

Article

Empirical Means on Pseudo-Orthogonal Groups

Jing Wang ¹, Huafei Sun ² and Simone Fiori ^{3,*} ¹ School of Information, Beijing Wuzi University, Beijing 101149, China; wangjingzzumath@163.com² School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, China; huafeisun@bit.edu.cn³ Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, 60026 Ancona, Italy

* Correspondence: s.fiori@univpm.it

Received: 4 July 2019; Accepted: 1 October 2019; Published: 11 October 2019



Abstract: The present article studies the problem of computing empirical means on pseudo-orthogonal groups. To design numerical algorithms to compute empirical means, the pseudo-orthogonal group is endowed with a pseudo-Riemannian metric that affords the computation of the exponential map in closed forms. The distance between two pseudo-orthogonal matrices, which is an essential ingredient, is computed by both the Frobenius norm and the geodesic distance. The empirical-mean computation problem is solved via a pseudo-Riemannian-gradient-stepping algorithm. Several numerical tests are conducted to illustrate the numerical behavior of the devised algorithm.

Keywords: pseudo-orthogonal group; pseudo-Riemannian geometry; gradient-based function minimization on manifolds; geodesic stepping

MSC: 15A16; 15B99; 22E70; 65K10; 90C30

1. Introduction

The present article aims at investigating a least-squares problem on the pseudo-orthogonal Lie group $O(p, q)$ (we follow the nomenclature in [1]) endowed with a pseudo-Riemannian metric that affords the computation of geodesic in closed form, with the aim of computing the empirical mean value μ of a finite collection of pseudo-orthogonal matrices. The empirical mean value μ of a distribution of points on a manifold is instrumental in several applications (see, for example, the article [2] that concerns the statistics of covariance matrices for biomedical engineering applications, the article [3] that deals with urban mobility analysis and the article [4] that concerns water quality assessment). The mean value μ is, by definition, close to all points in the distribution and this makes the tangent space at μ a good candidate as a reference space for numerical calculations (see, for example, [5]) as well as empirical statistical evaluations (see, for example, [6] and, for a recent account, [7]).

Since, in general, an empirical mean of a collection of mathematical objects living in a metric space is the closest point in the space to all points in the collection, the problem of defining and computing an empirical mean may be formulated in terms of minimization of a sum-of-squared distance problem. Function minimization over smooth matrix manifolds has received considerable attention due to its broad application range [8–13]. In general, smooth manifolds in function optimization serve to conveniently represent non-linear constraints. Constrained optimization arises in several branches of science, ranging from applied mathematics [14–16] to information sciences [14,17,18]. In [19,20], optimization-based mean-computation problems over the space of the symmetric positive-definite matrices, the special Euclidean group and the space of unipotent matrices were studied. The authors of [21] investigated the problem of computing empirical arithmetic averages over the Stiefel manifolds by tangent-bundle maps. Similarly, in [22], the authors studied empirical arithmetic-mean computation

algorithms over the Grassmann manifolds. In [23], the author studied non-compact matrix-type manifolds (for example, the real symplectic group).

A special case of pseudo-orthogonal group is the Lorentz group $O(1,3)$, which describes Lorentz transformations of Minkowski space-time, the classical setting for non-gravitational physical phenomena. Lorentz groups are likewise relevant in classical ray optics [24]. Moreover, the Lorentz group is related with Poincaré groups, which may be seen as the semidirect product of Lorentz and translation groups [25]. In engineering, the Lorentz group plays a fundamental role, e.g., in the analysis of motion of charged particles in electromagnetism. For instance, the article [26] investigates on the computation of the trajectory of a charged particle in a magnetic field (such as an electron trapped in a cusp-shaped magnetic field) in terms of the Lorentz-group’s Lie-algebra map.

The present article is organized as follows. Section 2 recalls some fundamental notions on the geometry of pseudo-Riemannian manifolds and the gradient-steepest-descent function-minimization method on pseudo-Riemannian manifolds. The present article takes the route of embedding a pseudo-orthogonal group $O(p,q)$ into the Euclidean ambient space $\mathbb{R}^{(p+q)\times(p+q)}$ and to formulate all equations without recurring to any specific coordinate system nor any specific base for the associated tangent bundle. Section 3 introduces the pseudo-Riemannian geometric structure of the pseudo-orthogonal group and presents the necessary calculations to tackle function-minimization problems over such matrix group. Section 4 presents results of several numerical tests about computing a mean matrix over a pseudo-orthogonal group. Section 5 concludes the article.

2. Function Minimization on Pseudo-Riemannian Smooth Manifolds

For a certain pseudo-Riemannian manifold M endowed with a distance function $d(\cdot, \cdot)$, the empirical mean $\mu \in M$ of a finite collection of points $x_1, x_2 \dots x_N \in M$ is defined as

$$\mu := \arg \min_{x \in M} \sum_k d^2(x, x_k). \tag{1}$$

In the present section, we summarize the notion of function minimization over a smooth pseudo-Riemannian manifold from previous research endeavors. Section 2.1 recalls fundamental notions from differential geometry, while Section 2.2 summarizes the main ideas behind pseudo-gradient-based function minimization on pseudo-Riemannian manifolds.

2.1. Notes on Pseudo-Riemannian Manifolds

Let M denote a p -dimensional pseudo-Riemannian submanifold of the general linear group and let $T_x M$ denote the tangent space of M at the point x . The symbol $TM := \{(x, v) \mid x \in M, v \in T_x M\}$ denotes the tangent bundle associated to the manifold M and the symbol $\Gamma(TM)$ denotes the set of vector fields on M . The cotangent space of M at a point x is denoted by $T_x^* M$. We denote by the symbol $\langle \cdot, \cdot \rangle^E$ the Euclidean inner product.

A pseudo-Riemannian manifold is endowed with an indefinite inner product $\langle \cdot, \cdot \rangle_x : T_x M \times T_x M \rightarrow \mathbb{R}, x \in M$. An indefinite inner product is a non-degenerate, smooth, symmetric, bilinear map which assigns a real number to pairs of tangent vectors at each tangent space of a manifold. In pseudo-Riemannian geometry, the metric tensor is symmetric and invertible *but not necessarily positive-definite*.

Let $f : M \rightarrow \mathbb{R}$ denote a differentiable function. The differential of a function $f : M \rightarrow \mathbb{R}$ at a point $x \in M$ is denoted by $df_x \in T_x^* M$ and satisfies the relationship $df_x(v) = \langle \nabla_x f, v \rangle_x$, where $\nabla_x f \in T_x M$ denotes the pseudo-Riemannian gradient. Once a metric is selected, the pseudo-Riemannian gradient of a smooth function can be computed by the condition $\langle \partial_x f, v \rangle^E = \langle \nabla_x f, v \rangle_x, \forall v \in T_x M$.

The covariant derivative of a vector field $v \in \Gamma(TM)$ in the direction of a tangent vector $w \in T_x M$ is denoted as $\nabla_w v$. In the present article, the notion of connection refers to a Levi–Civita connection. The calculation of the expression of geodesics associated to a given metric on a pseudo-Riemannian manifold may be formulated in variational terms as recalled in the following

Lemma 1 ([23]). *Let $\gamma : [0, 1] \rightarrow M$ be a sufficiently regular curve with fixed endpoints, then the geodesic equation $\nabla_{\dot{\gamma}}\dot{\gamma} = 0$ is equivalent to the equation*

$$\delta \int_0^1 \langle \dot{\gamma}, \dot{\gamma} \rangle_{\gamma} dt = 0, \tag{2}$$

where δ denotes the variation of the integral functional. Namely, γ satisfies the geodesic equation if and only if it is a critical point of the energy functional (define on the space of C^1 -paths with fixed endpoints).

This is, indeed, a well-known result: although the metric is pseudo-Riemannian, its proof is the same as in the Riemannian case where there exist several references (see, e.g., [27]).

The variational formulation to compute the explicit expression of geodesics in embedded manifolds is easier to use in practical calculations than the covariant-derivative-based definition.

2.2. Gradient-Based Function Minimization on Pseudo-Riemannian Manifolds

On a geodesically complete Riemannian manifold M , the metric is positive definite, hence the squared geodesic distance connecting two points $x_1, x_2 \in M$ is defined as

$$d^2(x_1, x_2) := \left(\int_0^1 \sqrt{\langle \dot{\gamma}, \dot{\gamma} \rangle_{\gamma}} dt \right)^2 = \langle \dot{\gamma}, \dot{\gamma} \rangle_{\gamma}|_{t=0}. \tag{3}$$

If the geodesic is denoted as $\gamma_{x,v}(t)$, which indicates that the geodesic arc is departing from the point $x \in M$ with initial direction $v \in T_x M$, then the squared geodesic distance equals $\|v\|_x^2$.

When the Riemannian manifold of interest M and a regular criterion function $f : M \rightarrow \mathbb{R}$ are specified, a geodesic-based Riemannian-gradient-steepest-descent function-minimization method is expressed as

$$x_{(\ell+1)} = \gamma_{x_{(\ell)}, -\nabla_{x_{(\ell)}} f}(t_{(\ell)}), \text{ with } t_{(\ell+1)} = \arg \min_{t>0} \{f(\gamma_{x_{(\ell)}, -\nabla_{x_{(\ell)}} f}(t))\}, \tag{4}$$

where $\ell \geq 0$ denotes an iteration step-counter, $t_{(\ell)}$ denotes a step-size schedule and the initial guess $x_{(0)} \in M$ is arbitrarily chosen.

