Tangent-Bundle Maps on the Grassmann Manifold: Application to Empirical Arithmetic Averaging

Simone Fiori, Tetsuya Kaneko and Toshihisa Tanaka (*Senior Member*)

***Abstract*— The present paper elaborates on tangent-bundle maps on the Grassmann manifold, with application to subspace arithmetic averaging. In particular, the present contribution elaborates on the work about retraction/lifting maps devised for the Stiefel manifold in the recently published paper T. Kaneko, S. Fiori and T. Tanaka, "Empirical arithmetic averaging over the compact Stiefel manifold," *IEEE Trans. on Signal Processing*, Vol. 61, No. 4, pp. 883 – 894, February 2013 and discusses the extension of such maps to the Grassmann manifold. Tangent-bundle maps are devised on the basis of the thin QR matrix decomposition, the polar matrix decomposition and the exponential map. Also, tangent-bundle pseudo-maps based on the matrix Cayley transform are devised. Theoretical and numerical comparisons about the devised tangent-bundle maps are performed in order to get an insight into their relative merits and demerits, with special emphasis to their computational burden. The averaging algorithms based on the thin-QR decomposition maps stands out as it exhibits the best trade off between numerical precision and computational burden. Such algorithm is further compared with two Grassmann averaging algorithms drawn from the scientific literature on an handwritten digits recognition data set. The thin-QR tangent-bundle maps-based algorithm exhibits again numerical features that make it preferable over such algorithms.**

I. Introduction

GRASSMANN MANIFOLDS are ubiquitous in mathematical engineering [8] and, in particular, in mathematical Signal Processing. Some known applications are to pattern recognition, as affine shape analysis [38] and face recognition [22] in computer vision, to motion analysis in artificial vision [27], to handwritten character classification [5], to object categorization and multiple motion segmentation [4], to quantum computation [12], to matrix completion for missing data imputation [29], to compressed sensing [40], to MIMO transmission design in telecommunications [6], to sensor reading prediction by subspace tracking [1], to model reduction by Galerkin projection [31] and to analyzing brain states of cognitive function and brain disorders [9].

S. Fiori (**corresponding author**) is with Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, Via Brecce Bianche, I-60131 Ancona (Italy).

T. Kaneko was with the Laboratory of Signal and Image Processing, Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology (TUAT), 2-24-16, Nakacho, Koganei-shi, Tokyo 184-8588 (Japan).

T. Tanaka is with the Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology (TUAT), 2-24-16, Nakacho, Koganei-shi, Tokyo 184-8588 (Japan).

In Signal Processing, Grassmann manifolds found widespread applications. The contribution [28], showed that the problem of quantization in an Euclidean space with unitary constraints (which arise in areas such as wireless communication with multiple antennas at the transmitter and at the receiver) can be formulated as an unconstrained problem on a Grassmann manifold. The paper [20] discusses the problem of redundancy reduction of high-dimensional noisy signals that may contain anomaly (or *rare*) vectors, which need to be preserved. Several important design problems in signal processing for communications can be cast as optimization problems on the complex Grassmann manifold [13]. Relationships between entities in datasets are often of multiple nature, as geographical distances and social relationships. Such relationships may be modeled by a set of weighted and undirected graphs that form a global multilayer graph. The paper [7], addresses the problem of analyzing multilayer graphs and propose methods for clustering by using tools from subspace analysis on a Grassmann manifold.

A statistical characterization of a set of data is their empirical mean, which appears as an object carrying on the same structure of the data. Averaging over a data-set is a good method to smooth out data and to alleviate measurement errors and random fluctuations. Moreover, data-processing algorithms often make use of averages of subsets of data, such as the k-means classification algorithm. In order to compute statistics on differentiable manifolds, it is necessary to build up algorithms that take into account the geometric structure of the (generally curved) space that the data belong to. Well-known examples are the computation of the *mean shift* on Riemannian matrix manifolds [34], the ample corpus of results about the space of symmetric positive-definite matrices [21] and the analysis of the statistics of subspaces in a Grassmann-manifold setting described in [3], [27].

The aim of the present paper is to extend to the Grassmann manifold the empirical arithmetic averaging algorithms devised in previous contributions by the same authors published on these Transactions, namely the algorithms to compute averages over Lie groups [10] and the algorithms to compute averages over the compact Stiefel manifold [19]. The structure of the empirical averaging algorithms initially proposed in [10] is well consolidated and may be replicated or extended to different (even non-metric) spaces. The paper [19] explained very clearly that the main efforts in the application of the well-developed averaging algorithms is to devise appropriate tangent-bundle maps for the space under consideration, which is the main contribution of the present paper, with reference to the Grassmann manifold.

Over the recent years, several contributions have appeared in the scientific literature about tangent-bundle maps from the tangent bundle to the base-manifold (namely, maps: $\mathrm{T}M \to M$) for a number of smooth manifolds M, mainly devoted to devise step-forward algorithms for optimization. The problem of averaging on manifold, however, is peculiar because the devised algorithms (see Section V for details) not only need tangent-bundle maps $\mathrm{T}M \to M$ but they also make intensive use of 'inverse' tangent-bundle maps $M \to \mathrm{T}M$, whose calculation is the main source of research efforts (as well as of computation burden). As it will be clear from Section II, a weaker notion of inverse map only will be defined, namely, 'right inverse'. The present paper introduces a number of tangent-bundle maps $\mathrm{T}M \leftrightarrow M$, discusses their analytic features and compares their numerical performances.

ORGANIZATION OF THE PAPER. The present paper is organized as follows. Section II recalls fundamental notions about the Grassmann manifold, which are instrumental in the development of the subsequent theory, and classical matrix decomposition/factorization methods. Section II also sets out a fundamental results about geodesic lines on the Grassmann manifold and recalls a measure of discrepancy between subspaces. Section III presents the devised tangent-bundle maps, namely, maps based on the thin QR matrix decomposition, the polar matrix decomposition and geodesic lines. Section III also presents details of the analytic features of such devised maps and of their (exact or approximate) interrelationships. Section IV presents two pairs of pseudo-tangent-bundle maps, which exploit the quotient space structure of the Grassmann manifold. Section V summarizes all the tangent-bundle maps devised in the previous Sections and introduces two empirical arithmetic averaging algorithms based on a fixed-point iteration scheme and illustrates their numerical features through numerical experiments. Section V also illustrates the results of a numerical comparison of the devised averaging algorithm with two algorithms from the scientific literature. Section VI concludes the paper.

II. NOTATION AND FUNDAMENTAL PROPERTIES

Let $\mathbb{V}$ be a finite-dimensional vector space. The Grassmann manifold (or Grassmannian) $\mathrm{Gr}(\mathbb{V}, n)$ is the set of all n-dimensional linear subspaces of $\mathbb{V}$. If $\mathbb{V} = \mathbb{R}^p$, then the Grassmann manifold is also denoted as $\mathrm{Gr}(p, n)$, namely:

$$\mathrm{Gr}(p,n) \stackrel{\mathrm{def}}{=} \{\mathbb{S} \subset \mathbb{R}^p \mid \dim(\mathbb{S}) = n\}, \tag{1}$$

with $n < p$. The Grassmannians are compact, differentiable manifolds. Grassmann manifolds are special cases of more general geometrical objects termed *flag manifolds* [11].

In the present paper, the symbol $\mathrm{St}(p,n) \subset \mathbb{R}^{p\times n}$ denotes a compact Stiefel manifold [19]. Often, an element of a Grassmann manifold, represented by a matrix $X \in \mathrm{St}(p,n)$, is denoted by the symbol $[X] \in \mathrm{Gr}(p,n)$, which indicates the subspace of $\mathbb{R}^p$ spanned by the columns of the matrix X. In the present paper, with a slight abuse of notation, both a subspace $[X]$ and its Stiefel-matrix representative X are denoted by the same symbol X. By definition of Grassmann manifold and by the relation between a Grassmann element and one of its Stiefel-matrix representative, two distinct matrices $X \in \mathrm{St}(p,n)$ and $Y \in \mathrm{St}(p,n)$ represent the same point on the Grassmann manifold $\mathrm{Gr}(p,n)$ (i.e., the same subspace) if they are related by $Y = XO$, with $O \in \mathrm{St}(n,n)$ being an arbitrary orthogonal matrix (n-dimensional rotation or reflection). In this case, the matrices X and Y are said to be *equivalent* and their equivalency is denoted by $X \sim Y$.

The tangent space $\mathrm{T}_X\mathrm{Gr}(p,n)$ is decomposed in the direct sum of the *horizontal space* and of the *vertical space* at $X \in \mathrm{Gr}(p,n)$. Starting from a point $X \in \mathrm{Gr}(p,n)$, moving along a horizontal direction causes the change of subspace, while moving along a vertical direction does not change the subspace X. The horizontal space at $X \in \mathrm{Gr}(p,n)$ reads

$$\mathrm{H}_X\mathrm{Gr}(p,n) \stackrel{\mathrm{def}}{=} \{V \in \mathbb{R}^{p\times n} \mid V^T X = 0\}. \tag{2}$$

A *tangent-bundle map*: $\mathrm{H}_X\mathrm{Gr}(p,n) \to \mathrm{Gr}(p,n)$ is denoted by $P_X \colon V \mapsto Y$, while a map: $\mathrm{Gr}(p,n) \to \mathrm{H}_X\mathrm{Gr}(p,n)$ is denoted by $P_X^{-1} \colon Y \mapsto V$. A review of the notion, the formal definition and the properties of a class of tangent-bundle maps from a horizontal space to a quotient manifold termed *retractions* may be found, e.g., in the publicly available report [26]. In the present contribution, a different definition is adopted, which is tailored to the problem of averaging over Grassmannians. In the present context, in fact, a tangent-bundle map $P_X \colon \mathrm{H}_X\mathrm{Gr}(p,n) \to \mathrm{Gr}(p,n)$ is any map that admits, at least *locally*, a *right inverse* (or *section*) $P_X^{-1} \colon \mathrm{Gr}(p,n) \to \mathrm{H}_X\mathrm{Gr}(p,n)$. Namely, the two maps satisfy the fundamental composition law

$$P_X(P_X^{-1}(Y)) \sim Y. \tag{3}$$

(Since the map P_X^{-1} is a right inverse, the property (3) on the composition $P_X \circ P_X^{-1}$ does *not* imply, in general, any similar property of the reverse composition $P_X^{-1} \circ P_X$.)

The identity matrix is denoted by I and the zero matrix is denoted by 0. When it is necessary to specify the size, the zero matrix is denoted by $0_{p,n}$ and the identity matrix is denoted by $I_{p,n}$. If $p = n$, the identity matrix is denoted by I_n. If $p > n$ the symbol $I_{p,n}$ denotes a pseudo-identity matrix $\left[\begin{smallmatrix} I_n \\ 0_{p-n,n} \end{smallmatrix}\right]$.

Recall the following kinds of *matrix decomposition*:

- **Thin QR decomposition** [14]: The 'thin QR decomposition' of a rectangular matrix A is the product QR of a Stiefel matrix Q of the same size of the matrix A and of an upper-triangular matrix R (whose in-diagonal elements are assumed to be positive so that the decomposition is unique). The notation $Q = \mathrm{qf}(A)$ is made use of to denote the 'Q' factor.
- **Polar decomposition** [18]: The 'polar decomposition' of a rectangular matrix A is the product QS of a Stiefel matrix Q (termed *polar factor*) of the same size of the matrix A and of a symmetric matrix S. The notation $Q = \mathrm{pf}(A)$ is made use of to denote a polar factor. If the matrix A is $p \times n$ with $p \geq n$, then the polar factor has the closed-form expression $\mathrm{pf}(A) = A(A^TA)^{-\frac{1}{2}}$.
- **Thin singular value decomposition** [37]: The 'thin SVD' of a $p \times n$ matrix A is the product $U\Sigma W^T$, where $U \in \mathrm{St}(p,n)$, Σ is $n \times n$ diagonal and $W \in \mathrm{St}(n,n)$.

- **Spectral decomposition** [25]: The 'spectral decomposition' of a real symmetric matrix A is the product $Q\Sigma Q^T$, where Q is an orthogonal matrix of the same size of the matrix A and Σ is a diagonal matrix.
- **Generalized singular value decomposition** [39]: The 'GSVD' is a matrix decomposition more general than the SVD. Let A and B be real-valued rectangular matrices: Their generalized singular value decomposition factors $A = UCW^T$ and $B = VSW^T$, with U, V, W being orthogonal matrices and $C^TC + S^TS = I$.

