Empirical Arithmetic Averaging over the Compact Stiefel Manifold

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

***Abstract*— The aim of the present research work is to investigate algorithms to compute empirical averages of finite sets of sample-points over the Stiefel manifold by extending the notion of Pythagoras' arithmetic averaging over the real line to a curved manifold. The idea underlying the developed algorithms is that sample-points on the Stiefel manifold get mapped onto a tangent space, where the average is taken, and then the average point on the tangent space is brought back to the Stiefel manifold, via appropriate maps. Numerical experimental results are shown and commented on in order to illustrate the numerical behaviour of the proposed procedure. The obtained numerical results confirm that the developed algorithms converge steadily and in a few iterations and that they are able to cope with relatively large-size problems.**

***Index Terms*— Arithmetic averaging, Matrix manifolds, Empirical averaging on matrix manifolds, Manifold retraction, QR-decomposition, Polar decomposition, Cayley transform, Orthographic projection.**

I. INTRODUCTION

REPRESENTATIONS involving structured matrices, such as orthogonal, symplectic, Toeplitz Hermitian-positive matrices, special Euclidean matrices and unitary matrices, arise in signal processing. Known situations are principal component analysis and independent component analysis by signal pre-whitening [15], radio interferometry [25] and optical system modelling [19]. Moreover, in statistical data processing, the data may appear under the form of random structured matrices (see, e.g., [16]). Random matrix theory is an important and active research area and it finds applications in fields as diverse as physics, wireless communications and information theory. A useful statistical characterization of a set of structured matrices is their empirical mean, which appears as an average matrix carrying on the same structure of the data themselves. Averaging over a data-set is a good method to smooth-out data and to alleviate measurement errors and random fluctuations. Occasionally, the median instead of the mean is made use of (see, e.g., [8]): Theorems of existence of medians on manifolds recently appeared in [3] with applications to Toeplitz Hermitian positive-definite matrices [7].

In the case of unconstrained data, such as, for example, in the case that the matrix-type data belong to the flat space $\mathbb{R}^{n\times n}$, simple arithmetic averaging produces the desired result.

T. Kaneko and T. Tanaka are 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).

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

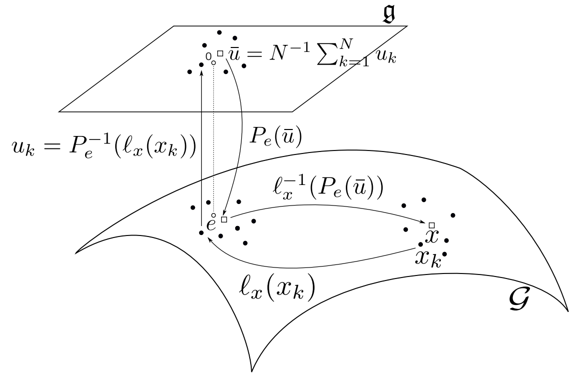

Fig. 1. Illustration of the averaging algorithm on the Lie group $\mathcal{G}$ proposed in [21]. The dots (•) denote sample elements and the box symbol (□) denotes the empirical average element. The procedure of computing an average element works as follows. The first step is to perform a left-translation ℓ_x of each sample $x_k \in \mathcal{G}$ about the sought-for average element $x \in \mathcal{G}$ so as to move all translated samples to a neighbourhood of the identity element e of the Lie group $\mathcal{G}$. The next step is, by means of a lifting map P_e^{-1}, to map all samples onto the Lie algebra $\mathfrak{g}$ and to compute their arithmetic mean denoted by $\overline{u}$. The final step consists in mapping the Lie-algebra element $\overline{u}$ onto $\mathcal{G}$ by exploiting a retraction map P_e and then to get the mean element x by the inverse left-translation ℓ_x^{-1}.

However, in the case that constraint conditions – such as orthogonality – are to be taken into account, arithmetic averaging does not produce any sensible result (for example, the result of entry-by-entry addition of two orthogonal matrices is not orthogonal, in general). Therefore, to compute statistics on spaces of matrices, it is necessary to build up algorithms that take into account the geometric structure of the (generally curved) space that those matrices belong to. See, for example, the computation of the *mean shift* on Riemannian matrix manifolds [38], the rich corpus of results about the space of symmetric positive-definite matrices [10], [31] and, as a generalization, the Bayesian analysis of the statistics of Stiefel-matrices that represent subspaces in a Grassmann-manifold setting [9], [34].

Fiori and Tanaka [21] presented a general-purpose averaging algorithm that works for matrix Lie groups and, in particular, for the space of special orthogonal matrices, that are square matrices with mutually orthogonal unitary-norm columns and such that their determinant is positive (namely, they represent high-dimensional rotations). A Lie group is an algebraic group endowed with a manifold structure compatible with its algebraic structure. The method presented in [21] exploits the relationship between a Lie group and its associated Lie algebra and is illustrated in Figure 1. Although there is a connection between the method proposed in [21] and the notion of Riemannian or Karcher mean [29], the substantial difference is that the Riemannian mean is defined on the basis of a dispersion criterion that involves the Riemannian (or geodesic) distance between two points, while the method proposed in [21] does not involve any metrics and is hence more general, in this regard. Other methods have been recently proposed in the literature to tackle the problem of average computation, as, for example, the method based on stochastic flow to compute averages on manifolds [4].