Since a pseudo-Riemannian manifold is a manifold endowed with a metric that is not necessarily positive-definite, on a pseudo-Riemannian manifold M , the quantity $\|v\|_x^2$ may be positive, negative or null even for $0 \neq v \in T_x M$. A key step in pseudo-Riemannian geometry is to decompose each tangent space $T_x M$ as

$$\begin{cases} T_x^+ M := \{v \in T_x M \mid \|v\|_x^2 > 0\}, \\ T_x^0 M := \{v \in T_x M \mid \|v\|_x^2 = 0\}, \\ T_x^- M := \{v \in T_x M \mid \|v\|_x^2 < 0\}. \end{cases} \tag{5}$$

Given a pseudo-Riemannian manifold M and a regular criterion function $f : M \rightarrow \mathbb{R}$, a “gradient steepest descent” function minimization rule is expressed as

$$\dot{x} = \begin{cases} -\nabla_x f, & \text{if } \nabla_x f \in T_x^+ M \cup T_x^0 M, \\ \nabla_x f, & \text{if } \nabla_x f \in T_x^- M, \end{cases} \tag{6}$$

which can be implemented numerically by a geodesic-based stepping method which moves a point along a short self-parallel arc in the direction of the pseudo-Riemannian gradient

$$x_{(\ell+1)} = \begin{cases} \gamma_{x_{(\ell)}, -\nabla_{x_{(\ell)}} f}(t_{(\ell)}), & \text{if } \nabla_{x_{(\ell)}} f \in T_{x_{(\ell)}}^+ M \cup T_{x_{(\ell)}}^0 M, \\ \gamma_{x_{(\ell)}, \nabla_{x_{(\ell)}} f}(t_{(\ell)}), & \text{if } \nabla_{x_{(\ell)}} f \in T_{x_{(\ell)}}^- M. \end{cases} \tag{7}$$

The step-size schedule $t_{(\ell)}$ may be defined on the basis of an optimality condition, which resembles Armijo’s line search method [28], under the assumption that $\nabla_{x_{(\ell)}} f \notin T_{x_{(\ell)}}^0 M$ (see [23]). Define the function

$$\tilde{f}_{(\ell)} := f \circ \gamma_{x_{(\ell)}, v_{(\ell)}}(t) : \mathbb{R}_+ \rightarrow \mathbb{R}, \tag{8}$$

where $v_{(\ell)} = -\nabla_{x_{(\ell)}} f$, if $\nabla_{x_{(\ell)}} f \in T_{x_{(\ell)}}^+ M$, while $v_{(\ell)} = \nabla_{x_{(\ell)}} f$, if $\nabla_{x_{(\ell)}} f \in T_{x_{(\ell)}}^- M$. The value $\tilde{f}_{(\ell)}(0) - \tilde{f}_{(\ell)}(t) \geq 0$ denotes the decrease of the criterion function f subject to a pseudo-geodesic step of size t . If the value of the geodesic step-size t is small enough, such decrease can be expanded in Taylor series as

$$\tilde{f}_{(\ell)}(0) - \tilde{f}_{(\ell)}(t) = -\tilde{f}_{1(\ell)}t - \frac{1}{2}\tilde{f}_{2(\ell)}t^2 + o(t^2), \tag{9}$$

where the coefficients $\tilde{f}_{1(\ell)}, \tilde{f}_{2(\ell)}$ of the Taylor expansion are defined as

$$\tilde{f}_{1(\ell)} = \left. \frac{d\tilde{f}_{(\ell)}}{dt} \right|_{t=0}, \quad \tilde{f}_{2(\ell)} = \left. \frac{d^2\tilde{f}_{(\ell)}}{dt^2} \right|_{t=0}. \tag{10}$$

Under such second-order Taylor approximation, the step-size value that maximizes the decrease rate is

$$t_{(\ell)} := \arg \max_t (-\tilde{f}_{1(\ell)}t - \frac{1}{2}\tilde{f}_{2(\ell)}t^2) = -\tilde{f}_{1(\ell)}\tilde{f}_{2(\ell)}^{-1}. \tag{11}$$

The above choice for the step-size schedule is optimal only if $\tilde{f}_{1(\ell)} \leq 0, \tilde{f}_{2(\ell)} > 0$, because these two conditions ensure that $t_{(\ell)} \geq 0$, hence that the algorithm is always seeking for the minimum of the criterion function f .

Given a precision value $\epsilon > 0$ and under the assumption that $\nabla_x f \notin T_x^0 M$, the numerical procedure stops at step $\bar{\ell}$ if

$$-\tilde{f}_{1(\bar{\ell})} < \epsilon. \tag{12}$$

This condition expresses the fact that the value taken by the first derivative of the function \tilde{f} is so small that no further improvements can be gained by continuing with the iteration. For the sake of consistency, it is assumed that the function f be bounded from below in M .

3. Function Minimization on the Pseudo-Orthogonal Group

Two criterion functions on the pseudo-orthogonal group that encode the notion of average squared distance may be constructed on the basis of the Frobenius norm and of the induced geodesic distance on the pseudo-orthogonal group $O(p, q)$. The related function minimization problem can be solved numerically by a pseudo-Riemannian-gradient-based algorithm.

In the present section, we deal with a matrix manifold, therefore it pays to recall that, on a matrix space, $\langle A, B \rangle^E := \text{tr}(A^\top B)$, where $^\top$ denotes matrix transpose and “tr” denotes a matrix trace operator. The matrix Frobenius norm is defined by $\|A\|_F := \sqrt{\langle A, A \rangle^E}$.

Section 3.1 surveys the geometric structure of the pseudo-orthogonal group and describes a pseudo-Riemannian metrization. Sections 3.2 and 3.3 describe two different criterion functions to compute the geodesic mean over a pseudo-orthogonal group and present the related numerical function minimization algorithms.

3.1. Pseudo-Riemannian Geometric Structure of the Pseudo-Orthogonal Group

A pseudo-orthogonal group $O(p, q)$, as a non-compact matrix Lie group [29], is an instance of quadratic groups and is defined by

$$O(p, q) := \{X \in \mathbb{R}^{(p+q) \times (p+q)} \mid X^\top R_{p,q} X = R_{p,q}\}, \quad R_{p,q} := \begin{pmatrix} I_p & O_{p \times q} \\ O_{q \times p} & -I_q \end{pmatrix}, \tag{13}$$

where the symbol I_p denotes a $p \times p$ identity matrix and the symbol $O_{p \times q}$ denotes a whole-zero $p \times q$ matrix. The matrix $R_{p,q}$ enjoys the properties $R_{p,q}^2 = I_{p+q}$, $R_{p,q}^\top = R_{p,q}^{-1} = R_{p,q}$. Pseudo-orthogonal matrices enjoy two properties that are recalled in the following.

Lemma 2. Any pseudo-orthogonal matrix $X \in O(p, q)$ is invertible. In addition, it holds that $\|X\|_F = \|X^{-1}\|_F$.

Proof. From the defining property $X^\top R_{p,q} X = R_{p,q}$, it follows that $\det(X^\top R_{p,q} X) = \det(R_{p,q})$, hence $\det(X)^2 = 1$, which proves the first assertion. From the definition of pseudo-orthogonal matrices, there follow the identities

$$X^\top = R_{p,q} X^{-1} R_{p,q}, \quad X^{-\top} = R_{p,q} X R_{p,q}, \tag{14}$$

which are often useful in the course of calculations. By the first identity in Equation (14), it is easy to see that

$$\|X\|_F^2 = \text{tr}(X X^\top) = \text{tr}(X R_{p,q} X^{-1} R_{p,q}) = \text{tr}(R_{p,q} X R_{p,q} X^{-1}),$$

by the circular shift property of the matrix trace operator. From the second identity in Equation (14), it follows that

$$\text{tr}((R_{p,q} X R_{p,q}) X^{-1}) = \text{tr}(X^{-\top} X^{-1}) = \|X^{-1}\|_F^2,$$

which proves the second assertion. \square

The tangent bundle of a pseudo-orthogonal Lie group has the structure

$$T_X O(p, q) = \{V \in \mathbb{R}^{(p+q) \times (p+q)} \mid V^\top R_{p,q} X + X^\top R_{p,q} V = 0\}, \tag{15}$$

and the tangent space at the identity of $O(p, q)$, namely the Lie algebra $\mathfrak{o}(p, q)$, has the structure

$$\mathfrak{o}(p, q) := \{H \in \mathbb{R}^{(p+q) \times (p+q)} \mid H^\top R_{p,q} + R_{p,q} H = 0\}. \tag{16}$$

By the embedding of $O(p, q)$ into the Euclidean space $\mathbb{R}^{(p+q)^2}$, the normal space at any point $X \in O(p, q)$ is defined by

$$N_X O(p, q) := \{P \in \mathbb{R}^{(p+q) \times (p+q)} \mid \text{tr}(P^\top V) = 0, \forall V \in T_X O(p, q)\}. \tag{17}$$

The tangent space, the Lie algebra and the normal space associated to the group-manifold $O(p, q)$ can be characterized as follows

$$\begin{aligned} T_X O(p, q) &= \{X R_{p,q} \Omega \mid \Omega \in \mathbb{R}^{(p+q) \times (p+q)}, \Omega^\top = -\Omega\}, \\ \mathfrak{o}(p, q) &= \{R_{p,q} \Omega \mid \Omega \in \mathbb{R}^{(p+q) \times (p+q)}, \Omega^\top = -\Omega\}, \\ N_X O(p, q) &= \{R_{p,q} X S \mid S \in \mathbb{R}^{(p+q) \times (p+q)}, S^\top = S\}. \end{aligned} \tag{18}$$

Let us consider the following indefinite inner product on the general linear group $GL(p, \mathbb{R})$

$$\langle U, V \rangle_X := \text{tr}(X^{-1} U X^{-1} V), \quad X \in GL(p, \mathbb{R}), \quad U, V \in T_X GL(p, \mathbb{R}), \tag{19}$$

referred to as Khvedelidze–Mladenov metric (see [30]). The following lemma proves two properties of such metric applied to the pseudo-orthogonal group.

