geodesic-based Riemannian-steepest-descent algorithm can be expressed as [25]

$$A_{p+1} = \gamma_{A_p, -\nabla_{A_p} f}(t_p), \tag{16}$$

where $p \geq 0$ is an iteration step-counter, the initial guess $A_0 \in O(n,k)$ is arbitrarily chosen, and t_p denotes a (possibly variable) stepsize schedule.

Note that the Riemannian gradient of the criterion function is crucial to solving the optimization problem by the geodesic-based Riemannian-steepest-descent algorithm (16).

Theorem 2. *The Riemannian gradient of the criterion function (14) is*

$$\nabla_A f = \frac{2}{N} \sum_q A \log(B_q^{-1} A). \tag{17}$$

Proof of Theorem 2. The differential with respect to A of the criterion function (14) is

$$\begin{aligned}
\mathrm{d}\, f(A) &= \frac{1}{N} \sum_q \mathrm{d}(\mathrm{tr}(\log^\top (B_q^{-1} A) \log(B_q^{-1} A))) \\
&= \frac{2}{N} \sum_q \mathrm{tr}(\log^\top (B_q^{-1} A) \,\mathrm{d} \log(B_q^{-1} A)) \\
&= \frac{2}{N} \sum_q \mathrm{tr}(\log^\top (B_q^{-1} A) A^{-1} \,\mathrm{d}\, A) \\
&= \left\langle \left(\frac{2}{N} \sum_q (\log^\top (B_q^{-1} A) A^{-1}) \right)^\top, \mathrm{d}\, A \right\rangle^E,
\end{aligned}$$

and, furthermore, according to Lemma 1, we get the Euclidean gradient

$$\partial_A f = \frac{2}{N} \sum_q A^{-\top} \log(B_q^{-1} A). \tag{18}$$

Substituting the Euclidean gradient (18) into the expression of the Riemannian gradient (4), we have

$$\begin{aligned}
\nabla_A f &= \frac{1}{N} \sum_q (A \log(B_q^{-1} A) - A I_{n,k} \log^\top (B_q^{-1} A) I_{n,k}) \\
&= \frac{1}{N} \sum_q (A \log(B_q^{-1} A) - A \log(I_{n,k} (B_q^{-1} A)^\top I_{n,k})) \\
&= \frac{1}{N} \sum_q (A \log(B_q^{-1} A) - A \log(B_q^{-1} A)^{-1}) \\
&= \frac{2}{N} \sum_q A \log(B_q^{-1} A).
\end{aligned}$$

□

Associated with the expressions of the geodesic (5) and the Riemannian gradient (17) of the criterion function, the geodesic-based Riemannian-steepest-descent algorithm (16) is recast as

$$A_{p+1} = A_p \, \mathrm{e}^{t_p U_p^\top} \, \mathrm{e}^{t_p (U_p - U_p^\top)}, \tag{19}$$

where $U_p := -A_p^{-1} \nabla_{A_p} f = -\frac{2}{N} \sum_q \log(B_q^{-1} A_p)$.

In order to compute a nearly-optimal stepsize, the function $f(A_{p+1})$ considered as a function of the stepsize t_p can be expanded in Taylor series at the point $t_p = 0$ as

$$f(A_{p+1}) = \frac{1}{2}f_{2,p}t_p^2 + f_{1,p}t_p + f(A_p) + o(t_p^2),$$

where the $f_{1,p}$ and $f_{2,p}$ as the coefficients of the Taylor expansion are computed by

$$f_{1,p} = \frac{\mathrm{d}\, f(A_{p+1})}{\mathrm{d}\, t_p}\bigg|_{t_p=0}, \quad f_{2,p} = \frac{\mathrm{d}^2\, f(A_{p+1})}{\mathrm{d}\, t_p^2}\bigg|_{t_p=0}.$$

Furthermore, the nearly-optimal stepsize value that minimizes the $f(A_{p+1})$ is $\hat{t}_p := -\frac{f_{1,p}}{f_{2,p}}$, provided that $f_{1,p} \le 0$, $f_{2,p} > 0$, hence $\hat{t}_p \ge 0$.