A symmetric positive-definite matrix A may be spliced as $U\Sigma^2U^T$, where $U \in \mathrm{St}(n,n)$ and $\Sigma^2 > 0$, by means of the spectral factorization. Its (principal) square root is denoted by $A^{\frac{1}{2}}$ and is defined as $U\Sigma U^T$, where $\Sigma > 0$ (it is, in fact, readily verified that $A^{\frac{1}{2}}A^{\frac{1}{2}} = U\Sigma(U^TU)\Sigma U^T = U\Sigma^2U^T = A$). Then, it holds that the determinant $\det(A^{\frac{1}{2}}) = \det(U\Sigma U^T) = \det(\Sigma) = \sqrt{\det(\Sigma^2)} = \sqrt{\det(U\Sigma^2U^T)} = \sqrt{\det(A)}$.

Also, let us recall the *Sherman-Morrison-Woodbury identity* (or *matrix inversion lemma*, see, e.g., [17]). For the purposes of the present paper, it is recalled in the simplified form $(I + AA^T)^{-1} = I - A(I + A^TA)^{-1}A^T$. Such formula is useful when the matrix A is of size $p \times n$ with $n \ll p$, because the inversion of a $p \times p$ matrix is converted into the inversion of a $n \times n$ matrix.

A geodesic arc on the Grassmannian $\mathrm{Gr}(p,n)$ with respect to its canonical metric, emanating from the point X with initial tangent $V \in \mathrm{H}_X\mathrm{Gr}(p,n)$, is described by:

$$\Gamma_{X,V}(t) = XW\cos(t\Sigma)W^T + U\sin(t\Sigma)W^T, \qquad (4)$$

where $U\Sigma W^T$ is the thin singular value decomposition of the horizontal vector V, $t \in [0,\ 1]$. (The trigonometric functions of a diagonal matrix apply to the diagonal entries only.)

As a compact, smooth Riemannian manifold, the Grassmannian $\mathrm{Gr}(p,n)$ is *geodesically complete*, which implies that any two points on $\mathrm{Gr}(p,n)$ may be connected by a geodesic arc.

A possible tangent-bundle map $\mathrm{H}_X\mathrm{Gr}(p,n) \to \mathrm{Gr}(p,n)$ is given by the exponential map $\exp_X(V) \stackrel{\mathrm{def}}{=} \Gamma_{X,V}(1)$. The corresponding map $\mathrm{Gr}(p,n) \to \mathrm{H}_X\mathrm{Gr}(p,n)$ is given by the logarithmic map $\log_X(\cdot)$. The logarithmic map may be calculated by the help of generalized singular-value decomposition [27] but, in order to carry on its calculation, an explicit choice of the orthogonal complement $X_\perp$ is needed. As X is a $p \times n$ matrix with, generally, $p \gg n$, the matrix $X_\perp$ is of size $p \times (p-n)$ with $p - n$ large, which makes the usage of logarithmic map inconvenient. Such consideration justifies the present research, which aims at envisaging a number of alternative tangent-bundle maps.

A fundamental feature of the geodesic line (4) is described in the following result.

Theorem 1: *If two points $X \in \mathrm{Gr}(p,n)$ and $Y \in \mathrm{Gr}(p,n)$ are connected by a geodesic arc, namely $Y = \Gamma_{X,V}(t)$ for some $t \in [0,\ 1]$ and $V \in \mathrm{H}_X\mathrm{Gr}(p,n)$, then the matrix product X^TY is symmetric.*

Proof: By definition of geodesic line (4), upon calculating the thin singular value decomposition $U\Sigma W^T = V$, it holds that $Y = XW\cos(t\Sigma)W^T + U\sin(t\Sigma)W^T$. By premultiplying both sides by X^T, it is readily obtained that

$$X^TY = W\cos(t\Sigma)W^T + X^TU\sin(t\Sigma)W^T.$$

The term $W\cos(t\Sigma)W^T$ is symmetric, while the term $X^TU\sin(t\Sigma)W^T$ reads:

$$\begin{aligned}
&X^TU\left(t\Sigma - \frac{t^3\Sigma\Sigma\Sigma}{3!} + \frac{t^5\Sigma\Sigma\Sigma\Sigma\Sigma}{5!} + \cdots\right)W^T \\
&= X^T\left(tU\Sigma W^T - \frac{t^3}{3!}U\Sigma W^TW\Sigma W^TW\Sigma W^T\right. \\
&\left.+\frac{t^5}{5!}U\Sigma W^TW\Sigma W^TW\Sigma W^TW\Sigma W^TW\Sigma W^T + \cdots\right) \\
&= X^T\left(tV - \frac{t^3}{3!}V(W\Sigma W^T)^2 + \frac{t^5}{5!}V(W\Sigma W^T)^4 + \cdots\right) \\
&= X^TV\sin(tW\Sigma W^T)(W\Sigma W^T)^{-1},
\end{aligned}$$

which equals zero since $X^TV = 0$. (In the last line, it was assumed that $\det(\Sigma) \neq 0$, although such assumption is not necessary to complete the proof). This implies that the product X^TY is symmetric. □

The above property is peculiar of the representation chosen for a geodesic arc. Since the matrix W is orthogonal, it is possible to post-multiply the above expression of the geodesic arc by W such that the product X^TY is *not* symmetric, yet $XW(\cos\Sigma) + U(\sin\Sigma) \sim (XW(\cos\Sigma) + U(\sin\Sigma))W^T$ by the Grassmann-manifold equivalency.

In order to numerically measure the discrepancy between two subspaces $X, Y \in \mathrm{Gr}(p,n)$, the matrix XX^T may be used instead of the matrix X [12]. If X and Y are equivalent (namely, they represent the same element of the Grassmann manifold), then $Y = XO$, with O being orthogonal, hence $YY^T = (XO)(XO)^T = X(OO^T)X^T = XX^T$. The matrix $\Pi = XX^T$ is an *orthogonal projector*, namely, it enjoys the properties $\Pi^T = \Pi$ and $\Pi^2 = \Pi$. A measure of discrepancy between two projectors Π_1 and Π_2 is $\|\Pi_1 - \Pi_2\|_{\mathrm{F}}$, where the symbol $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. Therefore, it is possible to define a measure of discrepancy between two Stiefel representatives of $\mathrm{Gr}(p,n)$ elements as:

$$\varepsilon(X,Y) \stackrel{\mathrm{def}}{=} \|XX^T - YY^T\|_{\mathrm{F}} = \sqrt{2n - 2\mathrm{tr}(Y^TXX^TY)}. \qquad (5)$$

Such measure of discrepancy has the advantage that it does not measure the dissimilarity between two Stiefel matrices but between two subspaces. It coincides with the "projection Frobenius norm" [15], except for an inessential constant factor. A discussion on the merits of such discrepancy measure in relation to other subspace distances may be found in [35].

III. Tangent-bundle maps

A first attempt to relieve the computational burden of geodesic-based optimization in (complex-valued) Grassmannians was presented in the early paper [23]. Since such paper is concerned with a steepest gradient descent method and a modified Newton method (with application to quadratic or quasi-quadratic problems), it focuses on the tangent-bundle map $T_X\mathrm{Gr}(p,n) \to \mathrm{Gr}(p,n)$ and not on its inverse. The present section introduces the devised tangent-bundle maps

$\mathrm{H}_X\mathrm{Gr}(p,n) \to \mathrm{Gr}(p,n)$ and $\mathrm{Gr}(p,n) \to \mathrm{H}_X\mathrm{Gr}(p,n)$. The first two tangent-bundle maps, namely, the one based on the thin QR decomposition and the one based on the polar decomposition, derive directly from the maps $\mathrm{T}_X\mathrm{St}(p,n) \to \mathrm{St}(p,n)$ and $\mathrm{St}(p,n) \to \mathrm{T}_X\mathrm{St}(p,n)$ discussed in the contribution [19]. They enjoy different properties, in the present context, which are studied and discussed analytically.

A. Thin-QR-decomposition-based maps

The map $\mathrm{H}_X\mathrm{Gr}(p,n) \to \mathrm{Gr}(p,n)$ based on the thin-QR-decomposition is defined as $P_X(V) \stackrel{\mathrm{def}}{=} \mathrm{qf}(X+V)$. It enjoys the following properties.

Theorem 2: *Let $Y = \mathrm{qf}(X+V)$ with $X \in \mathrm{Gr}(p,n)$ and $V \in \mathrm{H}_X\mathrm{Gr}(p,n)$. Then, the product X^TY is nonsingular. In addition, it holds that the matrix X^TY is upper triangular.*

Proof: By definition of QR decomposition, there exists an upper-triangular matrix R such that $X+V = YR$. Pre-multiplying both sides by X^T gives

$$X^TX + X^TV = X^TYR.$$

Since $X^TX = I$ and $X^TV = 0$, the above equation recast as $(X^TY)R = I$. Since both R and I are upper-triangular, then X^TY must be upper-triangular. From the equality $X+V = YR$, it further follows that

$$(X+V)^T(X+V) = R^TY^TYR.$$

By the properties of X, Y and V, the above equality becomes $I+V^TV = R^TR$. Taking the determinant of both sides yields $\det^2(R) = \det(I+V^TV)$. The matrix $I+V^TV$ is positive definite, hence $\det(R) \neq 0$, which implies that R is invertible and so is the product X^TY. □

The corresponding map $\mathrm{Gr}(p,n) \to \mathrm{H}_X\mathrm{Gr}(p,n)$ is defined as $P_X^{-1}(Y) \stackrel{\mathrm{def}}{=} Y(X^TY)^{-1} - X$. It is compatible with the geometry of the Grassmannians, as shown in the following.

Theorem 3: *Given a matrix $X \in \mathrm{Gr}(p,n)$ and an arbitrary matrix $Y \in \mathbb{R}^{p\times n}$ such that the product X^TY is nonsingular, then $Y(X^TY)^{-1} - X \in \mathrm{H}_X\mathrm{Gr}(p,n)$.*

Proof: Set $V = Y(X^TY)^{-1} - X$. It holds that $X^TV = X^TY(X^TY)^{-1} - X^TX$. The matrix X^TY is invertible, hence the first term exists and equals I. The matrix X is Stiefel, hence the second term equals I. Consequently, $X^TV = 0$ and therefore $V \in \mathrm{H}_X\mathrm{Gr}(p,n)$. □

In view of arithmetic subspace averaging, it is necessary to analyze to what extent the two devised maps satisfy the fundamental composition law (3). The following result answers such fundamental question.

Theorem 4: *The composition $P_X(P_X^{-1}(Y))$ exists for every pair $X, Y \in \mathrm{Gr}(p,n)$ such that the product X^TY is invertible. If, in addition, the two points $X, Y \in \mathrm{Gr}(p,n)$ are connected by a geodesic arc (4), then it holds that $P_X(P_X^{-1}(Y)) \sim Y$.*

Proof: The matrix X^TY is non-singular, hence, the quantity $P_X^{-1}(Y) = Y(X^TY)^{-1} - X$ exists and $P_X(P_X^{-1}(Y)) = \mathrm{qf}(Y(X^TY)^{-1})$. (As the product X^TY is not upper-triangular, in general, the Q-factor of the QR decomposition of the matrix $Y(X^TY)^{-1}$ does not coincide with Y.)

In order to prove the second implication, set $Q = \mathrm{qf}(Y(X^TY)^{-1})$ where, by Theorem 1, the quantity $S \stackrel{\mathrm{def}}{=} (X^TY)^{-1}$ is symmetric. The QR decomposition implies that $YS = QR$, with Q orthogonal and R upper triangular. Such equality implies that $(YS)^T(YS) = (QR)^T(QR)$, namely, that $S^2 = R^TR$. As $YS = QR$, then $Q = YO$ with $O = SR^{-1}$. The matrix O is orthogonal, in fact $O^TO = (R^{-1})^TS^2R^{-1} = (R^T)^{-1}R^TRR^{-1} = I$. This proves that $Y \sim Q$. □

In the context of averaging, it is typically requested that the points to average be sufficiently close to each other [19]. Such condition serves to ensure that the product X^TY be non-singular, in fact, if X and Y are sufficiently close to each other, then the product X^TY is close to an orthogonal matrix (whose absolute determinant equals 1). If two matrices $X, Y \in \mathrm{St}(p,n)$ are *not* sufficiently close to each other, then the product X^TY is not necessarily invertible. Consider the following example:

$$X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \; Y = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{6}$$

Although $X^TX = I$ and $Y^TY = I$, it results that:

$$X^TY = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \tag{7}$$

The inverse matrix $(X^TY)^{-1}$ does not exist, in this example.

B. Polar-decomposition-based maps

The polar-decomposition map $\mathrm{H}_X\mathrm{Gr}(p,n) \to \mathrm{Gr}(p,n)$ is defined as $P_X(V) \stackrel{\mathrm{def}}{=} \mathrm{pf}(X+V) = (X+V)(I+V^TV)^{-\frac{1}{2}}$. It enjoys the following property.