Lemma 3. The Khvedelidze–Mladenov metric on the pseudo-orthogonal group $O(p, q)$ is: (i) non-degenerate; and (ii) indefinite.

Proof. Let us prove the two parts of the lemma separately.

Proof of Part (i): A metric on a finite-dimensional space is non-degenerate if and only if $\langle U, V \rangle_X = 0$ for every U implies $V = 0$. Given $X \in O(p, q)$ and $U, V \in T_X O(p, q)$, by the structure of the tangent

space $T_X O(p, q)$, it is known that $X^{-1}U = R_{p,q}\Omega$, with $\Omega^\top = -\Omega$ and $X^{-1}V = R_{p,q}\Psi$, with $\Psi^\top = -\Psi$, therefore, $\langle U, V \rangle_X = \text{tr}(R_{p,q}\Omega R_{p,q}\Psi)$. Let us define $\Omega_\star := R_{p,q}\Omega R_{p,q}$. Calculations show that $\Omega_\star^\top = R_{p,q}^\top \Omega^\top R_{p,q}^\top = R_{p,q}^\top (-\Omega) R_{p,q}^\top = -\Omega_\star$, hence, Ω_\star is a skew-symmetric matrix. The proof of the claim follows from the observation that

$$\langle U, V \rangle_X = \text{tr}(\Omega_\star \Psi) = -\langle \Omega_\star, \Psi \rangle^E, \tag{20}$$

and that the Euclidean inner product is non-degenerate on the Lie algebra of skew-symmetric matrices.

Proof of Part (ii): Given $X \in O(p, q)$ and $V \in T_X O(p, q)$, it is known that $X^{-1}V = R_{p,q}\Omega$, with $\Omega^\top = -\Omega$ and

$$R_{p,q}\Omega = \begin{pmatrix} I_p & O_{p \times q} \\ O_{q \times p} & -I_q \end{pmatrix} \begin{pmatrix} A_{p \times p} & B_{p \times q} \\ -B_{p \times q}^\top & C_{q \times q} \end{pmatrix} = \begin{pmatrix} A_{p \times p} & B_{p \times q} \\ B_{p \times q}^\top & -C_{q \times q} \end{pmatrix} \tag{21}$$

with $A^\top = -A$, $C^\top = -C$ and B arbitrary. Hence, $\|V\|_X^2 = \text{tr}((X^{-1}V)^2) = 2 \text{tr}(BB^\top) - \text{tr}(AA^\top) - \text{tr}(CC^\top)$ has indefinite sign. \square

The sign of the squared norm $\|V\|_X^2$ may be positive, negative or even zero whenever $2 \text{tr}(BB^\top) + \text{tr}(A^2) + \text{tr}(C^2) = 0$. Let us take, as a special case, the pseudo-orthogonal group $O(1, 1)$, which is the group-manifold of choice in some of the numerical examples presented in Section 4 thanks to its low dimensionality, for which the following result holds:

Lemma 4. *The Khvedelidze–Mladenov metric in Equation (19) on $O(1, 1)$ is positive-definite.*

Proof. Every element of $O(1, 1)$ can be written in one of the four forms:

$$X_1 = \begin{pmatrix} \cosh(s) & \sinh(s) \\ \sinh(s) & \cosh(s) \end{pmatrix}, \quad X_2 = \begin{pmatrix} -\cosh(s) & \sinh(s) \\ \sinh(s) & -\cosh(s) \end{pmatrix},$$

$$X_3 = \begin{pmatrix} \cosh(s) & -\sinh(s) \\ \sinh(s) & -\cosh(s) \end{pmatrix}, \quad X_4 = \begin{pmatrix} -\cosh(s) & -\sinh(s) \\ \sinh(s) & \cosh(s) \end{pmatrix},$$

where s is any real number. Moreover, the inverses of the above representations, which are necessary in the evaluation of the norms, read

$$X_1^{-1} = \begin{pmatrix} \cosh(s) & -\sinh(s) \\ -\sinh(s) & \cosh(s) \end{pmatrix}, \quad X_2^{-1} = \begin{pmatrix} -\cosh(s) & -\sinh(s) \\ -\sinh(s) & -\cosh(s) \end{pmatrix},$$

$$X_3^{-1} = \begin{pmatrix} \cosh(s) & -\sinh(s) \\ \sinh(s) & -\cosh(s) \end{pmatrix}, \quad X_4^{-1} = \begin{pmatrix} -\cosh(s) & -\sinh(s) \\ \sinh(s) & \cosh(s) \end{pmatrix}.$$

The tangent vectors corresponding to the above four representations take the form

$$V_1 = \begin{pmatrix} t \sinh(s) & t \cosh(s) \\ t \cosh(s) & t \sinh(s) \end{pmatrix} \in T_{X_1} O(1, 1), \quad V_2 = \begin{pmatrix} t \sinh(s) & -t \cosh(s) \\ -t \cosh(s) & t \sinh(s) \end{pmatrix} \in T_{X_2} O(1, 1),$$

$$V_3 = \begin{pmatrix} -t \sinh(s) & t \cosh(s) \\ -t \cosh(s) & t \sinh(s) \end{pmatrix} \in T_{X_3} O(1, 1), \quad V_4 = \begin{pmatrix} -t \sinh(s) & -t \cosh(s) \\ t \cosh(s) & t \sinh(s) \end{pmatrix} \in T_{X_4} O(1, 1),$$

where t is any real number. Hence, straightforward calculations lead to the following values for the tangent vector norms

$$\langle V_i, V_i \rangle_{X_i} = 2t^2, \quad i = 1, 2, 3, 4.$$

By direct calculations, the assertion follows. \square

It is worth remarking that, in particular, from the proof of the above lemma, it follows that $T_X^0 O(1, 1) = \{O_{2 \times 2}\}$, for every $X \in O(1, 1)$.

Under the pseudo-Riemannian metric in Equation (19), it is possible to compute the expression of geodesic over the pseudo-orthogonal group in closed form. To compute the expression of a geodesic curve on $O(p, q)$, we invoke the variational formulation recalled in Section 2.1.

Theorem 1. *The geodesic $\gamma_{X,V} : [0, 1] \rightarrow O(p, q)$, with $X \in O(p, q)$, $V \in T_X O(p, q)$ corresponding to the indefinite Khvedelidze–Mladenov metric (19) has expression*

$$\gamma_{X,V}(t) = X \exp(tX^{-1}V). \tag{22}$$

Proof. On the strength of Lemma 1, the geodesic equation expressed in variational form reads

$$\delta \int_0^1 \text{tr}(\gamma^{-1} \dot{\gamma} \gamma^{-1} \dot{\gamma}) dt = 0,$$

where the natural parametrization of the curve is assumed. By computing the variation above, we have that

$$\int_0^1 \text{tr}(\delta \gamma (\gamma^{-1} \dot{\gamma} \gamma^{-1} - \gamma^{-1} \dot{\gamma} \gamma^{-1} \dot{\gamma} \gamma^{-1})) dt = 0.$$

Since the variation $\delta \gamma \in T_\gamma O(p, q)$ is arbitrary, the sum within the innermost parentheses must belong to the normal space at γ . By the structure of the normal space $N_\gamma O(p, q)$, we have

$$\dot{\gamma} - \dot{\gamma} \gamma^{-1} \dot{\gamma} = \gamma S R_{p,q}, \quad S^\top = S.$$

The curve γ must belong entirely to the pseudo-orthogonal group, therefore $\gamma^\top R_{p,q} \gamma = R_{p,q}$. Deriving this condition twice with respect to t gives:

$$\dot{\gamma}^\top R_{p,q} \gamma + 2 \dot{\gamma}^\top R_{p,q} \dot{\gamma} + \gamma^\top R_{p,q} \ddot{\gamma} = 0.$$

Substituting $\ddot{\gamma} = \dot{\gamma} \gamma^{-1} \dot{\gamma} + \gamma S R_{p,q}$ into the equation above yields $S = 0$. Hence, the geodesic equation reads

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

Its solution, with initial conditions $\gamma(0) = X \in O(p, q)$ and $\dot{\gamma}(0) = V \in T_X O(p, q)$, is found to be $\gamma_{X,V}(t) = X \exp(tX^{-1}V)$. \square

In the above result, the symbol “exp” denotes matrix exponential, defined on the basis of a Taylor series expansion. For low dimensions $p + q$, the matrix exponential may be computed through special Rodrigues-like closed-form expressions [31].

As an essential ingredient in the formulation of a pseudo-Riemannian-gradient stepping algorithm to minimize a smooth function on a pseudo-orthogonal group, the structure of the pseudo-Riemannian gradient associated to the Khvedelidze–Mladenov metric in Equation (19) in $O(p, q)$ is given by the following.