The method presented in [21] relies on a class of functions, termed 'lifting maps', that map an element of the Lie group to the Lie algebra and on a class of functions, termed 'retraction maps', that map an element of the Lie algebra onto the Lie group. The problem of *averaging on non-Lie-group-type manifolds is substantially more difficult* because the calculation of appropriate retraction (and especially of the lifting) maps is a substantially more involved and less studied problem. Hence, averaging on non-Lie-group-type manifolds cannot be achieved by any trivial extension of the method proposed in [21]. It could be tackled as a Riemannian-mean or Karcher-mean computation problem, but in some cases a distance function on manifolds of interest may be unavailable in closed form. In particular, the problem of averaging on the compact Stiefel manifold (the space of orthogonal rectangular matrices), which is not a Lie group, is worth analyzing because a number of signal-processing applications require statistical computation over the Stiefel manifold, such as data clustering [13], image and video-based recognition [40], Bayesian filtering [39] and manifold learning for expression analysis and human motion analysis [24]. The aim of the present research work is to extend the algorithm introduced in the paper [21] to compute averages over the compact Stiefel manifold. The idea behind the developed algorithms is that points on the Stiefel manifold are mapped onto a tangent space where the average over mapped points is taken, and then the average point on the tangent space is brought back to the Stiefel manifold. To summarize the method in a sentence, *the average of samples on the Stiefel manifold is computed by applying a retraction to the arithmetic average of the lifted samples*.

Most of the research work described in this paper concerns the individuation of appropriate retraction/lifting maps for the Stiefel manifold and of efficient ways to implement them, as explained in section II. In particular, in the present research work, a QR-decomposition based retraction map, a polar-decomposition-based retraction map, an orthographic retraction map and two Cayley-transform-based pseudo-retraction maps, along with their associated lifting and pseudo-lifting maps, are studied and tested numerically. The results of several numerical tests are illustrated in the section III. Section IV draws the conclusions.

II. Arithmetic averaging algorithms on the compact Stiefel Manifold

The aim of this section is to build up arithmetic averaging algorithms on the Stiefel manifold based on the notion of manifold retraction. In particular, the proposed method is based on a fast fixed-point algorithm. The compact Stiefel manifold is defined by:

$$\mathrm{St}(p,n) \stackrel{\text{def}}{=} \{X \in \mathbb{R}^{p\times n} | X^T X = I_n\}, \quad (1)$$

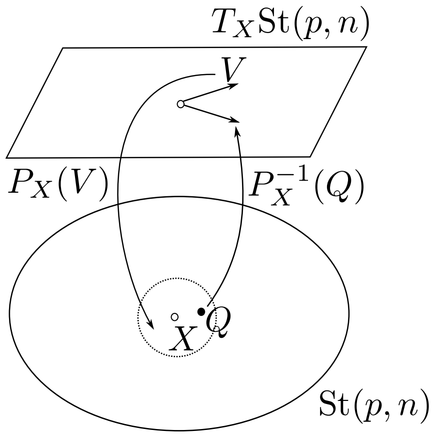

Fig. 2. Illustration of the notion of retraction map P_X and of lifting map P_X^{-1} about a point $X \in \mathrm{St}(p,n)$. The dashed-circle represents the domain of definition of the lifting map.

where symbol I_n denotes a $n \times n$ identity matrix and $n \leq p$, namely, the manifold $\mathrm{St}(p,n)$ is the space of the 'tall-skinny' orthogonal matrices. The compact Stiefel manifold has dimension $pn - \frac{n(n+1)}{2}$. Its tangent space at a point $X \in \mathrm{St}(p,n)$ may be expressed as:

$$T_X\mathrm{St}(p,n) = \left\{V \in \mathbb{R}^{p\times n} | X^T V + V^T X = 0\right\}. \quad (2)$$

Each tangent space is a vector space of dimension $pn - \frac{n(n+1)}{2}$ under standard matrix addition and multiplication by a real-valued scalar.

A *retraction* at a point $X \in \mathrm{St}(p,n)$ of a Stiefel manifold is a map $P_X : T_X\mathrm{St}(p,n) \to \mathrm{St}(p,n)$ such that, for each tangent space $T_X\mathrm{St}(p,n)$, it holds that [11]:

1) The retraction P_X is defined in some open ball about $0 \in T_X\mathrm{St}(p,n)$.
2) It holds that $P_X(0) = X$.
3) It holds that $\frac{d}{dt}P_X(tV)\big|_{t=0} = V$.

A retraction induces local coordinates on the manifold $\mathrm{St}(p,n)$. A map $P_X^{-1} : \mathrm{St}(p,n) \to T_X\mathrm{St}(p,n)$ such that $P_X(P_X^{-1}(Q)) = Q$, for $Q \in \mathrm{St}(p,n)$, is termed *lifting* map. A lifting map is defined only locally and is not unique, in general. The notions of retraction map and of lifting map are illustrated in Figure 2.