Theorem 5: *Let $Y \stackrel{\mathrm{def}}{=} (X+V)(I+V^TV)^{-\frac{1}{2}}$ with $X \in \mathrm{Gr}(p,n)$ and $V \in \mathrm{H}_X\mathrm{Gr}(p,n)$. Then, the product X^TY is nonsingular for $\|V\|_{\mathrm{F}}$ finite.*

Proof: It holds that

$$X^TY = (X^TX + X^TV)(I+V^TV)^{-\frac{1}{2}} = (I+V^TV)^{-\frac{1}{2}}.$$

Taking the determinant of both sides yields $\det(X^TY)\sqrt{\det(I+V^TV)} = 1$. For $\|V\|_{\mathrm{F}}$ finite, the last equality implies that X^TY is invertible. □

The map $\mathrm{Gr}(p,n) \to \mathrm{H}_X\mathrm{Gr}(p,n)$ based on the polar decomposition is defined as $P_X^{-1}(Y) \stackrel{\mathrm{def}}{=} Y(X^TY)^{-1} - X$. The fact that it is a valid map follows again from Theorem 3 above.

In view of subspace averaging, it is important to note that a result similar to Theorem 4 holds even for the polar-decomposition-based maps pair, as explained below.

Theorem 6: *The composition $P_X(P_X^{-1}(Y))$ exists for every pair $X, Y \in \mathrm{Gr}(p,n)$ such that the product X^TY is invertible. The identity $P_X(P_X^{-1}(Y)) = Y$ holds if $Y = \Gamma_{X,V}(t)$ for $\|tV\|_{\mathrm{F}}$ sufficiently small, with Γ given by equation (4).*

Proof: By hypothesis, the quantity $P_X^{-1}(Y) = Y(X^TY)^{-1} - X$ exists. The quantity $P_X(P_X^{-1}(Y)) = \mathrm{pf}(Y(X^TY)^{-1})$. (As the product X^TY is not symmetric, in general, the polar-factor of the polar decomposition of the matrix $Y(X^TY)^{-1}$ does not coincide with Y.)

In order to prove the second implication, assume that two points are joined by a geodesic arc (4). If $\|tV\|_{\mathrm{F}}$ is sufficiently

small, the points X and Y are close to each other, hence the quantity $P_X^{-1}(Y)$ exists and the composition $P_X(P_X^{-1}(Y))$ is well defined. By the Theorem 1, the product X^TY yields a symmetric matrix. Then, $\mathrm{pf}(Y(X^TY)^{-1}) = Y$ because the factor $(X^TY)^{-1}$ coincides with the symmetric factor of the polar decomposition, hence the stated property holds true. □

C. Geodesic-based tangent-bundle maps

The geodesic-based tangent-bundle map $\mathrm{H}_X\mathrm{Gr}(p,n) \to \mathrm{Gr}(p,n)$ is given by the exponential map, namely, $P_X(V) \stackrel{\mathrm{def}}{=} \exp_X(V)$, while the associated tangent-bundle map $\mathrm{Gr}(p,n) \to \mathrm{H}_X\mathrm{Gr}(p,n)$ is given by the logarithmic map, namely, $P_X^{-1}(Y) \stackrel{\mathrm{def}}{=} \log_X(Y)$. The computation of the logarithmic map is numerically cumbersome, however, a noteworthy finding of the present contribution is that, whenever two points $X, Y \in \mathrm{Gr}(p,n)$ are such that $X^TY = Y^TX$, the computation of the logarithmic map $\log_X(Y)$ is facilitated, as shown in the following result.

Theorem 7: *If two points $X, Y \in \mathrm{Gr}(p,n)$ are such that the product X^TY is symmetric, then $\log_X(Y) = U\Sigma W^T$, where $W(\cos\Sigma)W^T$ denotes the spectral decomposition of the matrix X^TY and $U \stackrel{\mathrm{def}}{=} (I - XX^T)YW(\sin\Sigma)^{-1}$.*

Proof: A geodesic arc that connects the points X and Y is of the form (4), therefore it must hold that:

$$Y = XW(\cos\Sigma)W^T + U(\sin\Sigma)W^T, \tag{8}$$

and the initial tangent vector is $V = U\Sigma W^T$. Pre-multiplying both sides of the equation (8) by X^T yields $X^TY = WDW^T$, where $D \stackrel{\mathrm{def}}{=} \cos\Sigma$. The orthogonal matrix W and the diagonal matrix D are computed by the spectral decomposition of the symmetric matrix X^TY. In particular, $D = W^TX^TYW$. The equation (8) recasts as $U(\sin\Sigma)W^T = Y - XW(\cos\Sigma)W^T$, from which it follows that $U = (Y - XW(\cos\Sigma)W^T)W(\sin\Sigma)^{-1}$. Computations yield:

$$\begin{aligned} U &= (YW - XWD)(\sin\Sigma)^{-1} \\ &= (YW - XW(W^TX^TYW))(\sin\Sigma)^{-1}, \end{aligned}$$

which leads to the stated result. □

The only required cumbersome operation is the evaluation of the spectral decomposition of a square symmetric matrix. The Theorem 7 suggests the following pair of tangent-bundle maps: $P_X(V) \stackrel{\mathrm{def}}{=} XW(\cos\Sigma)W^T + U(\sin\Sigma)W^T$, where $U\Sigma W^T$ is the thin singular value decomposition of the matrix V, and $P_X^{-1}(Y) \stackrel{\mathrm{def}}{=} (I - XX^T)YW(\sin\Sigma)^{-1}\Sigma W^T$, where $W(\cos\Sigma)W^T$ is the spectral decomposition of the matrix X^TY.

The map P_X^{-1} is well-defined even when the product X^TY is not invertible as, for instance, if $X^TY = 0$ (as in the example shown at the end of Subsection III-A). In such exemplary case, in fact, the spectral decomposition of the matrix X^TY in the form $W(\cos\Sigma)W^T$ gives W arbitrary orthogonal and $\Sigma = \frac{\pi}{2}I$. Therefore, $P_X^{-1}(Y) = (YW - 0)\left(\sin\frac{\pi}{2}I\right)^{-1}\left(\frac{\pi}{2}I\right)W^T = \frac{\pi}{2}YWW^T$. Hence, direct substitution yields $\log_X(Y) = \frac{\pi}{2}Y$. The map P_X^{-1} is well-defined even when the diagonal matrix Σ is singular, provided that one interprets $\frac{0}{\sin 0} = 1$.

D. Interrelationships between the tangent-bundle maps

A close examination of the proofs of the Theorem 4 and of the Theorem 6 reveals that, given two points $X \in \mathrm{Gr}(p,n)$ and $Y \in \mathrm{Gr}(p,n)$, the fact that the fundamental composition law $P_X(P_X^{-1}(Y)) \sim Y$ is met is indeed guaranteed by the condition that the product X^TY results in a symmetric matrix. However, given two representative matrices $X, Y \in \mathrm{St}(p,n)$ of two Grassmann-manifold elements, such condition is not necessarily met. Such a numerical problem may be solved by addressing the following:

Question: *Given two matrices $X, Y \in \mathrm{St}(p,n)$, does there exist an orthogonal matrix $O \in \mathrm{St}(n,n)$ such that the product X^TY_1, with $Y_1 = YO$, is symmetric ?* □

If such an orthogonal matrix exists, then the matrices Y and Y_1 represent the same point on the Grassmann manifold. The answer to the above problem is as follows.

Theorem 8 (Symmetrization trick): *Given two matrices $X, Y \in \mathrm{St}(p,n)$, if the product X^TY is invertible, then the matrix $O = \mathrm{pf}(Y^TX)$ is the only orthogonal matrix that makes the product X^TYO symmetric.*

Proof: An orthogonal matrix O is sought such that the equality $X^TYO = S$ holds true, with S symmetric. Transposing both sides of the above equation gives $O^TY^TX = S$. Pre-multiplying both sides of the above equation by the matrix O gives $Y^TX = OS$, namely, O is the polar factor of the polar factorization of the matrix Y^TX. Since Y^TX is nonsingular, the polar factorization is unique. □

An important finding is that the maps P_X based on the thin QR decomposition and on the polar decomposition are equivalent to each other, as stated in the following result.

Theorem 9: *Let $X \in \mathrm{Gr}(p,n)$ and $V \in \mathrm{H}_X\mathrm{Gr}(p,n)$. For both the Thin-QR-decomposition-based map and the Polar-decomposition-based map P_X it holds that*:

$$P_X(V)P_X^T(V) = (X+V)(I+V^TV)^{-1}(X+V)^T. \tag{9}$$

Proof: Let $Y \stackrel{\mathrm{def}}{=} \mathrm{qf}(X+V)$. By definition of thin QR decomposition, it holds that $X + V = YR$, with R upper-triangular with positive entries on the main diagonal. Hence, $Y = (X+V)R^{-1}$ and

$$YY^T = (X+V)(R^TR)^{-1}(X+V)^T.$$

The equality $X + V = YR$ also implies that

$$(X+V)^T(X+V) = R^TY^TYR,$$

hence that $R^TR = I + V^TV$. Therefore,

$$YY^T = (X+V)(I+V^TV)^{-1}(X+V)^T.$$

Let $Z \stackrel{\mathrm{def}}{=} \mathrm{pf}(X+V)$. By definition of polar decomposition, $Z = (X+V)(I+V^TV)^{-\frac{1}{2}}$, therefore, it holds that

$$ZZ^T = (X+V)(I+V^TV)^{-1}(X+V)^T. \ \square$$

Using the operator $\mathrm{span} : \mathbb{R}^{p\times n} \to \mathrm{Gr}(p,n)$, the above property states that $\mathrm{span}(\mathrm{qf}(X+V)) = \mathrm{span}(\mathrm{pf}(X+V))$.

The geodesic-based map P_X does *not* enjoy the equivalency property (9). In order to compare the Thin-QR-decomposition-based maps and the Polar-decomposition-based maps with the Geodesic map, it is convenient to rewrite the property (9) in

a different (although equivalent) way. Let $U\Sigma W^T$ be the thin singular-value decomposition of V. Note that $(I+V^TV)^{-1} = [I + (W\Sigma U^T)(U\Sigma W^T)]^{-1} = (I + W\Sigma^2 W^T)^{-1}$. As $I = WW^T$, because the matrix W is orthogonal, it holds that $(I + V^TV)^{-1} = [W(I + \Sigma^2)W^T]^{-1} = W(I + \Sigma^2)^{-1}W^T$. Consequently, the equivalency property (9) recast as:

$$\begin{aligned} &P_X(V)P_X^T(V) = \\ &(X + U\Sigma W^T)W(I + \Sigma^2)^{-1}W^T(X^T + W\Sigma U^T) = \\ &(XW + U\Sigma)(I + \Sigma^2)^{-1}(W^TX^T + \Sigma U^T). \end{aligned} \tag{10}$$

The above expression is equivalent to: $P_X(V)P_X^T(V) =$

$$[XW \quad U] \begin{bmatrix} (I + \Sigma^2)^{-1} & (I + \Sigma^2)^{-1}\Sigma \\ \Sigma(I + \Sigma^2)^{-1} & \Sigma(I + \Sigma^2)^{-1}\Sigma \end{bmatrix} \begin{bmatrix} W^TX^T \\ U^T \end{bmatrix}. \tag{11}$$

For comparison purposes, it is convenient to recall that the Geodesic-based tangent-bundle map P_X is defined as:

$$P_X(V) = [XW \quad U] \begin{bmatrix} \cos\Sigma \\ \sin\Sigma \end{bmatrix} W^T, \text{ with } U\Sigma W^T = V. \tag{12}$$

Computing the product $P_X(V)P_X^T(V)$ for a Geodesic map yields: $P_X(V)P_X^T(V) =$

$$[XW \quad U] \begin{bmatrix} \cos^2\Sigma & \frac{1}{2}\sin(2\Sigma) \\ \frac{1}{2}\sin(2\Sigma) & \sin^2\Sigma \end{bmatrix} \begin{bmatrix} W^TX^T \\ U^T \end{bmatrix}. \tag{13}$$

A direct comparison of the expressions (11) and (13) reveals that *the Geodesic-based maps are not equivalent to the Thin-QR-decomposition-based and the Polar-decomposition-based maps*. In other words, the Theorem 9 *cannot* be extended to include the Geodesic map. Numerically, the Geodesic map and the Thin-QR/Polar-decomposition-based maps behave similarly only if, for every diagonal entry of the matrix Σ (i.e., for every singular value σ_{ii}), it holds that:

$$\cos\sigma_{ii} \approx \frac{1}{\sqrt{1+\sigma_{ii}^2}} \text{ and } |\sin\sigma_{ii}| \approx \frac{|\sigma_{ii}|}{\sqrt{1+\sigma_{ii}^2}}. \tag{14}$$

Such conditions will hold only if the norm $\|\Sigma\|_{\mathrm{F}}$ (that coincides with the norm $\|V\|_{\mathrm{F}}$) is sufficiently small. In order to substantiate the above quantitative comparison, the discrepancy between the Thin-QR-decomposition-based/Polar-decomposition-based maps and the Geodesic-based maps is evaluated analytically as follows.