Theorem 2. *The pseudo-Riemannian gradient of a sufficiently regular function $f : O(p, q) \rightarrow \mathbb{R}$ associated to the Khvedelidze–Mladenov metric in Equation (19) reads*

$$\nabla_X f = \frac{1}{2} (X \partial_X^\top f X - R_{p,q} \partial_X f R_{p,q}).$$

Proof. According to the relationship in Equation (19), the gradient $\nabla_X f$ is computed as the solution of the following system of equations

$$\begin{cases} \text{tr}(X^{-1}\nabla_X f X^{-1}V) = \text{tr}(\partial_X^\top f V), \quad \forall V \in T_X O(p, q), \\ \nabla_X^\top f R_{p,q} X + X^\top R_{p,q} \nabla_X f = 0. \end{cases} \tag{23}$$

Note that the first equation in Equation (23) can be rewritten as

$$\text{tr}((\partial_X^\top f - X^{-1}\nabla_X f X^{-1})V) = 0. \tag{24}$$

Since $V \in T_X O(p, q)$ is arbitrary, the condition above implies that $(\partial_X^\top f - X^{-1}\nabla_X f X^{-1})^\top \in N_X O(p, q)$, hence that $(\partial_X^\top f - X^{-1}\nabla_X f X^{-1})^\top = R_{p,q} X S$, with $S = S^\top$. Therefore, the pseudo-Riemannian gradient of the function f has the expression

$$\nabla_X f = X \partial_X^\top f X - X S R_{p,q}. \tag{25}$$

Substituting the relation in Equation (25) into the second equation of Equation (23) gives

$$S = \frac{1}{2}(R_{p,q} X^\top \partial_X f + \partial_X^\top f X R_{p,q}). \tag{26}$$

Substituting back Equation (26) into the relation in Equation (25) completes the proof. \square

A Riemannian setting for the metrization of the pseudo-orthogonal group was proposed and studied in [12]. We believe that both Riemannian and pseudo-Riemannian metrizations are worth investigating as they lead to quite different analytic results.

3.2. A Criterion Function Based on the Frobenius Norm over $O(p, q)$

In the present article, the pseudo-orthogonal group is treated as a pseudo-Riemannian manifold. Although it is possible to introduce a pseudo-distance function that is compatible with the pseudo-Riemannian metric, such function is not positive definite and cannot be interpreted as a distance function.

Therefore, as a first attempt in the construction of a criterion function to define an empirical mean, we consider the distance function on the pseudo-orthogonal group suggested in the research work [32], which is defined as

$$D_{\mathbb{F}}^2(X, Y) := \|X - Y\|_{\mathbb{F}}^2 = \text{tr}((X - Y)^\top (X - Y)), \quad X, Y \in O(p, q). \tag{27}$$

The criterion function $f : O(p, q) \rightarrow \mathbb{R}$ to be minimized to compute an average point out of a collection $\{X_1, X_2, \dots, X_N\}$ of $O(p, q)$ -samples is

$$f(X) := \frac{1}{2N} \sum_k \|X - X_k\|_{\mathbb{F}}^2 \quad X, X_k \in O(p, q), \quad k = 1, 2, \dots, N. \tag{28}$$

For the sake of notational convenience, set $C := \frac{1}{N} \sum_k X_k$, which is the empirical arithmetic average of the collection of points. The criterion function in Equation (28) can be recast as $f(X) = \frac{1}{2} \|X - C\|_{\mathbb{F}}^2 + \text{constant}$. On the basis of such expression, it is straightforward to verify that

$$\begin{aligned} df(X) &= d\left(\frac{1}{2} \text{tr}((X - C)^\top (X - C))\right) = \frac{1}{2} \text{tr}(dX^\top (X - C) + (X - C)^\top dX) \\ &= \text{tr}((X - C)^\top dX) = \langle X - C, dX \rangle^E, \end{aligned} \tag{29}$$

therefore the Euclidean gradient of the function f with respect to X is given by $\partial_X f = X - C$. According to Theorem 2, the pseudo-Riemannian gradient of the criterion function in Equation (28) on a pseudo-orthogonal group endowed with the Khvedelidze–Mladenov metric is given by

$$\nabla_X f = \frac{1}{2} X(X - C)^\top X - \frac{1}{2} R_{p,q}(X - C)R_{p,q}. \tag{30}$$

The double squared pseudo-Riemannian norm of the pseudo-Riemannian gradient $\nabla_X f$ reads

$$\begin{aligned} 2\langle \nabla_X f, \nabla_X f \rangle_X &= 2 \operatorname{tr} \left((X^{-1} \nabla_X f)^2 \right) \\ &= \operatorname{tr} \left((X - C)^\top X(X - C)^\top X - (X - C)^\top R_{p,q}(X - C)R_{p,q} \right). \end{aligned} \tag{31}$$

To prove the consistency of the pseudo-Riemannian function minimization algorithm with the function minimization problem at hand, it is necessary to evaluate the sign of the coefficients of the step-size schedule, as discussed in Section 2.2.

Lemma 5. *The coefficients \tilde{f}_1 and \tilde{f}_2 of the function $\tilde{f} = f \circ \gamma_{X,V}(t)$ with $(X, V) \in TO(p, q)$ are given by*

$$\tilde{f}_1 = \operatorname{tr}((X - C)^\top V), \quad \tilde{f}_2 = \operatorname{tr}(VV^\top + (X - C)^\top VX^{-1}V). \tag{32}$$

Proof. Since $\tilde{f} = \frac{1}{2} \|X \exp(tX^{-1}V) - C\|_F^2 + \text{constant}$, it holds that

$$\begin{aligned} \tilde{f}_1 &= \left. \frac{d\tilde{f}}{dt} \right|_{t=0} = \operatorname{tr} \left((X \exp(tX^{-1}V) - C)^\top V \exp(tX^{-1}V) \right) \Big|_{t=0}, \\ \tilde{f}_2 &= \left. \frac{d^2\tilde{f}}{dt^2} \right|_{t=0} = \|V \exp(tX^{-1}V)\|_F^2 \Big|_{t=0} + \operatorname{tr} \left((X \exp(tX^{-1}V) - C)^\top VX^{-1}V \exp(tX^{-1}V) \right) \Big|_{t=0}, \end{aligned}$$

which proves the assertion. \square

The following two lemmas examine the signs of the coefficients \tilde{f}_1 and \tilde{f}_2 .

Lemma 6. *The coefficient \tilde{f}_1 in Equation (32) is non-positive.*

Proof. In the case that $\nabla_X f \in T_X^+ O(p, q) \cup T_X^0 O(p, q)$, the algorithm in Equation (7) takes $V = -\nabla_X f$, hence, by Equation (31),

$$\begin{aligned} \tilde{f}_1 &= -\operatorname{tr}((X - C)^\top \nabla_X f) \\ &= -\frac{1}{2} \operatorname{tr} \left((X - C)^\top X(X - C)^\top X - (X - C)^\top R_{p,q}(X - C)R_{p,q} \right) \\ &= -\operatorname{tr} \left((X^{-1} \nabla_X f)^2 \right) \leq 0. \end{aligned} \tag{33}$$

Conversely, when $\nabla_X f \in T_X^- O(p, q)$, the algorithm in Equation (7) takes $V = \nabla_X f$, hence

$$\tilde{f}_1 = \operatorname{tr} \left((X^{-1} \nabla_X f)^2 \right) = \langle \nabla_X f, \nabla_X f \rangle_X < 0, \tag{34}$$

which proves the assertion. \square

Lemma 7. *Fixing $\|C\|_F$, for $\|X - C\|_F$ sufficiently small, the coefficient \tilde{f}_2 in Equation (32) is positive.*

Proof. The coefficient \tilde{f}_2 is computed as the sum of two terms, $\|V\|_F^2$ and $\operatorname{tr}((X - C)^\top VX^{-1}V)$. The first term is nonnegative for every $V \in T_X O(p, q)$, while the second term is indefinite. Note that $\|X^{-1}\|_F = \|X\|_F$, therefore we have that

$$\begin{aligned}
 |\operatorname{tr}((X - C)^\top V X^{-1} V)| &\leq \|X - C\|_F \|V X^{-1} V\|_F \\
 &\leq \|X - C\|_F \|V\|_F^2 \|X\|_F \\
 &\leq \|X - C\|_F \|V\|_F^2 (\|X - C\|_F + \|C\|_F) \\
 &= \|V\|_F^2 (\|X - C\|_F^2 + \|X - C\|_F \|C\|_F).
 \end{aligned}
 \tag{35}$$

As a consequence, fixing $\|C\|_F$, for $\|X - C\|_F$ sufficiently small, the coefficient \tilde{f}_2 is non-negative. \square

A consequence of Lemma 7 is that the initial point $X_{(0)}$ may be chosen or randomly generated in $O(p, q)$, provided it meets the condition

$$\operatorname{tr}(\nabla_{X_{(0)}}^\top f \nabla_{X_{(0)}} f + (X_{(0)} - C)^\top \nabla_{X_{(0)}} f X_{(0)}^{-1} \nabla_{X_{(0)}} f) > 0.
 \tag{36}$$

The proposed procedure to minimize the criterion function in Equation (28) can be summarized by the pseudo-code listed in Algorithm 1, where it is assumed that the sequence $\ell \rightarrow X_{(\ell)}$ satisfies $\nabla_{X_{(\ell)}} f \notin T_{X_{(\ell)}}^0 O(p, q)$. In Algorithm 1, the quantity ℓ denotes a step counter, the matrix $J_{(\ell)}$ represents the Euclidean gradient of the criterion function in Equation (28), the matrix $U_{(\ell)}$ represents its pseudo-Riemannian gradient and the sign of the scalar quantity $s_{(\ell)}$ determines whether the matrix $U_{(\ell)}$ belongs to the space $T_{X_{(\ell)}}^+ O(p, q)$ or to the space $T_{X_{(\ell)}}^- O(p, q)$.