The compact Stiefel manifold $\mathrm{St}(p,n)$ is a submanifold of the Euclidean space $\mathbb{R}^{p\times n}$. As such, a normal space $N_X\mathrm{St}(p,n)$ may be associated to each point $X \in \mathrm{St}(p,n)$, which is given by:

$$N_X\mathrm{St}(p,n) = \left\{XS | S^T = S \in \mathbb{R}^{n\times n}\right\}. \quad (3)$$

The special orthogonal group of matrices, denoted by $\mathrm{SO}(p)$, is defined as:

$$\mathrm{SO}(p) \stackrel{\text{def}}{=} \{G \in \mathbb{R}^{p\times p} | G^T G = I_p, \det(G) = 1\}. \quad (4)$$

It is a Lie group under standard matrix multiplication and inversion, with the matrix I_p being its identity element. Its associated Lie algebra is:

$$\mathfrak{so}(p) \stackrel{\text{def}}{=} \{\Omega \in \mathbb{R}^{p\times p} | \Omega^T = -\Omega\}, \tag{5}$$

namely, it is the set of $p \times p$ skew-symmetric matrices. The Lie algebra $\mathfrak{so}(p)$ is a vector space of dimension $\frac{p(p-1)}{2}$.

For a general reference on differential geometry, readers might consult, e.g., the set of books [37].

As it is instrumental in the development of the averaging algorithms in the following subsections, it is worth recalling the notion of *Continuous-time Algebraic Riccati Equation* (CARE):

$$F^T S + SF - SGS + H = 0, \tag{6}$$

where all matrices are $n \times n$ and G, H are symmetric and S denotes a symmetric unknown matrix. Let B denote the factor in the decomposition $G = BB^T$ such that the rank of B equals the rank of G and C denote the factor in the decomposition $H = CC^T$ such that the rank of C equals the rank of H. Recall that:

- If the pair (F, B) is *stabilizable*, then there exists a matrix D such that all the eigenvalues of $F + BD$ have only negative real parts.
- If the pair (C, F) is *detectable*, then there exists a matrix E such that all the eigenvalues of $C + EF$ have only negative real parts.

Under the condition that (F, B) is a stabilizable pair and (C, F) is a detectable pair, the CARE has a unique positive-semidefinite solution [30]. Algorithmic details about the solution of the CARE are available in [5].

Denote the sample matrices to average as $X_k \in \mathrm{St}(p, n)$, with $k \in \{1, \ldots, N\}$, and assume that the samples X_k are distributed in a neighbourhood of a center of mass $C \in \mathrm{St}(p, n)$. In the present manuscript, it is assumed that $p > n$ strictly, as the case $p = n$ leads to averaging over the orthogonal group $\mathrm{O}(p)$, which may be given the structure of a Lie group, hence such a case may be treated by the method proposed in [21].

A. Fixed-point arithmetic averaging algorithms on $\mathrm{St}(p, n)$

The following steps lead to an equation characterizing the empirical mean matrix $X \in \mathrm{St}(p, n)$, that represents an estimate of the actual center of mass $C \in \mathrm{St}(p, n)$ on the basis of the available information:

1) Map the points $X_k \in \mathrm{St}(p, n)$ belonging to a neighbourhood of the sought-for mean-matrix $X \in \mathrm{St}(p, n)$ onto $T_X\mathrm{St}(p, n)$ by applying a lifting map. Denote such points as $V_k \stackrel{\text{def}}{=} P_X^{-1}(X_k)$.
2) Compute the linear combination $\overline{V} = \alpha \sum_{k=1}^N V_k$, with $\alpha > 0$. If $\alpha = \frac{1}{N}$, then the vector $\overline{V}$ coincides with the arithmetic mean of the vectors V_k.
3) Bring back the mean vector $\overline{V}$ to $\mathrm{St}(p, n)$ by the retraction $P_X(\overline{V})$ and get an empirical mean matrix $X = P_X(\overline{V})$.

Such a procedure is illustrated in Figure 3. Summarizing the above procedure, a mean matrix $X \in \mathrm{St}(p, n)$ is the solution of the non-linear, matrix-type equation:

$$X = P_X\left(\alpha \sum_{k=1}^N P_X^{-1}(X_k)\right) \tag{7}$$

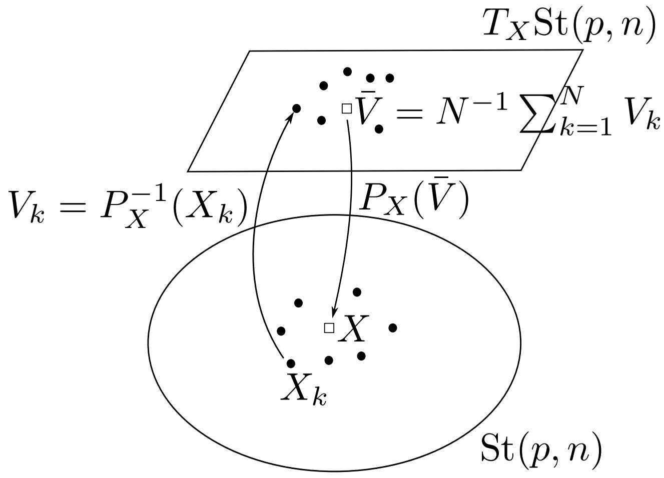

Fig. 3. Computation of an average matrix by a Stiefel-manifold retraction. The dots (•) denote sample matrices to average and the box symbol (□) denotes their empirical mean-Stiefel-matrix.