Thus, to obtain the $f_{1,p}$ and $f_{2,p}$, the function $f(A_{p+1})$ is recast as

$$f(A_{p+1}) = \frac{1}{N}\sum_q \mathrm{tr}(\log^\top(B_q^{-1}A_p\, e^{t_p U_p^\top}\, e^{t_p(U_p - U_p^\top)})\log(B_q^{-1}A_p\, e^{t_p U_p^\top}\, e^{t_p(U_p - U_p^\top)})).$$

Calculation leads to the first-order derivative of the function $f(A_{p+1})$

$$\frac{\mathrm{d}\, f(A_{p+1})}{\mathrm{d}\, t_p} = \frac{2}{N}\sum_q \mathrm{tr}(\log^\top(B_q^{-1}A_p\, e^{t_p U_p^\top}\, e^{t_p(U_p - U_p^\top)})\, e^{t_p(U_p^\top - U_p)}\, U_p\, e^{t_p(U_p - U_p^\top)}),$$

and setting $t_p = 0$ yields the $f_{1,p}$

$$f_{1,p} = \frac{2}{N}\sum_q \mathrm{tr}(\log^\top(B_q^{-1}A_p)U_p) = -\mathrm{tr}(U_p^\top U_p) \le 0.$$

Under the first-order derivative of the function $f(A_{p+1})$ above, direct calculations lead to the second-order derivative of the function $f(A_{p+1})$

$$\frac{\mathrm{d}^2\, f(A_{p+1})}{\mathrm{d}\, t_p^2} = \frac{2}{N}\sum_q \mathrm{tr}\left(e^{t_p(U_p^\top - U_p)}\, U_p^\top U_p\, e^{t_p(U_p - U_p^\top)} + \log^\top(B_q^{-1}A_p\, e^{t_p U_p^\top}\, e^{t_p(U_p - U_p^\top)})\right.$$
$$\left. e^{t_p(U_p^\top - U_p)}(U_p^\top U_p - U_p U_p^\top)e^{t_p(U_p - U_p^\top)}\right),$$

and, setting $t_p = 0$, we get

$$\begin{aligned}
f_{2,p} &= \frac{2}{N}\sum_q \mathrm{tr}\left(U_p^\top U_p + \log^\top(B_q^{-1}A_p)(U_p^\top U_p - U_p U_p^\top)\right)\\
&= 2\,\mathrm{tr}(U_p^\top U_p) - \mathrm{tr}(U_p^\top(U_p^\top U_p - U_p U_p^\top))\\
&= 2\,\mathrm{tr}(U_p^\top U_p) \ge 0.
\end{aligned}$$

Hence, the nearly-optimal stepsize of the geodesic-based Riemannian-steepest-descent algorithm (19) is

$$\hat{t}_p = -\frac{f_{1,p}}{f_{2,p}} = \frac{1}{2}. \tag{20}$$

Therefore, the iteration algorithm (19) with the stepsize (20) may be recast explicitly as

$$A_{p+1} = A_p\, e^{\frac{1}{N}\sum_q \log^\top(A_p^{-1}B_q)}\, e^{\frac{1}{N}\sum_q(\log(A_p^{-1}B_q) - \log^\top(A_p^{-1}B_q))}. \tag{21}$$

3.2. Extended Hamiltonian Algorithm on the Lorentz Group

References [16,17] introduced a general theory of extended Hamiltonian (second order) learning on Riemannian manifold, especially, as an instance of learning by constrained criterion function optimization on the matrix manifolds. For the Lorentz group, the extended Hamiltonian algorithm can be expressed by

$$
\begin{cases}
\dot{A} = X, \\
\dot{X} = -\Gamma_A(X, X) - \nabla_A f - \mu X,
\end{cases}
\tag{22}
$$

where $A \in O(n,k)$, $X \in T_A O(n,k)$ denotes the instantaneous learning velocity, $f : O(n,k) \to \mathbb{R}$ denotes a criterion function, and the constant $\mu > 0$ denotes a viscosity parameter.

Inserting the the expressions of the Christoffel matrix function Γ_A in Remark 1 and the Riemannian gradient $\nabla_A f$ (4) into the general extended Hamiltonian system (22), calculations lead to the following expressions:

$$
\begin{cases}
\dot{A} = AH, \\
\dot{H} = H^\top H - HH^\top + \frac{1}{2} I_{n,k}(\partial_A^\top f A I_{n,k} - I_{n,k} A^\top \partial_A f) - \mu H.
\end{cases}
$$

Hence, the second equation may be integrated numerically by a Euler stepping method, while the first one may be integrated via the geodesic. Namely, system (22) may be implemented by

$$
\begin{cases}
A_{p+1} = A_p \, e^{\eta H_p^\top} \, e^{\eta H_p - \eta H_p^\top}, \\
H_{p+1} = \eta(H_p^\top H_p - H_p H_p^\top) + (1 - \eta\mu)H_p + \frac{1}{2}\eta I_{n,k}(\partial_{A_p}^\top f A_p I_{n,k} - I_{n,k} A_p^\top \partial_{A_p} f),
\end{cases}
\tag{23}
$$

where $\eta > 0$ denotes the learning rate. The effectiveness of the algorithm is ensured if and only if

$$
\sqrt{2\lambda_{max}} < \mu < 1/\eta,
$$

where λ_{max} denotes the largest eigenvalue of the Hessian matrix of the criterion function $f(A)$ (see [17]).