Algorithm 1 Pseudocode to implement mean-computation over $O(p, q)$ according to the function minimization rule (7) endowed with the step-size-selection rule in Equation (11) and the stopping criterion in Equation (12).

```

Set  $R_{p,q} = \begin{pmatrix} I_p & O_{p \times q} \\ O_{q \times p} & -I_q \end{pmatrix}$ 
Set  $\ell = 0$ 
Set  $C = \frac{1}{N} \sum_k X_k$ 
Set  $X_{(0)}$  to an initial point in  $O(p, q)$ 
Set  $\epsilon$  to desired precision
repeat
  Compute  $J_{(\ell)} = X_{(\ell)} - C$ 
  Compute  $U_{(\ell)} = \frac{1}{2}(X_{(\ell)} J_{(\ell)}^\top X_{(\ell)} - R_{p,q} J_{(\ell)} R_{p,q})$ 
  Compute  $s_{(\ell)} = \operatorname{tr}((X_{(\ell)}^{-1} U_{(\ell)})^2)$ 
  if  $s_{(\ell)} > 0$  then
    Set  $V_{(\ell)} = -U_{(\ell)}$ 
  else
    Set  $V_{(\ell)} = U_{(\ell)}$ 
  end if
  Compute  $\tilde{f}_{1(\ell)} = \operatorname{tr}(J_{(\ell)}^\top V_{(\ell)})$ 
  Compute  $\tilde{f}_{2(\ell)} = \operatorname{tr}(V_{(\ell)}^\top V_{(\ell)}) + \operatorname{tr}(J_{(\ell)}^\top V_{(\ell)} X_{(\ell)}^{-1} V_{(\ell)})$ 
  Set  $\hat{t}_{(\ell)} = -\tilde{f}_{1(\ell)} / \tilde{f}_{2(\ell)}$ 
  Set  $X_{(\ell+1)} = X_{(\ell)} \exp(\hat{t}_{(\ell)} X_{(\ell)}^{-1} V_{(\ell)})$ 
  Set  $\ell = \ell + 1$ 
until  $-\tilde{f}_{1(\ell)} < \epsilon$ 

```

3.3. A Criterion Function Based on the Geodesic Distance over $O(p, q)$

We may consider a second instance of distance between two points in the group-manifold $O(p, q)$ defined as follows

$$D_g^2(X, Y) := \text{tr} |\log^2(X^{-1}Y)|, \quad X, Y \in O(p, q), \tag{37}$$

where the symbol $|\cdot|$ denotes the entry-wise absolute value of the argument matrix.

On the basis of the above distance function, the criterion function $f : O(p, q) \rightarrow \mathbb{R}$ to be minimized in the context of computing a mean matrix out of a set of pseudo-orthogonal matrix-samples is defined by

$$f(X) := \frac{1}{2N} \sum_k D_g^2(X, X_k) \quad X, X_k \in O(p, q), \quad k = 1, 2, \dots, N. \tag{38}$$

Let us fix an element $Y \in O(p, q)$ and compute the pseudo-Riemannian gradient of the map $X \rightarrow \tilde{D}_g^2(X, Y)$, where we define an auxiliary function as $\tilde{D}_g^2(X, Y) := \text{tr}(\log^2(X^{-1}Y))$. According to Proposition 2.1 in [20], the differential of the auxiliary function may be written as

$$\begin{aligned} d\tilde{D}_g^2(X, Y) &= \text{tr}(d \log^2(X^{-1}Y)) \\ &= 2 \text{tr}(\log(X^{-1}Y)(X^{-1}Y)^{-1}d(X^{-1}Y)) \\ &= -2 \text{tr}(X^{-1}Y \log(X^{-1}Y)Y^{-1}dX) \\ &= -2 \text{tr}(\log(X^{-1}Y)X^{-1}dX) \\ &= 2 \text{tr}(\log(Y^{-1}X)X^{-1}dX) \\ &= \langle (2 \log(Y^{-1}X)X^{-1})^\top, dX \rangle^E. \end{aligned} \tag{39}$$

Therefore, the Euclidean gradient of the auxiliary function is given by

$$\partial_X^\top \tilde{D}_g^2(X, Y) = 2 \log(Y^{-1}X)X^{-1}.$$

According to Theorem 2, the following expression for the pseudo-Riemannian gradient of the distance function in Equation (37) is readily obtained

$$\nabla_X D_g^2(X, Y) = \begin{cases} X \log(Y^{-1}X) - XR_{p,q} \log^\top(Y^{-1}X)R_{p,q}, & \text{for } \tilde{D}_g^2(X, Y) \geq 0, \\ XR_{p,q} \log^\top(Y^{-1}X)R_{p,q} - X \log(Y^{-1}X), & \text{for } \tilde{D}_g^2(X, Y) < 0. \end{cases} \tag{40}$$

The proposed procedure to minimize the criterion function (38) is summarized by the pseudocode listed in Algorithm 2, where the notation is the same as in Algorithm 1. In this empirical mean computation algorithm, a fixed step-size η has been selected, as opposed to Algorithm 1 that utilizes a variable step-size schedule.

Algorithm 2 Pseudocode to implement mean-computation over $O(p, q)$ according to the function minimization rule in Equation (7).

Set $R_{p,q} = \begin{pmatrix} I_p & O_{p \times q} \\ O_{q \times p} & -I_q \end{pmatrix}$
 Set $\ell = 0$
 Set $X_{(0)}$ to an initial point in $O(p, q)$
 Set η to a step-size value
repeat
 Compute $J_{(\ell)} = \frac{1}{N} \sum_{k=1}^N (\log(X_k^{-1} X_{(\ell)}) X_{(\ell)}^{-1})^\top \text{sign tr}(\log^2(X_{(\ell)}^{-1} X_k))$
 Compute $U_{(\ell)} = \frac{1}{2} (X_{(\ell)} J_{(\ell)}^\top X_{(\ell)} - R_{p,q} J_{(\ell)} R_{p,q})$
 Compute $s_{(\ell)} = \text{tr}((X_{(\ell)}^{-1} U_{(\ell)})^2)$
 if $s_{(\ell)} > 0$ **then**
 Set $V_{(\ell)} = -U_{(\ell)}$
 else
 Set $V_{(\ell)} = U_{(\ell)}$
 end if
 Set $X_{(\ell+1)} = X_{(\ell)} \exp(\eta X_{(\ell)}^{-1} V_{(\ell)})$
 Set $\ell = \ell + 1$
until $X_{(\ell)}$ is close enough to a critical point of f

4. Numerical Tests

The present section shows the results of several numerical tests effected on the iterative algorithm in Equation (7). Section 4.1 shows results about gradient-based minimization of a criterion function induced by the Frobenius norm. Section 4.2 shows results about the gradient-based minimization of a criterion function induced by the geodesic distance.

The numerical experiments rely on the availability of a way to generate pseudo-random samples on the pseudo-orthogonal group. Given a point $X \in O(p, q)$, which is referred to in the following as “center of mass” or simply center of the random distribution, it is possible to generate a random sample $Y \in O(p, q)$ in a neighbor of a matrix X by the rule

$$Y = X \exp(X^{-1}V), \tag{41}$$

where the direction $V \in T_X O(p, q)$ is randomly generated around $0 \in T_X O(p, q)$. A set of points Y generated by the exponential rule in Equation (41) distributes around the point X , as confirmed by the following.

Lemma 8. *Let $(X, V) \in TO(p, q)$ and define $Y := X \exp(X^{-1}V)$. Then, it holds that $\|Y - X\|_F \leq \|X\|_F [\exp(\|X^{-1}V\|_F) - 1]$.*

According to the structure of the tangent space $T_X O(p, q)$, Equation (41) can be written $Y = X \exp(R_{p,q}(A - A^\top))$, where $A \in \mathbb{R}^{(p+q) \times (p+q)}$ has zero-mean randomly-generated entries. Note that $\|R_{p,q}(A - A^\top)\|_F^2 = 2(\|A\|_F^2 - \text{tr}(A^2))$, hence, by Lemma 8, it holds that

$$\|Y - X\|_F \leq \|X\|_F [\exp(\sqrt{2(\|A\|_F^2 - \text{tr}(A^2))}) - 1].$$

Namely, the variance of the entries a_{ij} of A controls the spread of the random pseudo-orthogonal sample-matrices Y around the center of mass X .

4.1. Gradient-Based Minimization of a Criterion Function Induced by the Frobenius Norm

The computation of the mean matrix from a dataset of matrices belonging to the pseudo-orthogonal group $O(1, 1)$ is tackled as a first test. Further numerical tests focus on the computation of the empirical mean of a set of pseudo-orthogonal matrices in $O(p, q)$ with $p > 1$ and $q > 1$. All tests in the present section refer to Algorithm 1.

Numerical experiment A: Mean matrix computation on the pseudo-orthogonal group $O(1, 1)$. From Lemma 4, we know that every element of the group $O(1, 1)$ can be written in one of four canonical forms: here the first form is studied. This choice was motivated by the observation that the elements of the group $O(1, 1)$ can be rendered graphically on a 2-dimensional drawing. Figure 1 shows the location of $N = 50$ samples X_k (circles) generated randomly around a randomly-selected center of mass (cross), the trajectory of the function minimization algorithm over $O(1, 1)$ (solid-dotted line), and the location of the final point computed by the algorithm (diamond). To emphasize the behavior of the function minimization method in Equation (7), in the present experiment, a constant step-size schedule $t_{(\ell)} = \frac{1}{2}$ is selected and no stopping criterion is made use of. Figure 1 confirms that the numerical function minimization algorithm is convergent toward the center of mass. (Because of finite sample size, the empirical center of mass differs from the actual one.) Figure 2 shows the values of the criterion function $\frac{1}{2N} \sum_k D_F^2(X, X_k)$ and of the Frobenius norm of its pseudo-Riemannian gradient during iteration, and the distances $D_F(X, X_k)$ before iteration (with initial point chosen as $X_{(0)} = I_2$) and after iteration. In particular, all the distances from the samples X_k and the matrix X decreased substantially.