Property: Let $X \in \mathrm{Gr}(p,n)$ and $V \in \mathrm{H}_X\mathrm{Gr}(p,n)$. Define $Y \stackrel{\mathrm{def}}{=} \mathrm{qf}(X+V)$ (for the Thin-QR-decomposition-based map) or $Y \stackrel{\mathrm{def}}{=} \mathrm{pf}(X+V)$ (for the Polar-decomposition-based map) and $Z \stackrel{\mathrm{def}}{=} \exp_X(V)$ (for the Geodesic-based map). The squared discrepancy between subspaces Y and Z reads:

$$\varepsilon^2(Y,Z) = 2n - \sum_{i=1}^{n} \varphi(\sigma_{ii}), \tag{15}$$

where the σ_{ii}'s denote the singular values of the horizontal vector V (namely, the diagonal entries of the matrix Σ in the thin singular value decomposition $U\Sigma W^T$ of the matrix V) and the auxiliary function $\varphi : \mathbb{R} \to \mathbb{R}$, defined by $\varphi(x) \stackrel{\mathrm{def}}{=} \frac{\cos x + x\sin x}{1+x^2}$, was made use of.

Proof: Since the matrix $[XW\ U]$ is orthogonal, the sought squared discrepancy reads as in Equation (16) on the next page. The evaluation of the Frobenius norm (lengthy but elementary calculations are omitted for brevity) gives the stated result. □

Since the auxiliary function $\varphi(x)$ assumes its maximum value in $x = 0$, the discrepancy $\varepsilon(Y,Z)$ gets smaller as the singular values σ_{ii} get smaller, coherently with the previous quantitative comparison.

IV. Cayley-transformation-based pseudo-maps

The *special orthogonal group* is defined as $\mathrm{SO}(p) \stackrel{\mathrm{def}}{=} \{O \in \mathrm{St}(p,p) \mid \det(O) = 1\}$. The special orthogonal group is a *Lie group* with associated *Lie algebra* $\mathfrak{so}(p) \stackrel{\mathrm{def}}{=} \{\Omega \in \mathbb{R}^{p\times p} \mid \Omega^T + \Omega = 0\}$. By using the relationships between the three spaces $\mathrm{Gr}(p,n)$, $\mathrm{SO}(p)$ and $\mathrm{St}(p,n)$, it is possible to extend the full-Cayley-based maps devised in [19]. The obtained maps are not properly tangent-bundle maps and they are here referred to as *pseudo-tangent-bundle maps*. In the following, two versions of such maps are presented: The first version is a brute-force application of the maps devised in [19], while the second version is based on an economical representation that takes into account the quotient-space structure of the Grassmann manifold. Both versions are based on the matrix Cayley transformation $\mathrm{cay} : \mathfrak{so}(p) \to \mathrm{SO}(p)$ defined as $\mathrm{cay}(\Omega) \stackrel{\mathrm{def}}{=} (I_p + \Omega)(I_p - \Omega)^{-1} = (I_p - \Omega)^{-1}(I_p + \Omega)$.

A. Brute-force Cayley-based maps

The following pseudo-tangent-bundle maps is defined: $\hat{P}_X : \mathfrak{so}(p) \to \mathrm{Gr}(p,n)$ as $\hat{P}_X(\Omega) \stackrel{\mathrm{def}}{=} \mathrm{cay}(\Omega)X$ and $\hat{P}_X^{-1} : \mathrm{Gr}(p,n) \to \mathfrak{so}(p)$ as $\hat{P}_X^{-1}(Y) \stackrel{\mathrm{def}}{=} (Y - X)(Y^TX - X^TY)^{-1}(Y - X)^T$, which is clearly well-defined when the matrix $Y^TX - X^TY$ is nonsingular. Recall that, as the matrix $Y^TX - X^TY$ is $n \times n$ skew-symmetric, it might be invertible only for n even. To what concerns averaging over the Grassmann manifold, the following statement is of use.

Theorem 10: *Given two elements $X, Y \in \mathrm{Gr}(p,n)$ such that the quantity $Y^TX - X^TY$ is nonsingular, the composition $\hat{P}_X(\hat{P}_X^{-1}(Y))$ is well-defined and it holds that $\hat{P}_X(\hat{P}_X^{-1}(Y)) = Y$.*

Proof. The above result is proven by direct substitution. □

The quantity $\hat{P}_X^{-1}(Y)$ is not well-defined if the matrices X and Y represent two points on a geodesic arc (4). In fact, Theorem 1 implies that X^TY is symmetric and hence that the matrix $Y^TX - X^TY$ is not invertible. In order to get rid of such problem, recall that if $O \in \mathrm{St}(n,n)$ is any arbitrary matrix, then $YO \sim Y$, hence, in the expression of $\hat{P}_X^{-1}(Y)$ one may replace, e.g., the matrix Y with a matrix YO with O arbitrary in $\mathrm{St}(n,n)$. In this case, then, it will hold that $\hat{P}_X(\hat{P}_X^{-1}(Y)) \sim Y$.

Two comments about the brute-force Cayley-based pseudo-tangent-bundle maps are in order:

- From a computational viewpoint, the parameterization in terms of the matrices Ω is based on large matrices (of size $p \times p$) compared to the parameterization in terms of the matrices V, which is based on matrices of size $p \times n$, hence the former does not look convenient.

$$\varepsilon^2(Y,Z)=\left\|\left[\begin{array}{cc}\cos^2\Sigma-(I+\Sigma^2)^{-1} & \frac{1}{2}\sin(2\Sigma)-(I+\Sigma^2)^{-1}\Sigma\\ \frac{1}{2}\sin(2\Sigma)-\Sigma(I+\Sigma^2)^{-1} & \sin^2\Sigma-\Sigma(I+\Sigma^2)^{-1}\Sigma\end{array}\right]\right\|_{\mathrm{F}}^2. \tag{16}$$

- From a theoretical point of view, the $\mathfrak{so}(p)$-parameterization does not exploit any knowledge about the tangent-bundle structure of the Grassmannians (it is the same parameterization that was used for the Stiefel manifold, in fact).

The next section presents a second version of Cayley-transform-based pseudo-tangent-bundle maps that overcomes such problems.

B. Economical Cayley-based maps

Due to the structure of the Grassmann manifold regarded as a quotient space and its relationship with the special orthogonal group of matrices, given a point $X\in\mathrm{Gr}(p,n)$ and a horizontal vector $V\in H_X\mathrm{Gr}(p,n)$, the horizontal vector may be decomposed as $V=X_\perp B$, where $X_\perp$ is a $p\times(p-n)$ orthogonal-complement matrix of X, namely, $[XX_\perp]\in\mathrm{St}(p,p)$, and $B\in\mathbb{R}^{(p-n)\times n}$ is arbitrary [27]. Fixing the point $X\in\mathrm{Gr}(p,n)$ that the horizontal space $\mathrm{H}_X\mathrm{Gr}(p,n)$ is referred to, and its orthogonal complement $X_\perp$, each horizontal vector is parameterized by a matrix B and the parameter space is $\mathbb{R}^{(p-n)\times n}$. Therefore, a pseudo-tangent-bundle map $\hat{P}_X:\mathbb{R}^{(p-n)\times n}\to\mathrm{Gr}(p,n)$ may be written as:

$$\hat{P}_X(B)\stackrel{\mathrm{def}}{=}[X\ X_\perp]\psi\left(\left[\begin{array}{cc}0 & -B^T\\ B & 0\end{array}\right]\right)I_{p,n}, \tag{17}$$

where $\psi:\mathfrak{so}(p)\to\mathrm{SO}(p)$. In [27], the coordinate map ψ is chosen as $\psi=\exp$. In the present paper, it is chosen as $\psi=\mathrm{cay}$, namely:

$$\hat{P}_X(B)=[X\ X_\perp]\mathrm{cay}(\Omega)I_{p,n},\ \Omega=\left[\begin{array}{cc}0 & -B^T\\ B & 0\end{array}\right], \tag{18}$$

in view of making the computation of the associated pseudo-tangent-bundle map $\hat{P}_X^{-1}:\mathrm{Gr}(p,n)\to\mathbb{R}^{(p-n)\times n}$ tractable.

In order to study an important property of the pseudo-tangent-bundle map (18), the following lemma is necessary.
Lemma: *Let $B\in\mathbb{R}^{(p-n)\times n}$ denote any arbitrary matrix such that $\det(I_n+B^TB)\neq 0$. The inverse of the matrix $\left[\begin{array}{cc}I_n & B^T\\ -B & I_{p-n}\end{array}\right]$ reads:*

$$\left[\begin{array}{cc}(I_n+B^TB)^{-1} & -B^T(I_{p-n}+BB^T)^{-1}\\ B(I_n+B^TB)^{-1} & (I_{p-n}+BB^T)^{-1}\end{array}\right]. \tag{19}$$

Proof: The lemma is proven by direct substitution. The existence of the inverse is guaranteed by the existence of the inverse matrices $(I_n+B^TB)^{-1}$ and $(I_{p-n}+BB^T)^{-1}$. By the *Sylvester's Determinant Theorem*, $\det(I_{p-n}+BB^T)=\det(I_n+B^TB)$, hence, the condition $\det(I_n+B^TB)\neq 0$ suffices. □

An important property of the pseudo-tangent-bundle map (18) is illustrated by the following result.

Theorem 11: *Let $Y\stackrel{\mathrm{def}}{=}[X\ X_\perp]\mathrm{cay}\left(\left[\begin{array}{cc}0 & -B^T\\ B & 0\end{array}\right]\right)I_{p,n}$. Then, it holds that the product X^TY is symmetric.*

Proof: By definition of orthogonal complement $X_\perp$ and of Cayley map, it holds that $X^TY=I_{p,n}^T(I_p-\Omega)^{-1}(I_p+\Omega)I_{p,n}$, with $\Omega=\left[\begin{array}{cc}0 & -B^T\\ B & 0\end{array}\right]$. By the above lemma, the product $I_{p,n}^T(I_p-\Omega)^{-1}$ simplifies as:

$$[I_n\ 0_{n\times(p-n)}]\times\left[\begin{array}{cc}(I_n+B^TB)^{-1} & -B^T(I_{p-n}+BB^T)^{-1}\\ B(I_n+B^TB)^{-1} & (I_{p-n}+BB^T)^{-1}\end{array}\right]=[(I_n+B^TB)^{-1} \quad -B^T(I_{p-n}+BB^T)^{-1}].$$

The product $(I_p+\Omega)I_{p,n}$ simplifies as:

$$\left[\begin{array}{cc}I_n & -B^T\\ B & I_{p-n}\end{array}\right]\left[\begin{array}{c}I_n\\ 0_{(p-n)\times n}\end{array}\right]=\left[\begin{array}{c}I_n\\ B\end{array}\right].$$

Therefore: $X^TY=$

$$I_{p,n}^T(I_p-\Omega)^{-1}(I_p+\Omega)I_{p,n}=\left[(I_n+B^TB)^{-1} \quad -B^T(I_{p-n}+BB^T)^{-1}\right]\left[\begin{array}{c}I_n\\ B\end{array}\right]=(I_n+B^TB)^{-1}-B^T(I_{p-n}+BB^T)^{-1}B,$$

which is symmetric. □

As a corollary of the Theorem 11, an explicit expression of the Cayley-transform-based pseudo-tangent-bundle map $\hat{P}_X$, useful in the numerical implementations, is found as follows.

Corollary: *The pseudo-tangent-bundle map (18) may be written explicitly as:*

$$\hat{P}_X(B)=X_\perp Z+XS, \tag{20}$$

with $S=(I_n+B^TB)^{-1}[I_n+(B^TB)^2]-B^TB$ and $Z=B[(I_n+B^TB)^{-1}(I_n-B^TB)+I_n]$.

Proof: By the above lemma, it holds that $\hat{P}_X(B)=$

$$[X\ X_\perp]\times\left[\begin{array}{cc}(I_n+B^TB)^{-1} & -B^T(I_{p-n}+BB^T)^{-1}\\ B(I_n+B^TB)^{-1} & (I_{p-n}+BB^T)^{-1}\end{array}\right]\left[\begin{array}{c}I_n\\ B\end{array}\right].$$

Computing the matrix-products gives the expression (20) with $S=(I_n+B^TB)^{-1}-B^T(I_{p-n}+BB^T)^{-1}B$ and $Z=B(I_n+B^TB)^{-1}+(I_{p-n}+BB^T)^{-1}B$. By virtue of the Sherman-Morrison-Woodbury matrix inversion formula, the inverse $(I_{p-n}+BB^T)^{-1}$ may be written equivalently as $I_{p-n}-B(I_n+B^TB)^{-1}B^T$. By substitution into the expressions of the matrix-coefficients S and Z, the stated result is readily obtained. □

In the applications, $n\ll p$, hence the computation burden of the expression (20) is much reduced with respect to the 'raw' expression (18). The component $X_\perp Z$ of the map $\hat{P}_X(B)$ represents a horizontal vector. The expression (20) has some resemblance with the *orthographic retraction* studied in the contribution [19]. The above result also allows verifying that the expression (20) is a pseudo-tangent-bundle map. In fact, setting $Y=X_\perp Z+XS$, it is readily verified that $Y^TY=Z^TZ+S^2$. By direct substitution of the expressions for S and Z given in the Corollary, it is verified that $Z^TZ+S^2=I$.