4. Numerical Experiments

In the present section, we give two numerical experiments on the Lorentz group $O(3,1)$ to demonstrate the effectiveness and performance of the proposed algorithms. The Lorentz group as a homogeneous space has four connected components. Now, we study the optimization on a connected component $SO_0(3,1)$ containing the identical matrix I_4.

In order to emphasize the behavior of the optimization methods RSDA, the EHA and the EGA, in the following experiments, $\eta = 0.5$ and $\mu = 1.5$ are used.

4.1. Numerical Experiments on Averaging Two Lorentz Matrices

In subsection 3.1, it is noticed that the geodesic distance $d(A, B)$ and the discrepancy $D(A, B)$ between two points A and B are different. Now, given $B_1, B_2 \in O(3,1)$, numerical experiment results are shown to compare $d(B_1, B_2)$ and $D(B_1, B_2)$. The following two points are sought:

$$
B_1 = \begin{pmatrix} \cosh(0.5) & 0 & 0 & \sinh(0.5) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh(0.5) & 0 & 0 & \cosh(0.5) \end{pmatrix}, \quad B_2 = \begin{pmatrix} \cosh(0.1) & 0 & 0 & \sinh(0.1) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh(0.1) & 0 & 0 & \cosh(0.1) \end{pmatrix}.
$$

It can be verified that $B_i^\top I_{3,1} B_i = I_{3,1}$, $B_i \in O(3,1)$, $i = 1,2$. The optimization problem to compute the geodesic distance $d(B_1, B_2)$ can be solved by means of a descent procedure along the negative Euclidean gradient (9) with a line search strategy. In Figure 1, the results show that the Frobenius norm

of U is decreasing steadily when the criterion function $J(U)$ tends to be constant 0 during iteration. By the numerical experiments, we can get

$$U^* = \begin{pmatrix} 0.0000 & 0.0000 & 0.0000 & -0.4000 \\ 0.0000 & -0.0000 & -0.0000 & 0.0000 \\ 0.0000 & -0.0000 & -0.0000 & 0.0000 \\ -0.4000 & 0.0000 & 0.0000 & -0.0000 \end{pmatrix}.$$
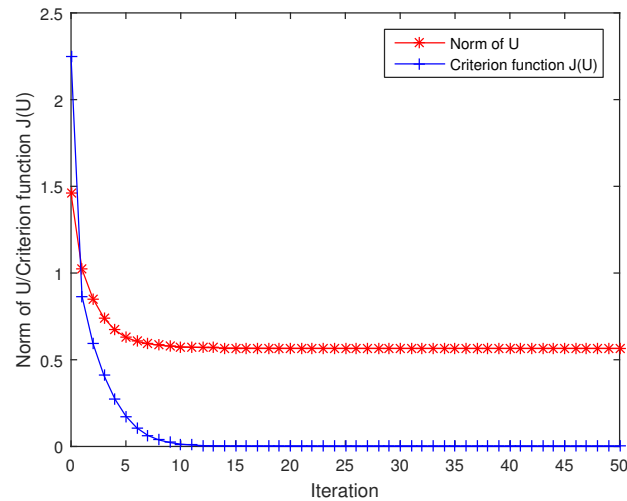

Figure 1. The criterion function $J(U)$ and the Frobenius norm of U during iteration, respectively.

It is interesting that the geodesic distance between B_1 and B_2 is $d(B_1, B_2) = \|U^*\| \approx 0.5657$, and the discrepancy between B_1 and B_2 is $D(B_1, B_2) = \|\log(B_1^{-1}B_2)\| \approx 0.5657$.