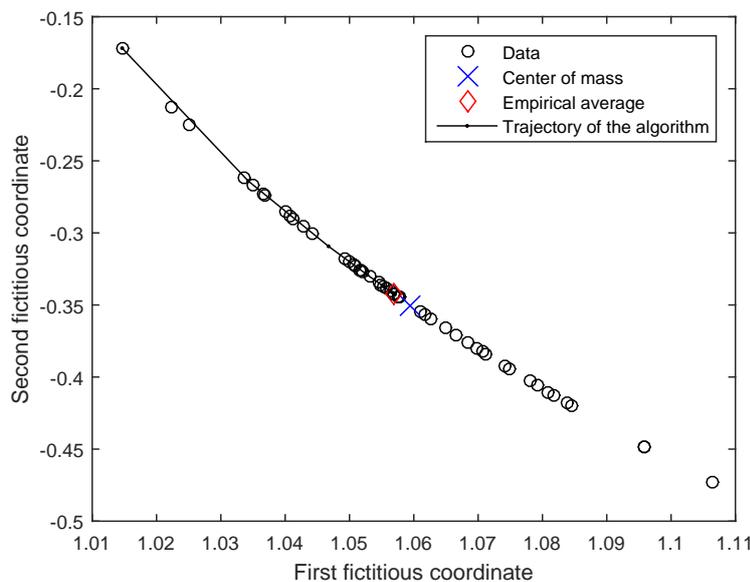


Figure 1. Numerical test of function minimization applied to the criterion function in Equation (28) over the pseudo-orthogonal group $O(1, 1)$: Data set and result of mean-matrix computation. (To obtain an easy-to-read graph, we chose to represent a pseudo-orthogonal matrix $X \in O(1, 1)$, which is a 2×2 matrix, by two fictitious coordinates, namely the entry $X^{(1,1)}$ as first coordinate and the entry $X^{(1,2)}$ as second coordinate.)

Numerical experiment B: Mean matrix computation on the pseudo-orthogonal group $O(p, q)$ with $p, q > 1$. Figure 3 illustrates a result obtained with the iterative algorithm in Equation (7) when $p = 5, q = 5$ and $N = 50$. Figure 3 shows the values of the criterion function $\frac{1}{2N} \sum_k D_F^2(X, X_k)$ and of the Frobenius norm of its pseudo-Riemannian gradient during iteration. Such numerical results were obtained by the adaptive step-size schedule explained in Section 2.2, the stopping condition in Equation (12) with precision $\epsilon = 10^{-6}$ and an initial point chosen as $X_{(0)} = I_{10}$. Figure 3 also displays the distances $D_F(X, X_k)$ before iteration (with initial point chosen as $X_{(0)} = I_{10}$) and after iteration. Figure 4 shows

the value of the index $\|X_{(\ell)}^T R_{p,q} X_{(\ell)} - R_{p,q}\|_F$ during iteration as well as the value of the step-size schedule. In Figure 4, the values of the coefficients \tilde{f}_1 and \tilde{f}_2 during iteration are displayed as well.

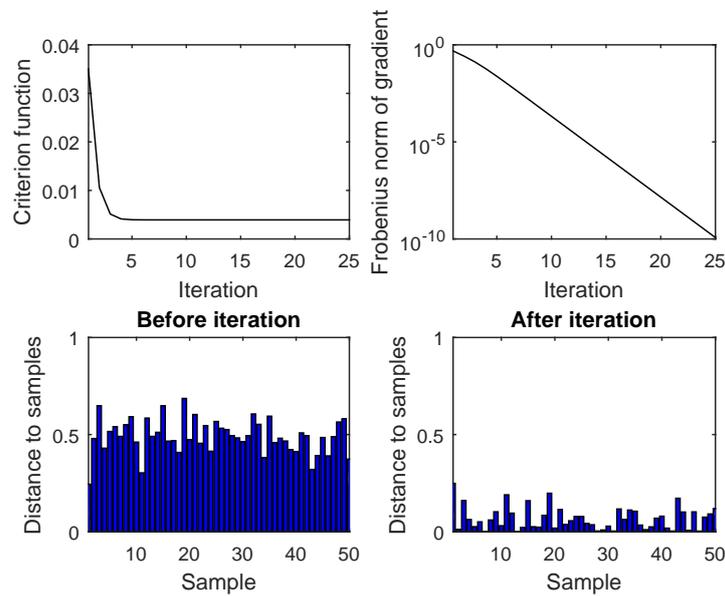


Figure 2. Mean-matrix computation over the pseudo-orthogonal group $O(1,1)$: Values of the criterion function in Equation (28) during iteration, Frobenius norm of its pseudo-Riemannian gradient, distances $D_F(X, X_k)$ before iteration ($X = X_{(0)}$) and after iteration ($X = \mu$).

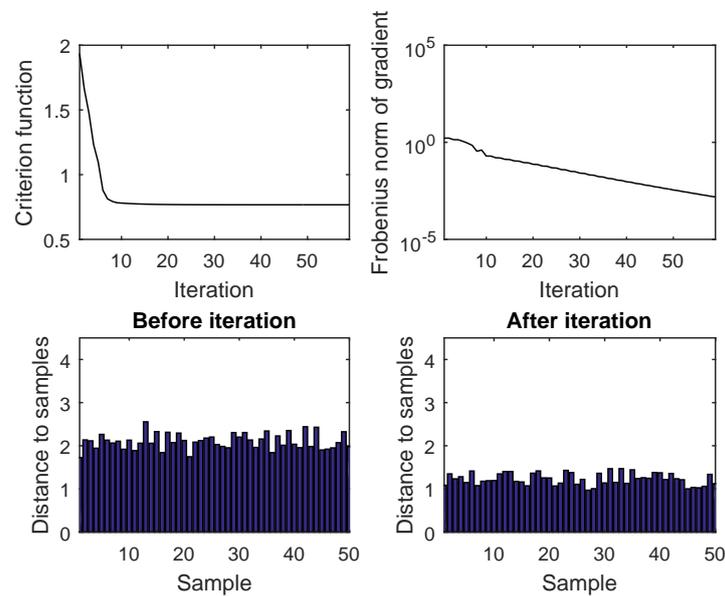


Figure 3. Mean-matrix computation over the pseudo-orthogonal group $O(5,5)$: Values of the criterion function in Equation (28) during iteration, Frobenius norm of its pseudo-Riemannian gradient, distances $D_F(X, X_k)$ before iteration ($X = X_{(0)}$) and after iteration ($X = \mu$).

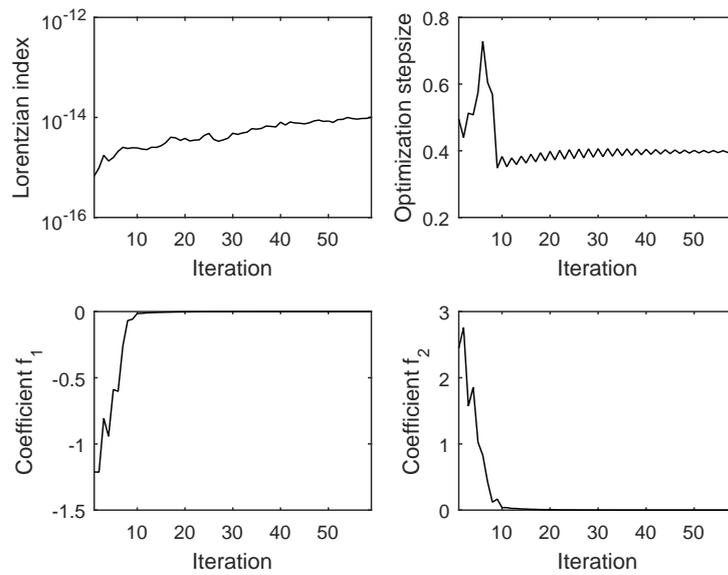


Figure 4. Minimization of the criterion function in Equation (28) over the pseudo-orthogonal group $O(5,5)$: Values of the index $\|X_{(\ell)}^T R_{p,q} X_{(\ell)} - R_{p,q}\|_F$ and of the coefficients $\hat{t}_{(\ell)}$, $\tilde{f}_{1(\ell)}$ and $\tilde{f}_{2(\ell)}$ during iteration.

Figures 5 and 6 illustrate the numerical results obtained by means of the iterative algorithm in Equation (7) for $p = 10, q = 5$ and $N = 50$. As may be readily observed, with the growth of the dimension of the underlying manifold, the minimization problems related to the search for a mean matrix becomes increasingly harder to solve. Compared to the results shown in Figures 3 and 4, the minimization algorithm takes a larger number of iterations to achieve comparable precision levels.