A pseudo-tangent-bundle map $\hat{P}_X^{-1}:\mathrm{Gr}(p,n)\to\mathbb{R}^{(p-n)\times n}$ associated with the map (18) is devised as follows.

Theorem 12: *Given two points $X, Y \in \mathrm{Gr}(p,n)$ such that $X^TY + I_n$ is symmetric nonsingular, a pseudo-tangent-bundle map $\hat{P}_X^{-1}(Y)$ associated with the pseudo-tangent-bundle map (18) is:*

$$\hat{P}_X^{-1}(Y) = X_\perp^T Y (X^TY + I_n)^{-1}. \quad (21)$$

Proof: A pseudo-tangent-bundle map $\hat{P}_X^{-1}(Y)$ should return a matrix B such that $\hat{P}_X(B) = Y$. Namely, the following equation, in the unknown B, needs to be set up:

$$Y = [X\ X_\perp](I_p - \Omega)^{-1}(I_p + \Omega)I_{p,n},\ \Omega = \begin{bmatrix} 0 & -B^T \\ B & 0 \end{bmatrix}. \quad (22)$$

Since $[X\ X_\perp]^{-1} = [X\ X_\perp]^T$, the above equation may be rewritten equivalently as:

$$(I_p - \Omega)\begin{bmatrix} X^TY \\ X_\perp^TY \end{bmatrix} = (I_p + \Omega)\begin{bmatrix} I_n \\ 0_{p-n,n} \end{bmatrix},\ \Omega = \begin{bmatrix} 0 & -B^T \\ B & 0 \end{bmatrix},$$

or, rearranging terms, as:

$$\begin{bmatrix} X^TY - I_n \\ X_\perp^TY \end{bmatrix} = \begin{bmatrix} 0 & -B^T \\ B & 0 \end{bmatrix}\begin{bmatrix} X^TY + I_n \\ X_\perp^TY \end{bmatrix}. \quad (23)$$

The above relation may be broken into two equations, namely:

$$X^TY - I_n = -B^T(X_\perp^TY), \quad (24)$$

$$X_\perp^TY = B(X^TY + I_n). \quad (25)$$

Provided that the matrix $X^TY + I_n$ is non-singular, a solution of the equation (25) is as in (21). Such solution must satisfy equation (24) too. If the product X^TY is symmetric, substituting $B = X_\perp^TY(X^TY + I_n)^{-1}$ into equation (24) gives:

$$X^TY - I_n = -(X^TY + I_n)^{-1}(X_\perp^TY)^T(X_\perp^TY).$$

Pre-multiply both sides by $X^TY + I_n$: The left-hand side will read $(X^TY + I_n)(X^TY - I_n) = (X^TY)^2 - I_n$. Since $X^TY = Y^TX$, the term $(X^TY)^2$ may be written as Y^TXX^TY. The above equality is hence equivalent to $Y^T(XX^T + X_\perp X_\perp^T)Y = I_n$. Since $X_\perp$ is an orthogonal complement of X, it holds that $XX^T + X_\perp X_\perp^T = I_p$. Since $Y^TY = I_n$, the above relation is an identity, hence the solution B satisfies both equations (24) and (25). □

By definition of the map $\hat{P}_X^{-1}$, it holds that $\hat{P}_X(\hat{P}_X^{-1}(Y)) = Y$. If, due to the ambiguity in the representation of two Grassmann elements, the product X^TY is not symmetric, it may be symmetrized by replacing the representative matrix Y with the equivalent representative matrix $Y\mathrm{pf}(Y^TX)$, as suggested in the Theorem 8.

V. Averaging over the Grassmann manifold

The problems of averaging over the Grassmann manifold was tackled, e.g., in [30] by a classical Karcher-mean approach. On the basis of the analyzed tangent-bundle maps, it is possible to devise an averaging method that retraces the technique proposed in the paper [19], namely, that extends Phytagoras' averaging method on the Grassmann manifold. Averaging over Grassmann manifold has known applications such as to signal subspace estimation [32], clustering by k-means for image classification under affine invariance [2][1], particle filtering technique on direction of arrival (DOA) tracking [33], and classification of 3D protein structures [36] (for a tutorial on clustering on the Grassmann manifold, see [15]). Likewise in the paper [19], in the present note a distinction is made between *fixed-point arithmetic averaging on* $\mathrm{Gr}(p,n)$ *based on tangent-bundle maps* and *fixed-point arithmetic averaging algorithms on* $\mathrm{Gr}(p,n)$ *based on pseudo-tangent-bundle maps*[2]. The maps devised in the Section III are summarized in the Table I. For comparison purpose, the logarithmic map associated with the exponential map has been recalled too from [27][3]. It is immediate to recognize that the logarithmic map P_X^{-1} is computationally cumbersome because it requires an explicit computation of an orthogonal-complement matrix $X_\perp$ and because it needs a generalized singular value decomposition. As a rule-of-thumb, in order to get an orthogonal-complement matrix $X_\perp$ from the matrix X, the following formula is used:

$$X_\perp = \mathrm{qf}([X\ I_{p,p-n}])\begin{bmatrix} 0_{n,p-n} \\ I_{p-n} \end{bmatrix}. \quad (26)$$

A. Fixed-point arithmetic averaging algorithms on $\mathrm{Gr}(p,n)$ by tangent-bundle maps

The following steps lead to an equivalence relation characterizing the empirical mean matrix $X \in \mathrm{Gr}(p,n)$ of a set of N available samples $X_k \in \mathrm{Gr}(p,n)$. The samples X_k are supposed to distribute around a center of mass $C \in \mathrm{Gr}(p,n)$.

1) Map the points $X_k \in \mathrm{Gr}(p,n)$ belonging to a neighborhood of the sought mean-point $X \in \mathrm{Gr}(p,n)$ onto $\mathrm{H}_X\mathrm{Gr}(p,n)$ by applying a tangent-bundle map P_X^{-1}. Denote such points as $V_k \stackrel{\mathrm{def}}{=} P_X^{-1}(X_k)$.
2) As $\mathrm{H}_X\mathrm{Gr}(p,n)$ is a vector space, it is possible to compute a linear combination of the elements $V_1, V_2, \ldots, V_N$. (This step represents indeed arithmetic averaging over a linear space.) Denote such a linear combination by $\overline{V} \stackrel{\mathrm{def}}{=} \mathcal{L}(V_1, V_2, \ldots, V_N)$ obtained by means of a linear operator $\mathcal{L} : (\mathrm{H}_X\mathrm{Gr}(p,n))^N \to \mathrm{H}_X\mathrm{Gr}(p,n)$.
3) Bring back the mean vector $\overline{V}$ to $\mathrm{Gr}(p,n)$ by the tangent-bundle map P_X and get an empirical mean-point $X \sim P_X(\overline{V})$.

(Perfect estimation of the empirical mean matrix would imply that $X \sim C$.) Summarizing the above procedure, a mean-point $X \in \mathrm{Gr}(p,n)$ is a solution of the relation

$$X \sim P_X\left(\mathcal{L}\left(P_X^{-1}(X_1), P_X^{-1}(X_2), \ldots, P_X^{-1}(X_N)\right)\right) \quad (27)$$

in the unknown X. In general, however, the relation (27) cannot be resolved in closed form. It may be solved by means of a fixed-point iteration algorithm that generates a sequence $X^{(i)} \in \mathrm{Gr}(p,n)$ of estimates of the sought empirical mean element X. When all the quantities are represented by matrices in $\mathrm{St}(p,n)$, the fixed-point scheme may be implemented as:

$$X^{(i+1)} = P_{X^{(i)}}\left(\alpha \sum_{k=1}^{N} P_{X^{(i)}}^{-1}(X_k)\right),\ i \geq 0, \quad (28)$$

[1]The contribution [2] defines the notion of covariance of a set of tangent vectors, which is made use of in the implementation of a clustering algorithm.

[2]It seems appropriate to underline that the wording "fixed-point arithmetic averaging" does not relate in any way whatsoever with the notion of "fixed-point arithmetic".

[3]Equation (23) in [27] contains a mistake: It should read $X_\perp X_\perp^T Y = X_\perp U_1 S W^T = X_\perp U_1 \left[\begin{smallmatrix} S_1 \\ 0 \end{smallmatrix}\right] W^T = U S_1 W^T$.

TABLE I

SUMMARY OF TANGENT-BUNDLE MAPS AND PSEUDO-TANGENT-BUNDLE MAPS FOR A GRASSMANN MANIFOLD $\mathrm{Gr}(p,n)$. THE COLUMN 'SPACE' RECALLS THE LINEAR SPACE WHERE ARITHMETIC AVERAGE IS ACTUALLY TAKEN ON.

Maps type	Maps	Notes	Space
Thin-QR-decomposition	$P_X(V) = \mathrm{qf}(X+V)$ $P_X^{-1}(Y) = Y(X^TY)^{-1} - X$	If X^TY is not symmetric, replace Y with $Y\mathrm{pf}(Y^TX)$ in $P_X^{-1}(Y)$	$\mathrm{H}_X\mathrm{Gr}(p,n)$
Polar-decomposition	$P_X(V) = \mathrm{pf}(X+V)$ $P_X^{-1}(Y) = Y(X^TY)^{-1} - X$	If X^TY is not symmetric, replace Y with $Y\mathrm{pf}(Y^TX)$ in $P_X^{-1}(Y)$	
Geodesic	$P_X(V) = XW(\cos\Sigma)W^T + U(\sin\Sigma)W^T$ $U\Sigma W^T$ is the thin singular value decomposition of V $P_X^{-1}(Y) = (I - XX^T)YW(\sin\Sigma)^{-1}\Sigma W^T$ $W(\cos\Sigma)W^T$ is the spectral decomposition of X^TY	If X^TY is not symmetric, replace Y with $Y\mathrm{pf}(Y^TX)$ in $P_X^{-1}(Y)$	
Exp/Log	$P_X(V) = XW(\cos\Sigma)W^T + U(\sin\Sigma)W^T$ $U\Sigma W^T$ is the thin singular value decomposition of V $P_X^{-1}(Y) = U\sin^{-1}(S_1)W^T$ $X^TY = VCW^T$, $X_\perp^TY = U_1SW^T$ denotes GSVD $X_\perp$ is an orthogonal complement of X U is formed by the first n columns of $X_\perp U_1$ S_1 is formed by the first n rows of S	—	
Brute-force Cayley	$\hat{P}_X(\Omega) = (I+\Omega)(I-\Omega)^{-1}X$ $\hat{P}_X^{-1}(Y) = (Y-X)(Y^TX - X^TY)^{-1}(Y-X)^T$	If X^TY is symmetric, replace Y with YO in $P_X^{-1}(Y)$	$\mathfrak{so}(p)$
Economical Cayley	$\hat{P}_X(B) = X_\perp Z + XS$ $S = (I_n + B^TB)^{-1}[I_n + (B^TB)^2] - B^TB$ $Z = B[(I_n + B^TB)^{-1}(I_n - B^TB) + I_n]$ $\hat{P}_X^{-1}(Y) = X_\perp^TY(X^TY + I_n)^{-1}$ $X_\perp$ is an orthogonal complement of X	If X^TY is not symmetric, replace Y with $Y\mathrm{pf}(Y^TX)$ in $P_X^{-1}(Y)$	$\mathbb{R}^{(p-n)\times n}$

where the matrix $X^{(0)} \in \mathrm{St}(p,n)$ denotes an initial guess and the parameter α needs to be nonzero.

B. Fixed-point arithmetic averaging algorithms on $\mathrm{Gr}(p,n)$ *by pseudo-tangent-bundle maps*

This section presents an algorithm for calculating arithmetic subspace averages on the basis of the relationship between the three spaces $\mathrm{Gr}(p,n)$, $\mathrm{SO}(p)$ and $\mathrm{St}(p,n)$. The present Subsection treats directly the involved quantities as matrices.