In the following experiment, we consider the average value of two points B_1 and B_2. Figure 2 displays the results of running the RSDA, the EHA, and the EGA. Note that the iteration sequence $\{A_p\}$ is expected to converge to a matrix that locates amid the samples $\{B_q\}$, and the condition $\|B_q^{-1}A_p - I_4\| < 1$ holds. The initial point is set as $A_0 = B_1$, which satisfies $\|B_2^{-1}A_0 - I_4\| \approx 0.5921 < 1$. More initial points can be chosen in a neighbor of the matrix B_1 by the following rule (25). It can be found that the averaging algorithms converge steadily and the RSDA has the fastest convergence rate among three algorithms and needs one iteration to obtain the average value as follows:

$$\bar{A} = \begin{pmatrix} 1.0453 & 0.0000 & 0.0000 & 0.3045 \\ 0.0000 & 1.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 & 0.0000 \\ 0.3045 & 0.0000 & 0.0000 & 1.0453 \end{pmatrix}.$$
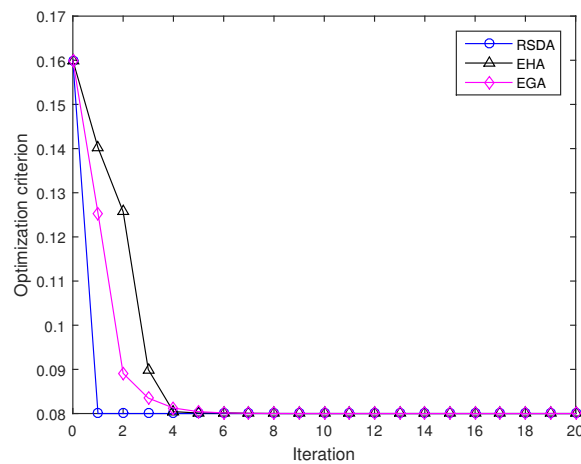

Figure 2. Convergence comparison among the Riemannian-steepest-descent algorithm (RSDA), the extended Hamiltonian algorithm (EHA) and the Euclidean gradient algorithm (EGA) during iteration.

However, the EHA and the EGA need 13 and 18 iterations to realize the same accuracy, respectively. It is interesting to compare the discrepancy $D(\bar{A}, B_1) \approx 0.2828$ with the discrepancy $D(\bar{A}, B_2) \approx 0.2828$, which confirms that the computed average value is truly a midpoint (i.e., it is located at approximately equal discrepancy from the two matrices).

4.2. Numerical Experiments on Averaging Several Lorentz Matrices

The numerical experiments rely on the availability of a way to generate pseudo-random samples on the Lorentz group. Given a point $B \in O(n, k)$ which is referred to as "center of mass" or simply center of the random distribution, it is possible to generate a random sample $B_q \in O(n, k)$ in a neighbor of a matrix B by the rule [26,27]

$$B_q = B\,e^{B^{-1}X}, \tag{24}$$

where the center matrix B can be generated by a curve departing from the identical matrix , $B = e^V$ with V generated randomly matrix in $\mathfrak{o}(n, k)$.

According to the structure of the tangent space $T_B O(n, k)$, expression (24) can be written

$$B_q = B\,e^{I_{n,k}(C - C^\top)}, \tag{25}$$

where $C \in \mathbb{R}^{(n+k) \times (n+k)}$ is any unstructured random matrix.