in the variable X. In general, however, the equation (7) cannot be solved in closed form. It may be solved by means of a fixed-point iteration algorithm, that generates a sequence $X^{(i)} \in \mathrm{St}(p, n)$ of estimates of the sought-for empirical mean matrix X, and that may be written as:

$$X^{(i+1)} = P_{X^{(i)}}\left(\alpha \sum_{k=1}^N P_{X^{(i)}}^{-1}(X_k)\right), \; i \geq 0, \tag{8}$$

where matrix $X^{(0)} \in \mathrm{St}(p, n)$ denotes an initial guess. When $\alpha = \frac{1}{N}$, the fixed point algorithm is a direct extension of the iteration rule proposed in the paper [21].

The following subsections investigate three instances of retraction/lifting maps and associated averaging algorithms.

1) QR-decomposition-type retraction map and its associated lifting map: In [1], one of the retractions P_X that map a tangent vector of $T_X\mathrm{St}(p, n)$ onto $\mathrm{St}(p, n)$ is given by:

$$P_X(V) \stackrel{\text{def}}{=} \mathrm{qf}(X + V), \tag{9}$$

where the symbol $\mathrm{qf}(\cdot)$ denotes the Q-factor of the thin QR decomposition of its $\mathbb{R}^{p\times n}$ matrix argument and the R-factor is a upper-triangular matrix with strictly positive elements on its main diagonal, so that the decomposition is unique.

In the present paper, it is proposed a way to calculate the lifting map associated to the above QR-decomposition-based retraction map. Given matrices $X, Q \in \mathrm{St}(p, n)$, if there exists an upper-triangular matrix R with strictly positive elements on its main diagonal such that $QR - X \in T_X\mathrm{St}(p, n)$, then the lifting map P_X^{-1} can be represented by:

$$P_X^{-1}(Q) = QR - X. \tag{10}$$

The $n \times n$ matrix R must satisfy the condition:

$$X^T(QR - X) + (QR - X)^T X = 0. \tag{11}$$

Namely, the matrix R may be calculated by solving the linear system of $\frac{n(n+1)}{2}$ independent equations:

$$MR + R^T M^T = 2I_n, \tag{12}$$

where $M \stackrel{\text{def}}{=} X^T Q$ is known. Note that when $X = Q$, it holds that $M = I_n$, hence the equation (12) reduces to $R + R^T = 2I_n$. As the matrix R is upper-triangular, the last condition implies that $R = I_n$. As the retraction $P_X^{-1}(Q)$ exists for $Q = X$ and as any entry r_{ij} computes as a rational function of the elements m_{ij}, by continuity it must exist in a neighbourhood of matrix X.

The $(i, j)^{\text{th}}$ entry of the linear system (12) reads:

$$\sum_{k=1}^{j} m_{ik} r_{kj} + \sum_{k=1}^{i} m_{jk} r_{ki} = 2\delta_{ij}, \tag{13}$$

where the symbol δ_{ij} represents the 'Kronecker delta'. It is immediate to verify that exchanging the index i with the index j in the above equation yields the same equation, hence, the values of the indices may be restricted to $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, i$. Define:

- The matrix $\tilde{M}_i$ as the i^{th} principal minor extracted from the matrix M, namely: $\tilde{M}_i \stackrel{\text{def}}{=} \begin{bmatrix} m_{11} & \cdots & m_{1i} \\ \vdots & \ddots & \vdots \\ m_{i1} & \cdots & m_{ii} \end{bmatrix}$.
- The vector $\tilde{r}_i$ as the column-vector formed by the first i elements of the i^{th} column of the matrix R, namely: $\tilde{r}_i \stackrel{\text{def}}{=} [r_{1i} \cdots r_{ii}]^T$.
- The vector b_i as the column vector whose j^{th} element equals $-[m_{i1}\ m_{i2}\ \cdots\ m_{ij}]\tilde{r}_j$, for $j = 1, 2, \ldots, i-1$, while its i^{th} entry equals 1.

Then, the equations (12) may be rearranged as follows: For any value of the index i ranging from 1 to n, the unknown vector $\tilde{r}_i$ may be found by solving the linear system:

$$\tilde{M}_i \tilde{r}_i = b_i. \tag{14}$$

Consistency of the solution requires not only that $\det(\tilde{M}_i) \neq 0$, but also that $r_{ii} > 0$. Note that, for a given value of the index i, the right-hand term b_i depends only on the vectors $\tilde{r}_j$ with $j < i$. The procedure to calculate the thin-QR-decomposition-based lifting $P_X^{-1}(Q)$ is outlined in the Algorithm 1. If the procedure stops before its natural end, then the matrix Q does not belong to the domain of definition of the lifting map P_X^{-1}.