1) Map the points $X_k \in \mathrm{Gr}(p,n)$ belonging to a neighborhood of the sought mean-point $X \in \mathrm{Gr}(p,n)$ onto $\mathfrak{so}(p)$ or $\mathbb{R}^{(p-n)\times n}$ by applying a pseudo-tangent-bundle map $\hat{P}_X^{-1}$. Denote such point-matrices as $\Omega_k \stackrel{\mathrm{def}}{=} \hat{P}_X^{-1}(X_k)$ or $B_k \stackrel{\mathrm{def}}{=} \hat{P}_X^{-1}(X_k)$.
2) As $\mathfrak{so}(p)$ and $\mathbb{R}^{(p-n)\times n}$ are vector spaces, it is possible to compute a linear combination $\overline{\Omega} \stackrel{\mathrm{def}}{=} \alpha \sum_k \Omega_k$ or $\overline{B} \stackrel{\mathrm{def}}{=} \alpha \sum_k B_k$, with $\alpha \neq 0$. (This step represents indeed the arithmetic averaging over a linear space.)
3) Bring back the mean vector $\overline{\Omega}$ or $\overline{B}$ to $\mathrm{Gr}(p,n)$ by the pseudo-tangent-bundle map $\hat{P}_X$ and get an empirical mean-point $X = \hat{P}_X(\overline{\Omega})$ or $X = \hat{P}_X(\overline{B})$.

Summarizing the above procedure, a mean-point $X \in \mathrm{Gr}(p,n)$ is a solution of the equation:

$$X = \hat{P}_X\left(\alpha \sum_k \hat{P}_X^{-1}(X_k)\right) \tag{29}$$

in the unknown X. In general, however, the relation (29) cannot be resolved in closed form. It may be solved by means of a fixed-point iteration algorithm that generates a sequence $X^{(i)} \in \mathrm{Gr}(p,n)$ of estimates of the sought empirical average element X. The fixed-point scheme may be implemented as:

$$X^{(i+1)} = \hat{P}_{X^{(i)}}\left(\alpha \sum_{k=1}^{N} \hat{P}_{X^{(i)}}^{-1}(X_k)\right),\ i \geq 0, \tag{30}$$

where, again, the matrix $X^{(0)} \in \mathrm{St}(p,n)$ denotes an initial guess and $\alpha \neq 0$.

C. Numerical test of the averaging algorithms

In order to test the arithmetic subspace averaging algorithms developed in the present Section, the following test design was used for the first three experiments:

- Generate a random point $C \in \mathrm{Gr}(p,n)$.
- Generate a set of N random points X_k by using a geodesic arc emanating from C and with random horizontal tangent $V_k \in \mathrm{H}_X\mathrm{Gr}(p,n)$ each, namely, $X_k = \Gamma_{C,V_k}(1)$.
- Run a subspace averaging algorithm that generates a sequence of ever-refined samples $X^{(i)}$, namely, (28) or (30), and measure the *discrepancy to center* $\varepsilon(X^{(i)}, C)$ as well as the step-wise *differential discrepancy* $\varepsilon(X^{(i)}, X^{(i-1)})$.

In all the performed simulations, the parameter α was set to $\frac{1}{4N}$. In the first three experiments, the initial guess $X^{(0)}$ was set to the pseudo-identity matrix $I_{p,n}$. Each random sample X_k was generated by the rule $X = CW\cos(\Sigma)W^T + U\sin(\Sigma)W^T$, where $U\Sigma W^T$ is the thin singular value decomposition of the random initial velocity matrix V, generated by

the rule $V = a(I - CC^T)A$, with A being a $p \times n$ matrix whose entries are random real-valued numbers drawn from a normal distribution.

The first numerical experiment aims at comparing the values of discrepancy-to-center pertaining to the averaging algorithm (28), where the tangent-bundle maps were selected as the Thin-QR-decomposition-based maps, Polar-decomposition-based maps, Geodesic-based maps and Exponential/Logarithm maps, and pertaining to the averaging algorithm (30) with both Brute-force Cayley-transform-based pseudo-tangent-bundle maps and Economical Cayley-transform-based pseudo-tangent-bundle maps. The following considerations are in order:

- The Thin-QR-decomposition-based maps and the Polar-decomposition-based maps are equivalent (by virtue of Theorem 9). Hence, only the Thin-QR-decomposition-based maps are considered in the comparison.
- The Geodesic-based maps arose as a special implementation that aims at simplifying the Exponential/Logarithm maps thanks to the *symmetrization trick*. Hence, only the Geodesic-based maps are considered in the comparison.
- The Brute-force Cayley-transform-based pseudo-tangent-bundle maps need the *un-symmetrization trick* to be put into effect. The post-multiplication of the argument of the map $\hat{P}_X(Y)$ by a *random* orthogonal matrix was chosen.

The number of fixed-point iterations was set to 100 in order to verify the numerical stability of the devised fixed-point algorithms and the spread parameter a was set to 0.05. The number of samples to average was set to $N = 100$ and the base-manifold was $\mathrm{Gr}(20, 4)$. The Figure 1 shows the values of discrepancy-to-center measure $\varepsilon(X^{(i)}, C)$, while the Figure 2 shows the differential discrepancy $\varepsilon(X^{(i)}, X^{(i-1)})$, pertaining to the same algorithms. While the algorithms based on the fixed-point iteration rule (28) perform satisfactorily, the algorithm based on the $\mathfrak{so}(p)$ parameterization does not behave well. From an algorithmic viewpoint, the fixed-point iteration (30) endowed with Brute-force Cayley-transform-based pseudo-tangent-bundle maps tries to achieve convergence but it becomes numerically unstable, eventually. A reason of instability is due to the structure of the tangent-bundle pseudo-map $\hat{P}_X^{-1}(Y)$ and, in particular, of the submatrix $Y^TX - X^TY$, that eventually comes too close to singularity. In order to confirm numerically such observation, at each iteration i of the fixed-point algorithm (30) endowed with Brute-force Cayley-transform-based pseudo-tangent-bundle maps, the following quantity was recorded:

$$\min_k \{\|X_k^T X^{(i)} - (X^{(i)})^T X_k\|_{\mathrm{F}}\}, \tag{31}$$

as an indicator of the minimal norm of the matrix $X_k^T X^{(i)} - (X^{(i)})^T X_k$ for each sample. The Figure 3 shows the minimal norm (31) versus the number of iterations. The values shown in the Figure 3 were recorded at the same time of the values shown in the Figures 1 and 2. It may be observed that the algorithm (30) becomes unstable after the minimal norm reaches sufficiently low values. In conclusion, the Thin-QR-decomposition-based tangent-bundle maps jointly with the Polar-decomposition-based tangent-bundle maps, the Geodesic-based tangent-bundle maps jointly with the Exponential/Logarithmic maps, and the economical Cayley-transform-based pseudo-tangent-bundle maps are to be preferred over the Cayley-transform-based pseudo-tangent-bundle maps, from a numerical stability viewpoint.

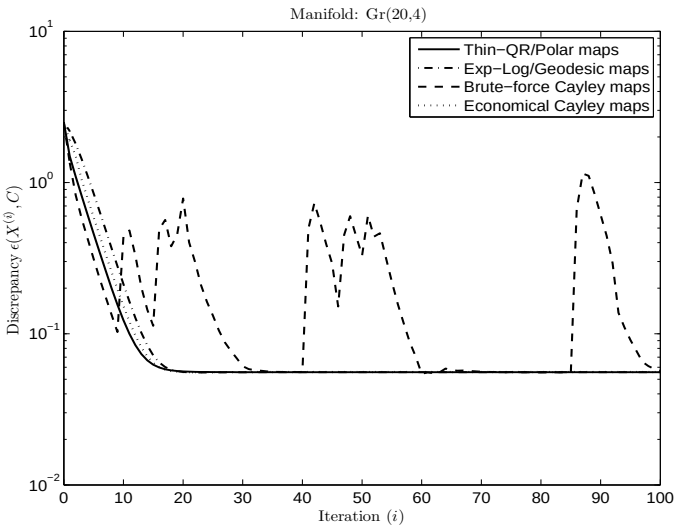

Fig. 1. Averaging over the Grassmann manifold $\mathrm{Gr}(20, 4)$: Discrepancy-to-center pertaining to the averaging algorithm (28), where the tangent-bundle maps were selected as the Thin-QR-decomposition-based maps and as the Geodesic-based maps, and to the averaging algorithm (30).

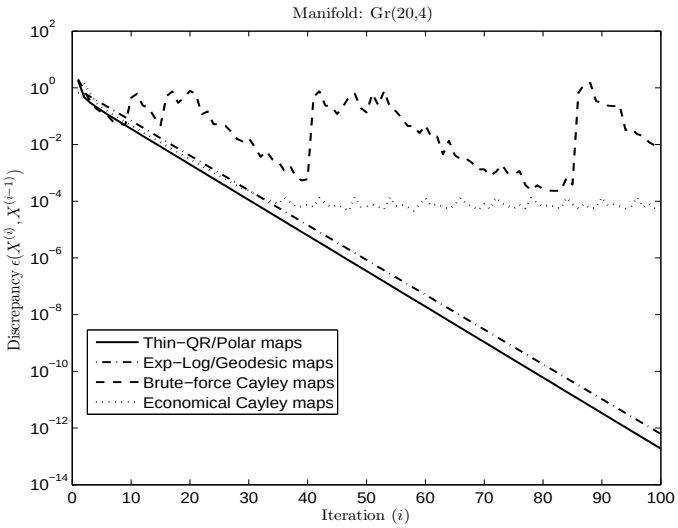

Fig. 2. Averaging over the Grassmann manifold $\mathrm{Gr}(20, 4)$: Differential discrepancy pertaining to the averaging algorithm (28), where the tangent-bundle maps were selected as the Thin-QR-decomposition-based maps and as the Geodesic-based maps, and to the averaging algorithm (30).

The second numerical experiment aims at evaluating the average performances of the algorithm (28) endowed with the Thin-QR/Polar-decomposition-based tangent-bundle maps, the Exponential/logarithmic/Geodesic-based tangent-bundle maps, as well as of the algorithm (30) endowed with the Brute-force Cayley-transform-based and the Economical Cayley-transform-based pseudo-tangent-bundle maps. The Figure 4 shows the discrepancy to center after completion of each run (of 50 iterations each), normalized to the discrepancy to

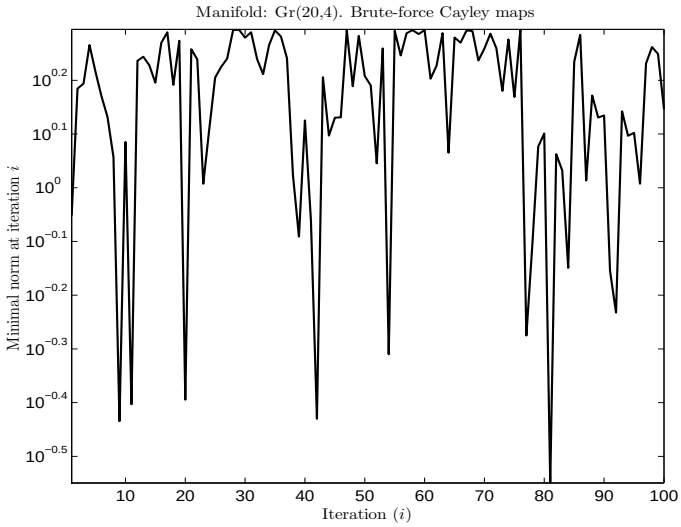

Fig. 3. Averaging over the Grassmann manifold $\mathrm{Gr}(20, 4)$ by the fixed-point algorithm (30) and Brute-Force Cayley map: Minimal norm (31) versus number of iterations.