Figure 3 displays the results on optimization over $O(3, 1)$ with $N = 50$ Lorentz matrices. The initial point A_0 is randomly generated,
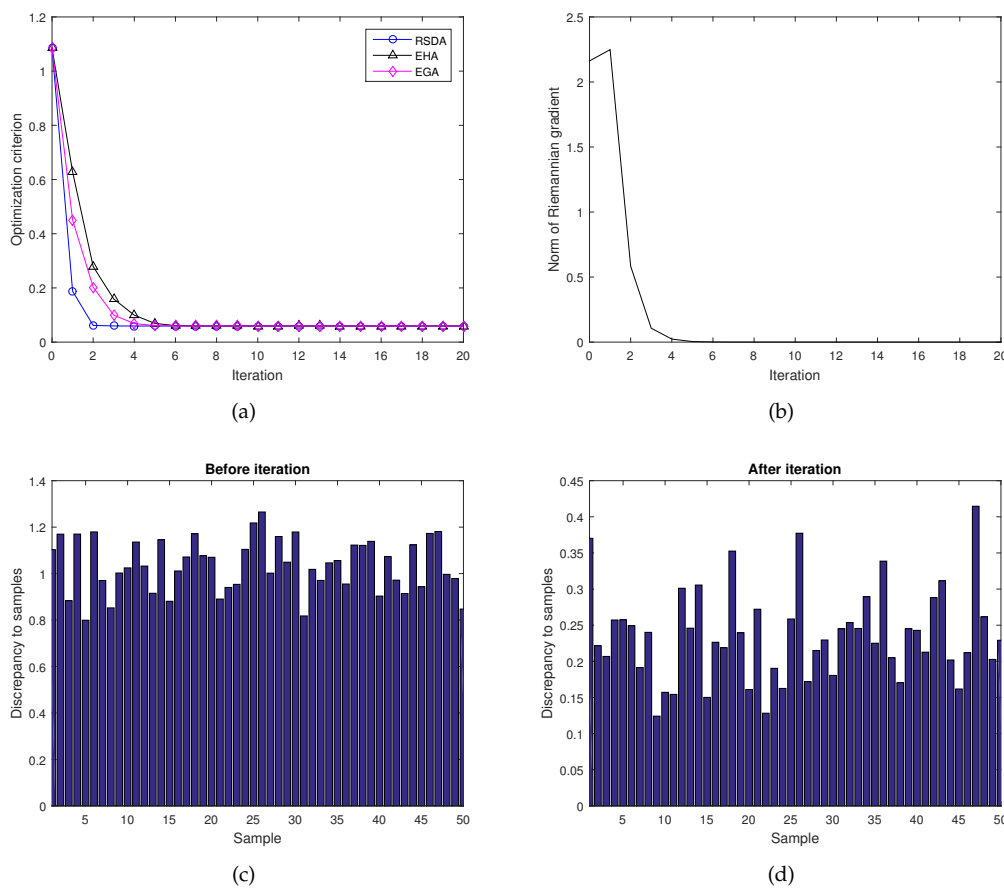
Figure 3. (**a**) convergence comparison among the RSDA, the EHA and the EGA during iteration; (**b**) the norm $\|A^{-1}\nabla_A f\|$ of the Riemannian gradient (17) during iteration by the Riemannian-steepest-descent algorithm (RSDA); (**c–d**) the discrepancy $D(A, B_q)$ before iteration ($A = A_0$) and after iteration ($A = \bar{A}$), respectively.

$$A_0 = \begin{pmatrix} 0.8412 & 0.2117 & 0.5487 & -0.2351 \\ -0.0045 & 1.0783 & -0.2160 & -0.4577 \\ -0.5410 & 0.2998 & 0.8456 & -0.3126 \\ -0.0203 & -0.5454 & -0.2508 & 1.1665 \end{pmatrix}.$$

Figure 3a shows the values of the criterion function (14) through the RSDA, the EHA and the EGA. Numerical experiments show that the average value of 50 Lorentz matrices is

$$\bar{A} = \begin{pmatrix} 0.9127 & -0.1134 & 0.3933 & 0.0235 \\ 0.0317 & 0.9853 & 0.2035 & -0.1154 \\ -0.4088 & -0.1675 & 0.8980 & -0.0407 \\ 0.0342 & -0.1087 & -0.0504 & 1.0077 \end{pmatrix},$$

which can be obtained after seven iterations using the RSDA, while the EHA and the EGA need 15 iterations. Figure 3b gives the norm $\|A^{-1}\nabla_A f\|$ of the Riemannian gradient (17) during iteration by the RSDA, from which we can obtain that the norm of the Riemannian gradient tends to be constant 0 after iteration 6. Figure 3c,d give the values of the discrepancy $D(A, B_q)$ before iteration and after iteration, respectively. In particular, note that all the discrepancy from the samples B_q and the matrix A are decreased substantially. The results show that the convergence rate of the RSDA is still the fastest among three algorithms.

5. Conclusions

In this paper, we consider the averaging optimization problem for the Lorentz group $O(n, k)$, namely, to measure the average of a set of Lorentz matrices. In order to tackle the related optimization problem, a geodesic-based Riemannian-steepest-descent algorithm is presented on the Lorentz group endowed with a left-invariant metric (Riemannian metric). The devised averaging algorithm is compared with the extended Hamiltonian algorithm and the Euclidean gradient algorithm. The results of numerical experiments over the Lorentz group $O(3, 1)$ show that the convergence rate of the geodesic-based Riemannian-steepest-descent algorithm is the best one among these three algorithms.