An alternative way to compute the solution of the equation (12) would be to recast it as a linear system of the form $A\text{vec}(R) = 2\text{vec}(I_n)$, where the matrix A is written in terms of Kronecker products of appropriate matrices. Such a method was employed to conduct the numerical tests presented in the conference paper [27], [28] but it was found to be numerically far more expensive than the one presented in the Algorithm 1 and hence soon abandoned in favour of the latter one.

On the basis of the thin-QR-type retraction map and of its associated lifting map, the fixed-point iteration algorithm (8) may be particularized to:

$$X^{(i+1)} = \text{qf}\left(\alpha \sum_{k=1}^{N} X_k R_k(X^{(i)}) + (1 - N\alpha) X^{(i)}\right),\ i \geq 0, \tag{15}$$

where matrix $X^{(0)} \in \text{St}(p, n)$ denotes an initial guess and the notation $R_k(X^{(i)})$ emphasizes the fact that the upper-triangular matrix R_k depends on the current estimate $X^{(i)}$ via the condition (11).

Algorithm 1 Procedure to calculate the thin-QR-decomposition-based lifting map P_X^{-1}.

Given matrices X and Q, compute $M = X^T Q$
if $m_{11} > 0$ **then**
  Set $\tilde{r}_1 = \frac{1}{m_{11}}$
else
  Stop
end if
Set $i = 2$
repeat
  Set $\tilde{M}_i$ to the i^{th} principal minor extracted from M
  if $\det(\tilde{M}_i) \neq 0$ **then**
    Set b_i as the column-vector whose j^{th} entry equals $-[m_{i1}\ m_{i2}\ \cdots\ m_{ij}]\tilde{r}_j$, for $j = 1, 2, \ldots, i-1$, while its i^{th} entry equals 1
    Compute $\tilde{r}_i = \tilde{M}_i^{-1} b_i$
  else
    Stop
  end if
  if $(\tilde{r}_i)_i \leq 0$ **then**
    Stop
  end if
  Set $i = i + 1$
until $i > n$
Construct the matrix R from the vectors $\tilde{r}_i$
Compute the lifting map $P_X^{-1}(Q)$ as $QR - X$

2) Polar-decomposition-based retraction map and its associated lifting map: Given a real-valued $p \times n$ matrix A, its polar decomposition is written as $A = QS$, where Q is a matrix in $\text{St}(p, n)$ termed *polar factor* of A, hereafter denoted by $Q = \text{pf}(A)$, and S is a symmetric positive-semidefinite $n \times n$ matrix [26]. The polar decomposition of a matrix always exists and if the matrix is full rank, then its polar factor is unique. Given a point $X \in \text{St}(p, n)$ and a vector $V \in T_X\text{St}(p, n)$, the polar-decomposition retraction on the Stiefel manifold may be written as [1]:

$$P_X(V) \stackrel{\text{def}}{=} \text{pf}(X + V). \tag{16}$$

Under appropriate conditions, the polar-decomposition retraction may be written in closed form. In fact, write $X + V = QS$. From the conditions $Q^T Q = I_n$ and $S^T = S$, it follows that:

$$(X + V)^T (X + V) = S^T Q^T Q S \Rightarrow$$
$$X^T X + X^T V + V^T X + V^T V = S^2.$$

Now, it holds that $X^T X = I_n$ and $X^T V + V^T X = 0$, and it may be readily verified that the matrix $I_n + V^T V$ is positive-definite, hence $S = (I_n + V^T V)^{\frac{1}{2}}$. From the equality $X + V = QS$, the following closed-form expression for the polar-decomposition-based retraction is obtained:

$$P_X(V) = (X + V)(I_n + V^T V)^{-\frac{1}{2}}. \tag{17}$$

A possible lifting map associated to the polar-decomposition-based-retraction (16) is proposed in the present paper on the basis of the following considerations.

Algorithm 2 Procedure to calculate the polar-decomposition-based lifting map P_X^{-1}.

Given matrices X and Q, compute $M = X^T Q$
Solve the CARE $(-M)S + S(-M^T) + 2I_n = 0$ for S
Compute the lifting map $P_X^{-1}(Q)$ as $QS - X$

Given matrices $Q, X \in \mathrm{St}(p,n)$, the lifting map associated to the retraction map (16) may be computed as:

$$P_X^{-1}(Q) = QS - X, \tag{18}$$

provided that there exists a symmetric positive-semidefinite $n \times n$ matrix S such that $QS - X \in T_X\mathrm{St}(p,n)$. The tangency condition reads $(QS - X)^T X + X^T(QS - X) = 0$. Setting $M \stackrel{\mathrm{def}}{=} X^T Q$, the above condition becomes:

$$(-M)S + S(-M^T) + 2I_n = 0. \tag{19}$$

When $X = Q$, it holds that $M = I_n$, hence the equation (19) reduces to $2S = 2I_n$. As the retraction $P_X^{-1}(Q)$ exists for $Q = X$ and as any entry s_{ij} computes as a rational function of the elements m_{ij}, by continuity it must exist in a neighbourhood of X. The equation (19) is linear in the unknown S and represents a special case of the CARE (6).