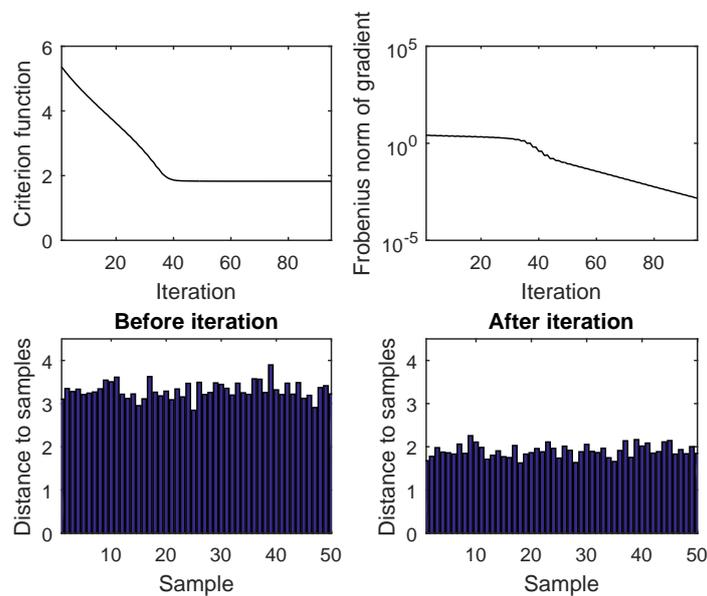


Figure 5. Mean-matrix computation over over the pseudo-orthogonal group $O(10,5)$: Values of the criterion function in Equation (28) during iteration, Frobenius norm of its pseudo-Riemannian gradient, distances $D_F(X, X_k)$ before iteration ($X = X_{(0)}$) and after iteration ($X = \mu$).

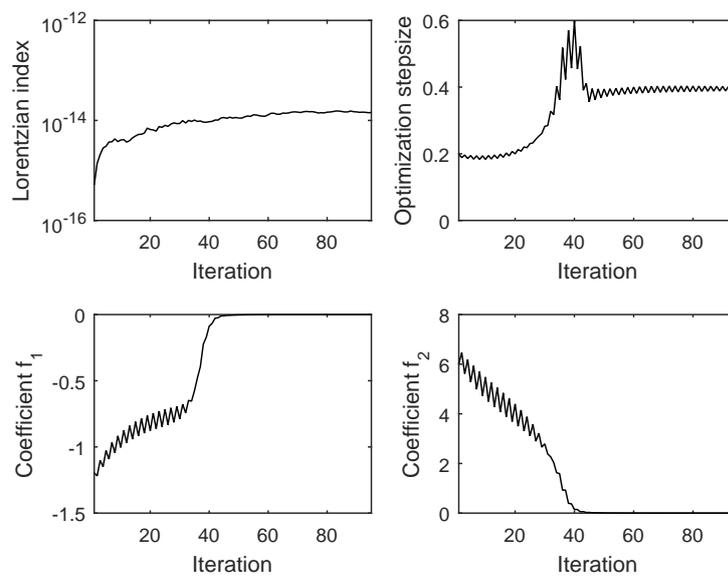


Figure 6. Minimization of the criterion function in Equation (28) over the pseudo-orthogonal group $O(10,5)$: Values of the index $\|X_{(\ell)}^T R_{p,q} X_{(\ell)} - R_{p,q}\|_F$ and of the coefficients $\hat{f}_{(\ell)}$, $\tilde{f}_{1(\ell)}$ and $\tilde{f}_{2(\ell)}$ during iteration.

Figure 7 shows the result of an empirical statistical analysis about mean-matrix computation over the pseudo-orthogonal group $O(3,1)$ on 500 independent trials. In each trial, the algorithm starts from a randomly generated initial point. In particular, Figure 7 displays the distribution of the number of iterations that the numerical minimization algorithm takes to achieve the desired precision level $\epsilon = 10^{-6}$ on each trial. The convergence speed varies with the initial point while the algorithm converges in every trial to the same value of the criterion function (namely, to about 0.0622 in the current test). The largest number of trials converge in about 20–30 iterations.

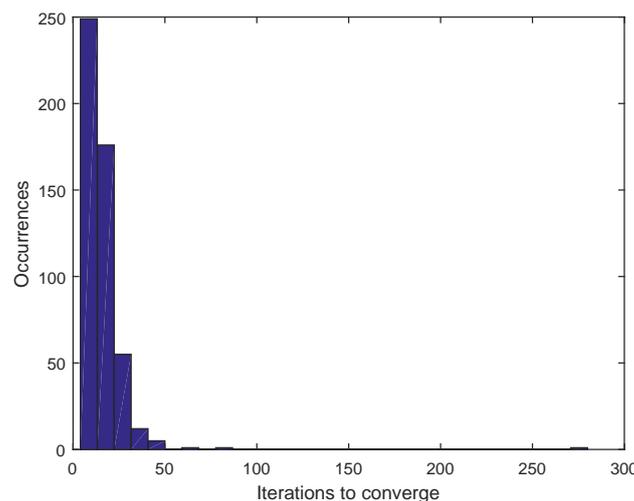


Figure 7. Minimization of the criterion function in Equation (28) over the pseudo-orthogonal group $O(3,1)$: Distribution of the number of iterations to converge on each trial on a total of 500 independent trials.

4.2. Gradient-Based Minimization of a Criterion Function Induced by the Geodesic Distance

In the following experiments, a constant step-size $\eta = \frac{1}{2}$ was chosen and, to evidence the numerical stability of the numerical minimization method, no stopping criterion was used. In this

set of experiments, empirical means are sought for within the pseudo-orthogonal groups $O(1, 1)$ and $O(5, 5)$ by means of Algorithm 2.

Numerical experiment C: Mean matrix computation on the pseudo-orthogonal group $O(1, 1)$. Figure 8 shows the results of optimization over the low-dimensional pseudo-orthogonal group $O(1, 1)$ rendered on a 2-dimensional drawing. Figure 8 shows the location of 50 samples (or target matrices) X_k (circles) generated randomly around a randomly-selected center of mass (cross), the location of the final point of the iteration, namely, the inferred empirical mean (diamond) and the trajectory of the function minimization algorithm over $O(1, 1)$ (solid-dotted line). Figure 9 shows that the algorithm is convergent toward the center of mass. (Again, because of finite sample size, the empirical mean differs from the actual center of mass.) Figure 9 shows the values of the criterion function $\frac{1}{2N} \sum_k D_g^2(X, X_k)$, the pseudo-norm of its pseudo-Riemannian gradient during iteration and the distances $D_g(X, X_k)$ before iteration (with initial point chosen as $X_{(0)} = I_2$) and after iteration.

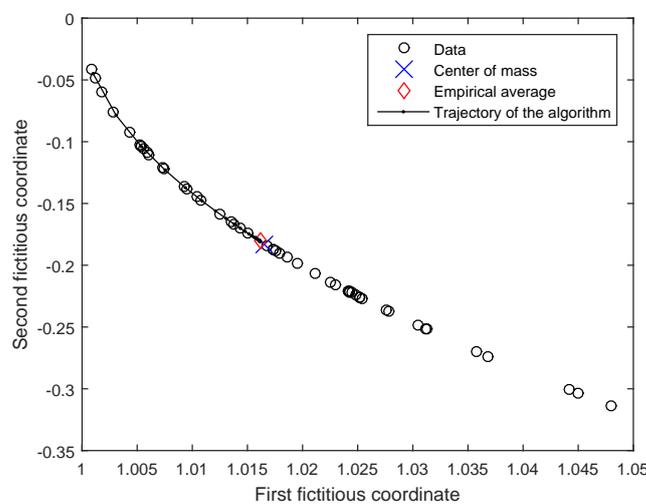


Figure 8. Minimization of the criterion function in Equation (38) over the pseudo-orthogonal group $O(1, 1)$: Data set and result of mean-matrix computation.

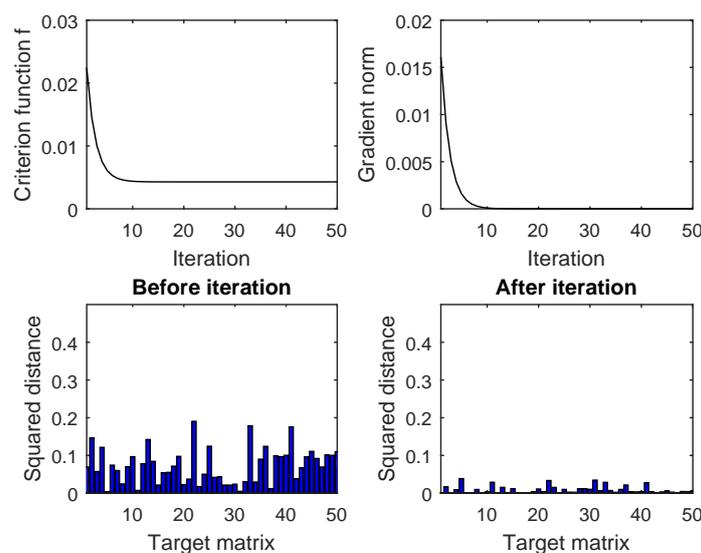


Figure 9. Mean-matrix computation over over the pseudo-orthogonal group $O(1, 1)$: Values of the criterion function in Equation (38) during iteration, of the pseudo-norm of its pseudo-Riemannian gradient, of the distances $D_g(X, X_k)$ before iteration ($X = X_{(0)}$) and after iteration ($X = \mu$).

The numerical function-minimization algorithm converges steadily and quickly to the minimum value of the criterion function and keeps in the vicinity of the empirical mean after convergence.

Numerical experiment D: Mean matrix computation on the pseudo-orthogonal group $O(p, q)$ with $p, q > 1$. Figure 10 shows a result obtained with the iterative algorithm (7) for $p = q = 5$ and $N = 50$. Figure 10 shows the values of the criterion function $\frac{1}{2N} \sum_k D_g^2(X, X_k)$ and the pseudo-norm of its pseudo-Riemannian gradient during iteration, and the distances $D_g(X, X_k)$ before iteration (with initial point chosen as $X_{(0)} = I_{10}$) and after iteration. The results in Figure 10 confirm that the numerical function minimization algorithm converges steadily toward the minimal distance configuration. In fact, the distances from the found empirical mean are much smaller than the distances from the initial point.