Author Contributions: Jing Wang conceived and designed the experiments; Jing Wang performed the experiments; Jing Wang and Huafei Sun analyzed the data; Didong Li contributed reagents/materials/analysis tools; Jing Wang, Huafei Sun and Didong Li wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Harris, W.F. Paraxial ray tracing through noncoaxial astigmatic optical systems, and a 5×5 augmented system matrix. *Optom. Vis. Sci.* **1994**, *71*, 282–285.
2. Barachant, A.; Bonnet, S.; Congedo, M.; Jutten, C. Multi-class brain computer interface classification by Riemannian geometry. *IEEE Trans. Bio. Med. Eng.* **2012**, *59*, 920–928.
3. Moakher, M. Means and averaging in the group of rotations. *SIAM J. Matrix Anal. A* **2002**, *24*, 1–16.
4. Mello, P.A. Averages on the unitary group and applications to the problem of disordered conductors. *J. Phys. A* **1990**, *23*, 4061–4080.
5. Duan, X.M.; Sun, H.F.; Peng, L.Y. Riemannian means on special Euclidean group and unipotent matrices group. *Sci. World J.* **2013**, *2013*, 292787.
6. Chakraborty, R.; Vemuri, B.C. Recursive Fréchet mean computation on the Grassmannian and its applications to computer vision. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
7. Fiori, S.; Kaneko, T.; Tanaka, T. Tangent-bundle maps on the Grassmann manifold: Application to empirical arithmetic averaging. *IEEE Trans. Signal Process.* **2015**, *63*, 155–168.
8. Kaneko, T.; Fiori, S.; Tanaka, T. Empirical arithmetic averaging over the compact Stiefel manifold. *IEEE Trans. Signal Process.* **2013**, *61*, 883–894.

9.   Pölitz, C.; Duivesteijn, W.; Morik, K. Interpretable domain adaptation via optimization over the Stiefel manifold. *Mach. Learn.* **2016**, *104*, 315–336.

10.   Fiori, S. Learning the Fréchet mean over the manifold of symmetric positive-definite matrices. *Cogn. Comput.* **2009**, *1*, 279–291.

11.   Moakher, M. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. A* **2005**, *26*, 735–747.

12.   Buchholz, S.; Sommer, G. On averaging in Clifford groups. In *Computer Algebra and Geometric Algebra with Applications*; Springer: Berlin/Heidelberg, Germany, 2005; Volume 3519, pp. 229–238.

13.   Kawaguchi, H. Evaluation of the Lorentz group Lie algebra map using the Baker-Cambell-Hausdorff formula. *IEEE Trans. Magn.* **1999**, *35*, 1490–1493.

14.   Heine, V. *Group Theory in Quantum Mechanics*; Dover: New York, NY, USA, 1993.

15.   Geyer, C.M. Catadioptric Projective Geometry: Theory and Applications. Ph.D. Thesis, University of Pennsylvania, Philadelphia, PA, USA, 2002.

16.   Fiori, S. Extended Hamiltonian learning on Riemannian manifolds: Theoretical aspects. *IEEE Trans. Neural Netw.* **2011**, *22*, 687–700.

17.   Fiori, S. Extended Hamiltonian learning on Riemannian manifolds: Numerical aspects. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *23*, 7–21.

18.   Zhang, X. *Matrix Analysis and Application*; Springer: Beijing, China, 2004.

19.   Andruchow, E.; Larotonda, G.; Recht, L.; Varela, A. The left invariant metric in the general linear group. *J. Geom. Phys.* **2014**, *86*, 241–257.

20.   Zacur, E.; Bossa, M.; Olmos, S. Multivariate tensor-based morphometry with a right-invariant Riemannian distance on $GL^+(n)$. *J. Math. Imaging Vis.* **2014**, *50*, 19–31.

21.   Goldberg, K. The formal power series for $\log(e^x e^y)$. *Duke Math. J.* **1956**, *23*, 1–21.

22.   Newman, M.; So, W.; Thompson, R.C. Convergence domains for the Campbell-Baker-Hausdorff formula. *Linear Algebra Appl.* **1989**, *24*, 30–310.

23.   Thompson, R.C. Convergence proof for Goldberg's expoential series. *Linear Algebra Appl.* **1989**, *121*, 3–7.

24.   Karcher, H. Riemannian center of mass and mollifier smoothing. *Commun. Pure Appl. Math.* **1977**, *30*, 509–541.

25.   Gabay, D. Minimizing a differentiable function over a differentiable manifold. *J. Optim. Theory App.* **1982**, *37*, 177–219.

26.   Fiori, S. Solving minimal-distance problems over the manifold of real symplectic matrices. *SIAM J. Matrix Anal. A* **2011**, *32*, 938–968.

27.   Fiori, S. A Riemannian steepest descent approach over the inhomogeneous symplectic group: Application to the averaging of linear optical systems. *Appl. Math. Comput.* **2016**, *283*, 251–264.