The procedure to calculate the polar-decomposition-based lifting $P_X^{-1}(Q)$ is outlined in the Algorithm 2. If the CARE does not admit any solution, then the matrix Q does not belong to the domain of definition of the lifting map P_X^{-1}.

On the basis of the polar-decomposition-based retraction map and of its associated lifting map, the fixed-point iteration algorithm (8) may be particularized to:

$$X^{(i+1)} = \mathrm{pf}\left(\alpha \sum_{k=1}^{N} X_k S_k(X^{(i)}) + (1 - N\alpha)X^{(i)}\right),\ i \geq 0, \tag{20}$$

where matrix $X^{(0)} \in \mathrm{St}(p,n)$ denotes an initial guess and the notation $S_k(X^{(i)})$ emphasizes the fact that the symmetric positive-semidefinite matrix S_k depends on the current estimate $X^{(i)}$.

An alternative solution to the problem of computing the lifting map associated to the polar-decomposition-based retraction, that would avoid solving a CARE sub-problem, is suggested by the observation that an equation of the form $ESF^T + FSE^T = G = G^T$ has always a symmetric solution and may be recast as $(E \otimes F + F \otimes E)\mathrm{vec}(S) = \mathrm{vec}(G)$, where symbol $\otimes$ denotes Kronecker product, as recalled, e.g., in the paper [36]. However, it is immediate to verify that such a solution is extremely expensive from a computational point of view and we are not about to pursue it.

3) Orthographic retraction map and its associated lifting map: The paper [2] studies *orthographic retractions* on sub-manifolds of Euclidean spaces. In the paper [2] it is proven that, given a pair $(X, V) \in T\mathrm{St}(p,n)$, if V is sufficiently close to $0 \in T_X\mathrm{St}(p,n)$, then there exists a normal vector $Z \in N_X\mathrm{St}(p,n)$ such that:

$$P_X(V) \stackrel{\mathrm{def}}{=} X + V + Z \tag{21}$$

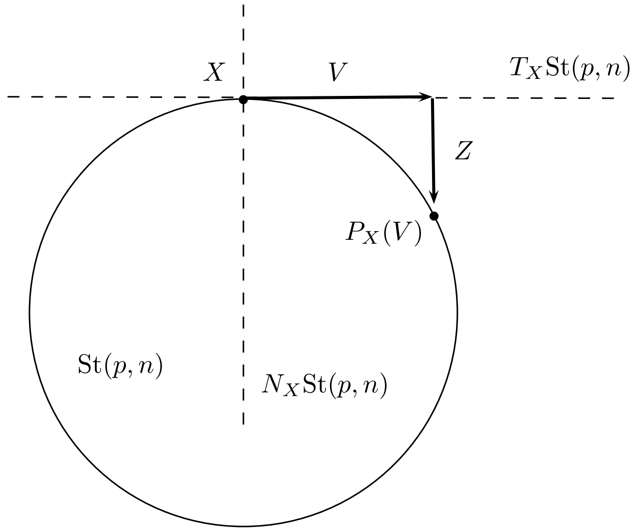

Fig. 4. Illustration of the notion of orthographic retraction on the manifold $\mathrm{St}(p,n)$ at a point $X \in \mathrm{St}(p,n)$, with $V \in T_X\mathrm{St}(p,n)$ and $Z \in N_X\mathrm{St}(p,n)$.

Algorithm 3 Procedure to calculate the orthographic retraction $P_X(V)$.

Given matrices X and V, compute $M = X^T V + I_n$
Solve the CARE $-S^2 - SM - M^T S - V^T V = 0$ for S
Compute the retraction map $P_X(V)$ as $X(I_n + S) + V$

is a retraction on $\mathrm{St}(p,n)$. Note that the quantities X, V, Z are regarded as elements of the Euclidean space $\mathbb{R}^{p\times n}$, hence their addition makes sense. Such a retraction is illustrated in Figure 4. From the figure, it is clear that if the tangent vector V is too large, there might not exist any normal vector Z such that the sum $X + V + Z \in \mathrm{St}(p,n)$. The orthographic retraction map on the Stiefel manifold $\mathrm{St}(p,n)$ reads:

$$P_X(V) = X + V + XS, \tag{22}$$

provided that there exists a $n \times n$ symmetric matrix S such that $V + X(I_n + S) \in \mathrm{St}(p,n)$, namely, such that:

$$(X + V + XS)^T(X + V + XS) = I_n. \tag{23}$$

The above equation in the unknown matrix S may be written in plain form as:

$$-S^2 - S(X^T V + I_n) - (V^T X + I_n)S - V^T V = 0. \tag{24}$$

The equation (24) represents an instance of the Continuous-time Algebraic Riccati Equation (6), where $F = -X^T V - I_n$, $G = I_n$ and $H = -V^T V$. The orthographic retraction map (22) may be computed numerically as shown in the Algorithm 3. It is worth noting that the explicit expression of the orthographic retraction map (22) and of the CARE (19) in terms of the vector V were not given in paper [2].