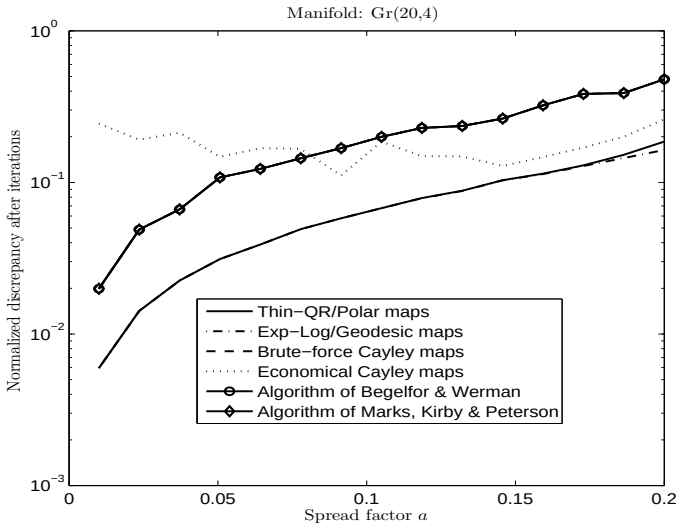

Fig. 4. Average performances of the fixed-point averaging algorithms for 15 increasing values of the spread parameter a: Discrepancy to center $\varepsilon(X^{(50)}, C)$, with $X^{(50)}$ denoting the mean-matrix estimated after $N = 50$ iterations of the averaging algorithms, normalized to the discrepancy to center at iteration 0.

center at iteration 0, averaged over 100 independent trials to get rid of random fluctuations, on 15 increasing values of the spread parameter a. The number of samples to average was set to $N = 50$ and the Grassmann manifold was again $\mathrm{Gr}(20, 4)$. For comparison purpose, the algorithm to average over a Grassmann manifold drawn from the paper [2] (referred to as 'Algorithm of Begelfor & Werman') and the algorithm drawn from the contribution [24] (referred to as 'Algorithm of Marks, Kirby & Peterson') were tested on the same data sets. The obtained results confirm that the Thin-QR/Polar-decomposition-based maps and the Exponential/logarithmic/Geodesic-based maps perform the best in conjunction with the fixed-point algorithm (28), although, for some values of the spread, the Economical Cayley-

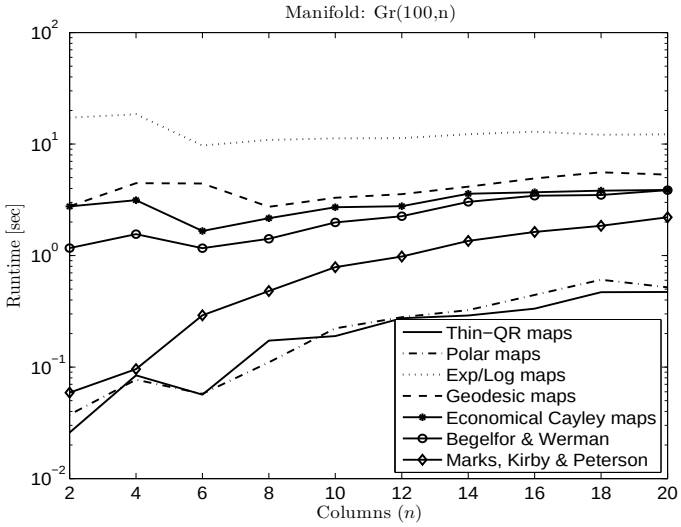

Fig. 5. Averaging over the Grassmann manifold $\mathrm{Gr}(100, n)$ by the fixed-point averaging algorithms: Runtime (CPU) versus increasing dimension n of subspaces. The number of samples to average is $N = 50$ and the shown results are averaged over 100 independent trials.

transform-based map performs comparably. In addition, the algorithms based on the Thin-QR/Polar-decomposition-based maps and the Exponential/logarithmic/Geodesic-based maps outperform the algorithm of Begelfor & Werman as well as the algorithm of Marks, Kirby & Peterson.

The third numerical experiment aims at evaluating the computational complexity of the algorithm (28) endowed with the Thin-QR-decomposition-based tangent-bundle maps, the Polar-decomposition-based tangent-bundle maps, the Exponential/logarithmic tangent-bundle maps and the Geodesic-based tangent-bundle maps, as well as of the algorithm (30) endowed with the Economical Cayley-transform-based pseudo-tangent-bundle maps, in comparison with the algorithm of Begelfor & Werman [2] and the algorithm of Marks, Kirby & Peterson [24]. The computational burden of the algorithms is evaluated in terms of CPU runtime. The Brute-Force Cayley-transform-based pseudo-tangent-bundle maps are excluded because of their poor numerical performances. The Figure 5 shows the CPU runtime taken by the algorithms to complete 20 iterations (averaged over 100 independent trials to get rid of random fluctuations), over a Grassmann manifold $\mathrm{Gr}(p, n)$ with index $p = 100$ fixed and index n increasing from 2 to 20 with step 2 and with the spread parameter a set to 0.05. From the obtained results, it may be concluded that the Thin-QR-decomposition-based tangent-bundle maps and the Polar-decomposition-based tangent-bundle maps are the most convenient ones to implement the averaging algorithm (28). The Exponential/Logarithmic maps are the most computationally demanding among the considered maps pairs and in the considered ranges, while the Geodesic-based maps, the Economical Cayley maps, the algorithm of Begelfor & Werman and the algorithm of Marks, Kirby & Peterson locate in an intermediate position as for their computational complexity. It should be noted that, while the computational burden of the Thin-QR-decomposition-based tangent-bundle maps, of

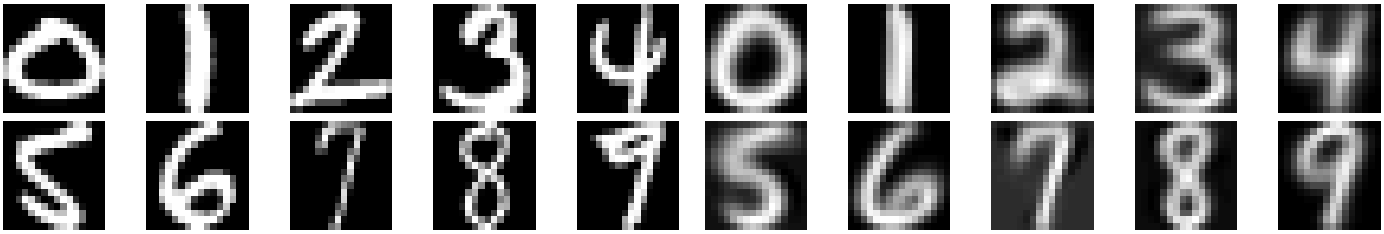

Fig. 6. Example of the data set for the *handwritten digit classification problem* used in the fourth numerical experiment.

Fig. 7. Example of digits reconstructed upon projection on $\mathrm{Gr}(256, 2)$.

the Polar-decomposition-based tangent-bundle maps and of the algorithm of Marks, Kirby & Peterson increases quickly with the increase of the number n, the computational complexity of the Exponential/Logarithmic maps, of the Geodesic-based maps, of the Economical Cayley maps and of the algorithm of Begelfor & Werman increases slowly. For a Grassmann manifold $\mathrm{Gr}(p, n)$ with index n of the same order of the index p, such advantage may, therefore, vanish.

The fourth numerical experiment aims at testing the empirical arithmetic averaging algorithm (28) endowed with the Thin-QR-decomposition-based tangent-bundle maps on a real-world data set, and to compare it with existing algorithms found in the scientific literature, namely, the algorithm to average over a Grassmann manifold drawn from the paper [2] (again referred to as 'Algorithm of Begelfor & Werman') and the algorithm drawn from the contribution [24] (again referred to as 'Algorithm of Marks, Kirby & Peterson'). The data-set of choice concerns a *handwritten digit classification problem*. In the recognition of images, such as handwritten characters and human faces, it is a common practice to extract a limited number of abstract features, such as the principal eigenvectors of covariance matrices of vectorized images. Such eigenvectors are obtained by, e.g., principal component analysis. In order to avoid storing in memory all the images and all the extracted eigenvectors, the average subspace may be calculated. The present experiment concerns handwritten digits data, with digit from 0 to 9, where each sample consists of 16×16 pixels and a total of 500 samples are available for each digit. The data were obtained from the USPS data set and are described in [16]. An example of the appearance of the ten handwritten digits is shown in the Figure 6. The data are pre-processed in the following way:

1) Within each digit subset, each 16×16 image gets vectorized (by a lexicographic order) to make a 1×256 vector.
2) Within each digit class, the 500 available 1×256 vectors are made zero-mean and their 256×256 covariance matrix is computed.
3) Within each digit class, the subspace spanned by the two principal eigenvectors of the covariance matrix is estimated (cf. [41], Ss. IV-E). Such subspaces are elements of the Grassmannian $\mathrm{Gr}(256, 2)$ and are represented by a 256×2 matrix denoted by X_ℓ^{all}, for $\ell = 0, \ldots, 9$.

The Grassmannian $\mathrm{Gr}(256, 2)$ is sufficient to capture the essence of the handwritten digits, as testified by the images reconstructed upon projection on $\mathrm{Gr}(256, 2)$ shown in the Figure 7. Next, for each digit class, the 500 available images were partitioned into 10 sub-subsets of 50 images each. Within each sub-subset, the subspace spanned by the first two principal eigenvectors were estimated from the sub-subset's covariance matrix, obtaining 10 subspace estimates for each class. Then, for each class ℓ separately, with $\ell = 0, \ldots, 9$, the average of the 10 subspaces in $\mathrm{Gr}(256, 2)$ was computed using:

- The arithmetic averaging algorithm (28) endowed with the Thin-QR-decomposition-based tangent-bundle maps, whose result is represented by X_ℓ^{qr},
- The algorithm of Begelfor & Werman, whose result is represented by X_ℓ^{bw},
- The algorithm of Marks, Kirby & Peterson, whose result is represented by X_ℓ^{mkp}.

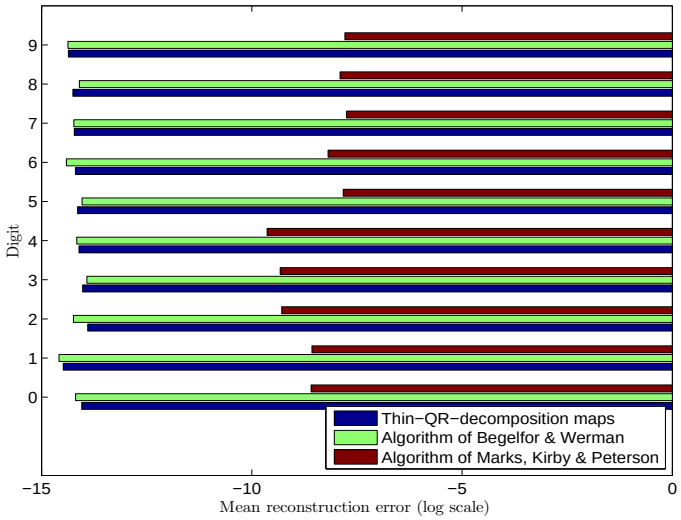

Fig. 8. Mean reconstruction error for each digit class pertaining to each of the three considered Grassmann averaging algorithms.

The cumulative result of the the above analysis are shown in the Figure 8, that shows the *mean reconstruction error* obtained by the projectors $X_\ell^{\mathrm{qr}}(X_\ell^{\mathrm{qr}})^T$, $X_\ell^{\mathrm{bw}}(X_\ell^{\mathrm{bw}})^T$ and $X_\ell^{\mathrm{mkp}}(X_\ell^{\mathrm{mkp}})^T$ applied to each digit-sample within the class ℓ, and in the Figure 9, that shows the discrepancy between the subspaces represented by X_ℓ^{qr}, X_ℓ^{bw} and with the subspace represented by X_ℓ^{all}. The arithmetic averaging algorithm (28) endowed with the Thin-QR-decomposition-based tangent-bundle maps and the algorithm of Begelfor & Werman perform comparably, while the algorithm of Marks, Kirby & Peterson compares unfavorably with the former two algorithms.

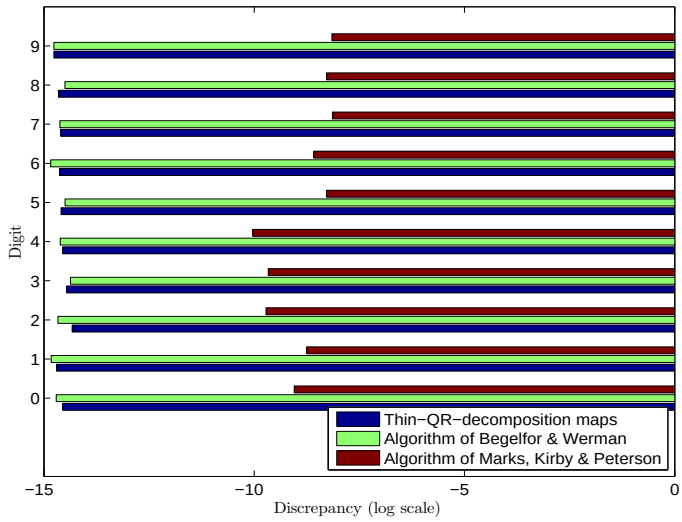

Fig. 9. Discrepancy between the average subspaces estimated by three algorithms and the subspace obtained by all samples for each digit class.

VI. Conclusions

The aim of the present contribution is to devise, discuss and compare a number of tangent-bundle (pseudo-) maps pairs to parameterize elements of the Grassmann manifold in view of computing average subspaces.

In particular, the present paper presents and discusses several tangent-bundle maps, namely, maps based on the thin QR matrix decomposition, on the polar matrix decomposition, on exponential/logarithmic maps and on geodesic lines, along with details about the analytic features of such devised maps and of their (exact or approximate) interrelationships. The paper also presents two pairs of pseudo-tangent-bundle maps, which exploit the quotient space structure of the Grassmann manifold. The present paper also introduces two empirical arithmetic averaging algorithms based on a fixed-point iteration scheme and illustrates their numerical features through numerical experiments that investigate on their numerical performances as well as on their computational complexity.