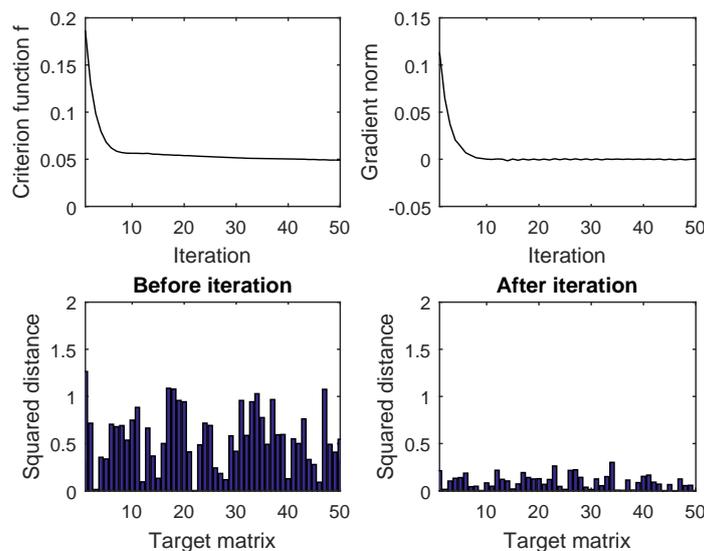


Figure 10. Mean-matrix computation over the pseudo-orthogonal group $O(5, 5)$: Values of the criterion function in Equation (38) during iteration, of the pseudo-norm of its pseudo-Riemannian gradient, of the distances $D_g(X, X_k)$ before iteration ($X = X_{(0)}$) and after iteration ($X = \mu$).

5. Conclusions

The present article investigated the computation of the empirical mean from a collection of pseudo-orthogonal matrices.

The pseudo-orthogonal group is regarded as a pseudo-Riemannian manifold and a metric is chosen that affords the computation of geodesic arcs in closed form as well as the pseudo-Riemannian gradient of a smooth function. Within such geometric setting, we consider a minimal-distance problem based on the Frobenius norm and on an induced geodesic distance. The related function minimization problem can be solved by a pseudo-Riemannian-gradient-based algorithm. Unlike the geodesic-based Riemannian gradient-steepest-descent function minimization method (see [11]), which converges under appropriate conditions, the geodesic-based stepping method (7) is not necessarily convergent, in general, due to the presence of the tangent-space component $T_x^0 M$. However, the function minimization method (7), endowed with the step-size selection method (11), is shown numerically to enjoy favorable convergence properties in all numerical tests.

In this article, we present numerical tests based on two criterion functions. To illustrate the numerical features of the proposed function-minimization method, several indicators are defined and their values are evaluated on a number of case studies over the pseudo-orthogonal group in low as well as high dimension.

Author Contributions: Conceptualization, J.W. and H.S.; methodology, H.S.; software, S.F. and J.W.; validation, J.W., H.S. and S.F.; formal analysis, J.W.; writing—original draft preparation, J.W. and H.S.; writing—review and editing, S.F.; funding acquisition, H.S. and S.F.

Funding: This research was funded by the National Natural Science Foundations of China (Grant No. 61179031) and by the Progetto di Ricerca Scientifica di Ateneo UNIVPM 2014 (RSA-A) “Elaborazione adattativa di segnali strutturati”.

Acknowledgments: S.F. wishes to gratefully thank Huafei Sun for his invitation to spend two weeks at the Beijing Institute of Technology in September 2015, during which part of the present research was conducted.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. O'Neill, B. *Semi-Riemannian Geometry. With Applications to Relativity. Pure and Applied Mathematics, 103*; Academic Press, Inc.: New York, NY, USA, 1983.
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. [[CrossRef](#)] [[PubMed](#)]
3. Gariazzo, C.; Pelliccioni, A.; Bogliolo, M.P. Spatiotemporal analysis of urban mobility using aggregate mobile phone derived presence and demographic data: A case study in the city of Rome, Italy. *Data* **2019**, *4*, 8. [[CrossRef](#)]
4. Zavareh, M.; Maggioni, V. Application of rough set theory to water quality analysis: A case study. *Data* **2018**, *3*, 50. [[CrossRef](#)]
5. Sen, S.K. Classification on Manifolds. Ph.D. Thesis, Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 2008.
6. Bhattacharya, A.; Bhattacharya, R. *Nonparametric Statistics on Manifolds with Applications to Shape Spaces. Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*; Institute of Mathematical Statistics: Beachwood, OH, USA, 2008.
7. Freifeld, O. Statistics on Manifolds with Applications to Modeling Shape Deformations. Ph.D. Thesis, Division of Applied Mathematics, Brown University, Providence, RI, USA, 2014.
8. Celledoni, E.; Fiori, S. Descent methods for optimization on homogeneous manifolds. *Math. Comput. Simulat.* **2008**, *79*, 1298–1323. [[CrossRef](#)]
9. Fan, J.; Nie, P. Quadratic problems over the Stiefel manifold. *Oper. Res. Lett.* **2006**, *34*, 135–141. [[CrossRef](#)]
10. Gabay, D. Minimizing a differentiable function over a differentiable manifold. *J. Optim. Theory Appl.* **1982**, *37*, 177–219. [[CrossRef](#)]
11. Luenberger, D. The gradient projection method along geodesics. *Manag. Sci.* **1972**, *18*, 620–631. [[CrossRef](#)]
12. Wang, J.; Sun, H.; Li, D. A geodesic-based Riemannian gradient approach to averaging on the Lorentz group. *Entropy* **2017**, *19*, 698. [[CrossRef](#)]
13. Zhang, Z.Y.; Qiu, Y.Y.; Du, K.Q. Conditions for optimal solutions of unbalanced Procrustes problem on Stiefel manifold. *J. Comput. Math.* **2007**, *25*, 661–671.
14. Balande, U.; Shrimankar, D. SRIFA: Stochastic ranking with improved-firefly-algorithm for constrained optimization engineering design problems. *Mathematics* **2019**, *7*, 250. [[CrossRef](#)]
15. Hong, D.H.; Han, S. The general least square deviation OWA operator problem. *Mathematics* **2019**, *7*, 326. [[CrossRef](#)]
16. Mills, E.A.; Yu, B.; Zeng, K. Satisfying bank capital requirements: A robustness approach in a modified Roy safety-first framework. *Mathematics* **2019**, *7*, 593. [[CrossRef](#)]
17. Guo, D.; Li, B.; Zhao, W. Physical layer security and optimal multi-time-slot power allocation of SWIPT system powered by hybrid energy. *Information* **2017**, *8*, 100. [[CrossRef](#)]
18. Omara, I.; Zhang, H.; Wang, F.; Hagag, A.; Li, X.; Zuo, W. Metric learning with dynamically generated pairwise constraints for ear recognition. *Information* **2018**, *9*, 215. [[CrossRef](#)]
19. 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. [[CrossRef](#)] [[PubMed](#)]
20. Moakher, M. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. A* **2005**, *26*, 735–747. [[CrossRef](#)]

21. Kaneko, T.; Fiori, S.; Tanaka, T. Empirical arithmetic averaging over the compact Stiefel manifold. *IEEE Trans. Signal Proces.* **2013**, *61*, 883–894. [[CrossRef](#)]
22. Fiori, S.; Kaneko, T.; Tanaka, T. Tangent-bundle maps on the Grassmann manifold: Application to empirical arithmetic averaging. *IEEE Trans. Signal Proces.* **2015**, *63*, 155–168. [[CrossRef](#)]
23. Fiori, S. Solving minimal-distance problems over the manifold of real symplectic matrices. *SIAM J. Matrix Anal. A* **2011**, *32*, 938–968. [[CrossRef](#)]
24. Başkal, S.; Georgieva, E.; Kim, Y.S.; Noz, M.E. Lorentz group in classical ray optics. *J. Opt. B Quantum Semiclass. Opt.* **2004**, *6*, S455–S472.
25. Podleś, P.; Woronowicz, S.L. Quantum deformation of Lorentz group. *Commun. Math. Phys.* **1990**, *130*, 381–431. [[CrossRef](#)]
26. Kawaguchi, H. Evaluation of the Lorentz group Lie algebra map using the Baker-Cambell-Hausdorff formula. *IEEE Trans. Magn.* **1999**, *35*, 1490–1493. [[CrossRef](#)]
27. Jost, J. *Riemannian Geometry and Geometric Analysis*; Springer: Berlin/Heidelberg, Germany, 2011.
28. Armijo, L. Minimization of functions having Lipschitz continuous first partial derivatives. *Pac. J. Math.* **1966**, *16*, 1–3. [[CrossRef](#)]
29. Baker, A. *Matrix Groups: An Introduction to Lie Group Theory*; Springer: Berlin, Germany, 2002.
30. Khvedelidze, A.; Mladenov, D. Generalized Calogero-Moser-Sutherland models from geodesic motion on $GL^+(n, R)$ group manifold. *Phys. Lett. A* **2002**, *299*, 522–530. [[CrossRef](#)]
31. Andrica, D.; Rohan, R.-A. A new way to derive the Rodrigues formula for the Lorentz group. *Carpathian J. Math.* **2014**, *30*, 23–29.
32. Wu, R.; Chakrabarti, R.; Rabitz, H. Critical landscape topology for optimization on the symplectic group. *J. Optim. Theory Appl.* **2010**, *145*, 387–406. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).