In the present paper, it is proposed a possible lifting map associated to the orthographic retraction map. It reads:

$$P_X^{-1}(Q) = Q - X - XS, \tag{25}$$

for $Q \in \mathrm{St}(p,n)$ (again thought of as an element of the Euclidean space $\mathbb{R}^{p\times n}$), which is well-defined provided that there exists a symmetric $n\times n$ matrix S such that $Q - X(I_n + S) \in T_X\mathrm{St}(p,n)$. The tangency condition of the lifting map reads, explicitly:

$$X^T(Q - X - XS) + (Q - X - XS)^T X = 0. \quad (26)$$

The above equation in the unknown matrix S is linear and admits the explicit solution:

$$S = \frac{1}{2}(Q^T X + X^T Q) - I_n. \quad (27)$$

Hence, the orthographic lifting map may be written in closed form as:

$$P_X^{-1}(Q) = Q - \frac{1}{2}X(Q^T X + X^T Q). \quad (28)$$

It is worth remarking that the orthographic lifting map is linear in its argument. The orthographic lifting map P_X^{-1} may be derived by applying a projection operator $\pi_X : \mathbb{R}^{p\times n} \to T_X\mathrm{St}(p,n)$ into the tangent space $T_X\mathrm{St}(p,n)$ corresponding to the Euclidean metric in $\mathbb{R}^{p\times n}$ [22] to the quantity $Q - X$. namely, $P_X^{-1}(Q) = \pi_X(Q - X)$.

The fixed-point averaging algorithm corresponding to the orthographic retraction/lifting pair reads:

$$\begin{cases} A \stackrel{\mathrm{def}}{=} \alpha \sum_k X_k, \\ V^{(i)} = A - \frac{1}{2}X^{(i)}A^T X^{(i)} - \frac{1}{2}X^{(i)}(X^{(i)})^T A, \\ X^{(i+1)} = X^{(i)}(I_n + S(X^{(i)}, V^{(i)})) + V^{(i)}, \end{cases} \quad (29)$$

for $i \geq 0$, where the notation $S(X^{(i)}, V^{(i)})$ emphasizes the fact that the symmetric factor S depends on the current iterate $X^{(i)}$ and the tangent vector $V^{(i)}$.

B. Fixed-point arithmetic averaging algorithms on $\mathrm{St}(p,n)$ via a Lie-group action

The special orthogonal group $\mathrm{SO}(p)$ *acts* on the Stiefel manifold $\mathrm{St}(p,n)$ via pre-multiplication. Namely, if $G \in \mathrm{SO}(p)$ and $X \in \mathrm{St}(p,n)$, then $GX \in \mathrm{St}(p,n)$. The pre-multiplication-based action may be exploited to design a retraction map for the Stiefel manifold, as it was suggested in [11] and subsequently applied to optimization problems on the Stiefel manifold in [12]. Such a result may be summarized as follows. Define:

- A coordinate map $\psi : \mathfrak{so}(p) \to \mathrm{SO}(p)$ such that $\psi(0) = I_p$.
- A function $\rho_X : \mathfrak{so}(p) \to T_X\mathrm{St}(p,n)$, defined by $\rho_X(U) \stackrel{\mathrm{def}}{=} \frac{d}{dt}(tU)X\big|_{t=0}$.
- A linear map $a_X : T_X\mathrm{St}(p,n) \to \mathfrak{so}(p)$ such that $\rho_X(a_X(V)) = V$ for any $V \in T_X\mathrm{St}(p,n)$.

Then, a retraction P_X which is a map from $T_X\mathrm{St}(p,n)$ to $\mathrm{St}(p,n)$ is given by:

$$P_X(V) = \psi(a_X(V))X = \psi(\Omega)X, \ \Omega \stackrel{\mathrm{def}}{=} a_X(V) \in \mathfrak{so}(p). \quad (30)$$

A key observation is that, in the present context, it is not necessary to utilize the full retraction P_X as it is sufficient to use the coordinate map combined with the special-orthogonal-group action and its inverse to parametrize Stiefel manifold matrices and tangent vectors to the Stiefel manifold.

In the following subsections, we introduce the notions of *pseudo-retraction* and *pseudo-lifting* maps and investigate two instances of such maps.

1) Cayley-type retraction/lifting pair: A difficulty related to the retraction (30) in the present context is that there are no known results about its inversion. A result published in the paper [23] that partially overcomes such a difficulty concerns the specific case that the retraction (30) is used to parametrize matrices in a neighbourhood of the matrix $[I_n \ 0]^T \in \mathrm{St}(p,n)$ and that the coordinate map ψ is chosen as the Cayley map $\mathrm{Cay} : \mathfrak{so}(p) \to \mathrm{SO}(p)$ defined by:

$$\mathrm{Cay}(\Omega) \stackrel{\mathrm{def}}{=} (I_p + \Omega)(I_p - \Omega)^{-1} = (I_p - \Omega)^{-1}(I_p + \Omega). \quad (31)$$