The performed numerical experiments on synthetic data revealed that the Thin-QR-decomposition-based tangent-bundle maps, the Polar-decomposition-based tangent-bundle maps, the Geodesic-based tangent-bundle maps, the Exponential/Logarithmic maps and the Economical Cayley-transform-based pseudo-tangent-bundle maps are to be preferred over the Brute-Force Cayley-transform-based pseudo-tangent-bundle maps, from a numerical stability viewpoint. Moreover, for Grassmann manifolds $\mathrm{Gr}(p,n)$ with index $n \ll p$, the Thin-QR-decomposition-based tangent-bundle maps and the Polar-decomposition-based tangent-bundle maps are the most convenient ones to implement the averaging algorithm (28), while the Exponential/Logarithmic maps are the most computationally demanding among the considered maps pairs. The numerical experiments allow the conclusion that the Thin-QR-decomposition- and the the Polar-decomposition-based tangent-bundle maps, which are equivalent to each other, are to be preferred in the actual implementations. A further test performed on an handwritten digit data set revealed that the arithmetic averaging algorithm (28) endowed with the Thin-QR-decomposition-based tangent-bundle maps and the algorithm of Begelfor & Werman perform comparably, in terms of reconstruction error as well as as subspace discrepancy, while the algorithm of Marks, Kirby & Peterson compares unfavorably in the analyzed test conditions.

Acknowledgments

We would like to gratefully thank the reviewers for the detailed and constructive comments, which helped improving the thoroughness and clarity of the manuscript. We also wish to acknowledge the help of Takashi Uehara who assisted with the computer experiments on the handwritten digits dataset.

References

[1] L. Balzano, R. Nowak and B. Recht, "Online identification and tracking of subspaces from highly incomplete information," in *Proceedings of the* 48^{th} *Annual Allerton Conference on Communication, Control, and Computing* (Allerton House, Monticello (IL, USA), September 29 - October 1, 2010), pp. 704 – 711, 2010.

[2] E. Begelfor and M. Werman, "Affine invariance revisited," in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (New York (NY, USA), June 17-22, 2006), Vol. 2, pp. 2087 – 2094, 2006.

[3] O. Besson, N. Dobigeon and J.-Y. Tourneret, "Minimum mean square distance estimation of a subspace," *IEEE Transactions on Signal Processing*, Vol. 59, No. 12, pp. 5709 – 5720, December 2011.

[4] H.E. Çetingül and R. Vidal, "Intrinsic mean shift for clustering on Stiefel and Grassmann manifolds," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR 2009, Center for Imaging Sciences, Johns Hopkins University, Baltimore (MD, USA), June 20-25, 2009), pp. 1896 – 1902, 2009.

[5] J.-M. Chang and J.I. Pacheco, "Classifying handwritten digits on the Grassmann manifold," in *Proceedings of the* 15^{th} *International Conference on Image Processing, Computer Vision and Pattern Recognition* (IPCV'11, Las Vegas (NV, USA), July 18-21, 2011), pp. 36 – 41, 2011.

[6] W. Dai, Y. Liu and B. Rider, "Quantization bounds on Grassmann manifolds and applications to MIMO communications," *IEEE Transactions on Information Theory*, Vol. 54, No. 3, pp. 1108 – 1123, March 2008.

[7] X. Dong, P. Frossard, P. Vandergheynst and N. Nefedov, "Clustering on multi-layer graphs via subspace analysis on Grassmann manifolds," *IEEE Transactions on Signal Processing*, Vol. 62, No. 4, pp. 905 – 918, February 2014.

[8] A. Edelman, T.A. Arias and S.T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Journal on Matrix Anaysis and Applictions*, Vol. 20, No. 2, pp. 303 – 353, 1998.

[9] Y. Fan, Y. Liu, H. Wu, Y. Hao, H. Liu, Z. Liu and T. Jiang, "Discriminant analysis of functional connectivity patterns on Grassmann manifold," *Neuroimage*, Vol. 56, No. 4, pp. 2058 – 2067, June 2011.

[10] S. Fiori and T. Tanaka, "An algorithm to compute averages on matrix Lie groups," *IEEE Transactions on Signal Processing*, Vol. 57, No. 12, pp. 4734 – 4743, December 2009.

[11] S. Fiori, "Extended Hamiltonian learning on Riemannian manifolds: Numerical aspects," *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 23, No. 1, pp. 7 – 21, January 2012.

[12] K. Fujii, "Introduction to Grassmann manifolds and quantum computation," *Journal of Applied Mathematics*, Vol. 2, No. 8, pp. 371 – 405, 2002.

[13] R.H. Gohary and T.N. Davidson, "An explicit expression for the Newton direction on the complex Grassmann manifold," *IEEE Transactions on Signal Processing*, Vol. 59, No. 3, pp. 1303 – 1309, March 2011.

[14] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 3^{rd} Edition, Johns Hopkins, 1996.

[15] P. Gruber and F.J. Theis, "Grassmann clustering," in *Proceedings of the* 14^{th} *European Signal Processing Conference* (EUSIPCO 2006, Florence (Italy), September 4-8, 2006), 2006.

[16] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning*, Springer, 2001.

[17] H.V. Henderson and S.R. Searle, "On deriving the inverse of a sum of matrices," *SIAM Review*, Vol. 23, No. 1, pp. 53 – 60, January 1981.

[18] N.J. Higham and R.S. Schreiber, "Fast polar decomposition of an arbitrary matrix", *SIAM Journal on Scientific and Statistical Computation*, Vol. 11, No. 4, pp. 648 – 655, July 1990.
[19] T. Kaneko, S. Fiori and T. Tanaka, "Empirical arithmetic averaging over the compact Stiefel manifold," *IEEE Transactions on Signal Processing*, Vol. 61, No. 4, pp. 883 – 894, February 2013.
[20] O. Kuybeda, D. Malah and M. Barzohar, "Anomaly preserving $\ell_{2,\infty}$-optimal dimensionality reduction over a Grassmann manifold," *IEEE Transactions on Signal Processing*, Vol. 58, No. 2, pp. 544 – 552, February 2010.
[21] J.D. Lawson and Y. Lim, "The geometric mean, matrices, metrics, and more," *The American Mathematical Monthly*, Vol. 108, No. 9, pp. 797 – 812, November 2001.
[22] Y.M. Lui and J.R. Beveridge, "Grassmann registration manifolds for face recognition," in *Proceedings of the 2008 European Conference on Computer Vision* (Palais des Congrès Parc Chanot, Marseille (France), October 12-18, 2008), 2008.
[23] J. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Transactions on Signal Processing*, Vol. 50, No. 3, pp. 635 – 650, March 2002.
[24] J. Marks, M. Kirby and C. Peterson, "A Normal/Tangent Bundle Algorithm for Representing Point Clouds on Grassmann and Stiefel Manifolds," Submitted (2013).
[25] C.D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM (Society for Industrial and Applied Mathematics), 2000.
[26] B. Mishra, G. Meyer, S. Bonnabel and R. Sepulchre, "Fixed-rank matrix factorizations and Riemannian low-rank optimization." Report publicly available on the arXiv archive `http://arxiv.org/abs/1209.0430`, April 2013.
[27] S. Mittal and P. Meer, "Conjugate gradient on Grassmann manifolds for robust subspace estimation," *Image and Vision Computing*, Vol. 30, pp. 417 – 427, 2012.
[28] B. Mondal, S. Dutta and R.W. Heath, "Quantization on the Grassmann Manifold," *IEEE Transactions on Signal Processing*, Vol. 55, No. 8, pp. 4208 – 4216, August 2007.
[29] T.T. Ngo and Y. Saad, "Scaled gradients on Grassmann manifolds for matrix completion," in *Proceedings of the Neural Information Processing Systems* (NIPS, Lake Tahoe (NV, USA), December 3-6, 2012), pp. 1421 – 1429, 2012.
[30] D. Nowicki and O. Dekhtyarenko, "Averaging on Riemannian manifolds and unsupervised learning using neural associative memory," in *Proceedings of the 2005 European Symposium on Artificial Neural Networks* (ESANN'2005, Bruges (Belgium), April 27-29, 2005), 2005.
[31] M. Rathinam and L.R. Petzold, "A new look at proper orthogonal decomposition," *SIAM Journal on Numerical Analysis*, Vol. 41, No. 5, pp. 1893 – 1925, 2003.
[32] A. Srivastava and E. Klassen, "Monte Carlo extrinsic estimators of manifold-valued parameters," *IEEE Transactions on Signal Processing*, Vol. 50, No. 2, pp. 299 – 308, February 2002.
[33] A. Srivastava and E. Klassen, "Bayesian and geometric subspace tracking," *Advances in Applied Probability*, Vol. 36, No. 1, pp. 43 – 56, 2004.
[34] R. Subbarao and P. Meer, "Nonlinear mean shift over Riemannian manifolds," *International Journal on Computer Vision*, Vol. 63, No. 84, pp. 1 – 20, 2009.
[35] X.-C. Sun, X. Wang, Q. Cheng and J. Feng, "On subspace distances," in *Proceeding of the Third International Conference on Image Analysis and Recognition* (ICIAR 2006, Póvoa de Varzim, Portugal, September 18-20, 2006), Part II, Lecture Notes in Computer Science, Volume 4142, pp 81 – 89, 2006.
[36] C.H. Suryanto, H. Saigo and K. Fukui, "Protein clustering on a Grassmann manifold," in *Proceedings of the* 7^{th} *IAPR International Conference* (Tokyo (Japan), November 8-10, 2012), Lecture Notes in Computer Science, Vol. 7632, pp. 71 – 81, 2012.
[37] L. N. Trefethen and D. Bau III, *Numerical Linear Algebra*, 1^{st} Edition, Society for Industrial and Applied Mathematics, 1997.
[38] P. Turaga, A. Veeraraghavan, A. Srivastava and R. Chellappa, "Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 11, pp. 2273 – 2286, 2011.
[39] C. Van Loan, "Computing the CS and the Generalized Singular Value Decomposition," *Numerische Mathematik*, Vol. 46, No. 4, pp. 479 – 491, 1985.
[40] W. Xu and B. Hassibi, "Compressed sensing over the Grassmann manifold: A unified analytical framework," in *Proceedings of the* 46^{th} *Annual Allerton Conference on Communication, Control, and Computing* (Allerton House, Monticello (IL, USA), September 23-26, 2008), pp. 562 – 567, 2008.
[41] Y. Washizawa and S. Hotta, "Mahalanobis distance on extended Grassmann manifolds for variational pattern analysis," *IEEE Transactions on Neural Networks and Learning Systems*. To appear.

Simone Fiori received the Italian Laurea (Dr. Eng.) with honors in electronics engineering in July 1996 from the University of Ancona (Italy), and the Ph.D. degree in electrical engineering (circuit theory) in March 2000 from the University of Bologna (Italy). In November 2005, he joined the Università Politecnica delle Marche. His research interests include linear and non-linear adaptive discrete-time filter theory and geometrical methods for machine learning and signal processing. He is author of more than 85 refereed journal papers and more than 75 conference/workshop papers. Dr. Fiori was the recipient of the 2001 "E.R. Caianiello Award" for the best Ph.D. dissertation in the artificial neural network field and the 2010 "Rector Award" as a proficient researcher. He is currently serving as Associate Editor of Neurocomputing (Elsevier) and Cognitive Computation (Springer). Dr. Fiori was awarded several scholarships to visit the AIST research institute (Tsukuba, Ibaraki prefecture, Japan), the RIKEN research institute (Wako-shi campus, Saitama prefecture, Japan), the Tokyo University of Agriculture and Technology (Koganei campus, Tokyo, Japan), the Trondheim Technical University (Trondheim, Norway), the University of Bari (Bari, Italy), the Bioinformatics Institute (Biopolis, Singapore) and the National Taiwan University (Taipei, Republic of China). He is a member of APSIPA.

Tetsuya Kaneko received the B.E. and the M.E. from the Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology, in 2011 and in 2013, respectively.

Toshihisa Tanaka (S'98-M'02-SM'10) received the B.E., M.E., and Ph.D. degrees from the Tokyo Institute of Technology, Tokyo, Japan, in 1997, 2000, and 2002, respectively. From 2000 to 2002, he was a Research Fellow at the Japan Society for the Promotion of Science. From October 2002 to March 2004, he was a Research Scientist at RIKEN Brain Science Institute, Saitama, Japan. In April 2004, he joined Department of Electrical and Electronic Engineering, the Tokyo University of Agriculture and Technology, Tokyo, where he is currently an Associate Professor. In 2005, he was a Royal Society Visiting Fellow at the Communications and Signal Processing Group, Imperial College London, U.K. From June 2011 to October 2011, he was a Visiting Faculty Member in the Department of Electrical Engineering, University of Hawaii at Manoa, Honolulu, HI, USA. His research interests include image and signal processing, statistical signal processing and machine learning, brain and biomedical signal processing, and adaptive systems. He is a coeditor of Signal Processing Techniques for Knowledge Extraction and Information Fusion (with Mandic, Springer), 2008. Dr. Tanaka served as a Guest Editor of special issues in journals including Neurocomputing. He served as an Associate Editor of the IEICE Transactions on Fundamentals. He was a Chair of the Technical Committee on Biomedical Signal Processing, Asia-Pacific Signal and Information Processing Association (APSIPA). He is a member of the Institute of Electronics, Information and Communication Engineers and APSIPA.