The fundamental limitation that the retraction be used in a neighbourhood of the matrix $[I_n \ 0]^T \in \mathrm{St}(p,n)$ is due to the fact that the domain of the coordinate map ψ is restricted in [23] to the set $\mathcal{W}$ of the skew-symmetric matrices with block structure:

$$\Omega = \begin{bmatrix} A & -B^T \\ B & 0 \end{bmatrix}, \quad (32)$$

where A is a $n \times n$ skew-symmetric matrix and B is an arbitrary $(p-n)\times n$ matrix, where the dimension of the space $\mathcal{W}$ is calculated as $\dim \mathcal{W} = pn - \frac{n(n+1)}{2}$. The paper [23] makes, thus, use of a *pseudo-retraction map* $\hat{P}_X : \mathcal{W} \to \mathrm{St}(p,n)$ defined by:

$$\hat{P}_X(\Omega) \stackrel{\mathrm{def}}{=} \mathrm{Cay}(\Omega)X. \quad (33)$$

The associated *pseudo-lifting map* $\hat{P}_X^{-1}$ returns a skew-symmetric matrix of the type (32), with:

$$\begin{aligned} A &= 2(X_u^T + Q_u^T)^{-1}\mathrm{sk}(Q_u^T X_u + X_l^T Q_l)(X_u + Q_u)^{-1}, \\ B &= (Q_l - X_l)(X_u + Q_u)^{-1}, \end{aligned}$$

where $\mathrm{sk}(M) \stackrel{\mathrm{def}}{=} \frac{1}{2}(M^T - M)$ for an arbitrary square matrix M and the following block-partitions were made use of: $X = \begin{bmatrix} X_u \\ X_l \end{bmatrix} \in \mathrm{St}(p,n)$ and $Q = \begin{bmatrix} Q_u \\ Q_l \end{bmatrix} \in \mathrm{St}(p,n)$, with $X_u, Q_u \in \mathbb{R}^{n\times n}$ and $X_l, Q_l \in \mathbb{R}^{(p-n)\times n}$, provided that the matrix $X_u + Q_u$ be nonsingular. Note that, when $X = Q$, it holds that $Q_u^T X_u + X_l^T Q_l = 2I_n$, whose skew-symmetric part is zero, hence, it holds that $A = 0$ and $B = 0$. As the retraction $P_X^{-1}(Q)$ exists for $Q = X$ and as any entry of matrices A and B computes as a rational function of the entries of matrices X_u, Q_u, X_l, Q_l, by continuity it must exist in a neighbourhood of the matrix X.

2) Proposed full-Cayley retraction/lifting pair: In the present paper, it is proposed that a full parametrization of the algebra $\mathfrak{so}(p)$ be used instead of the block structure (32). In this case, the retraction map takes again the form $\hat{P}_X(\Omega) = \mathrm{Cay}(\Omega)X$ but now $\Omega \in \mathfrak{so}(p)$. The calculation of the pseudo-lifting map $\hat{P}_X^{-1}$ implies the solution of the equation $Q = \hat{P}_X(\Omega)$ for $\Omega \in \mathfrak{so}(p)$ for $Q, X \in \mathrm{St}(p,n)$ given, or, equivalently, of the equation $(I_p - \Omega)^{-1}(I_p + \Omega)X = Q$. Rearranging terms, the latter equation may be written as:

$$\Omega(Q + X) = Q - X. \quad (34)$$

There appear to be no known closed-form solutions of equations of the type (34). It is to be noted that the equation (34) does not possess a unique solution, in general. A closed-form solution to the problem (34) was found in the special case that the dimension index n is an even number. Recall the following preliminary facts about skew-symmetric matrices:

- Any skew-symmetric matrix of odd size is not invertible (as a consequence of Jacobi's Theorem [17]).
- If a skew-symmetric matrix Ω is invertible, its inverse is also skew-symmetric, in fact $(\Omega^{-1})^T = (\Omega^T)^{-1} = (-\Omega)^{-1} = -\Omega^{-1}$.

In the case that a Stiefel manifold $\mathrm{St}(p,n)$ with n being an even number is considered, a possible closed-form solution to the problem (34) may be found as follows. The sought-for skew symmetric matrix may be written by the *ansatz*:

$$\Omega = Z(Q-X)^T, \tag{35}$$

with Z of size $p \times n$ unknown. Substituting equation (35) in equation (34) yields $Q - X = Z(Q-X)^T(Q+X) = Z(Q^TX - X^TQ)$. If the $n \times n$ matrix $\mathrm{sk}(X^TQ)$ is nonsingular, then $Z = \frac{1}{2}(Q-X)\mathrm{sk}^{-1}(X^TQ)$. Replacing such an expression for Z in equation (35) gives the final result:

$$\hat{P}_X^{-1}(Q) = \frac{1}{2}(Q-X)\mathrm{sk}^{-1}(X^TQ)(Q-X)^T. \tag{36}$$

It is straightforward to check that $(Q-X)\mathrm{sk}^{-1}(X^TQ)(Q-X)^T$ is skew-symmetric and that $P_X(P_X^{-1}(Q)) = Q$. The above setting may be referred to as *full-Cayley* pseudo-retraction/lifting case.

3) Averaging algorithm based on the Cayley map: The averaging algorithm related to the use of a Cayley-map-based retraction, and of its associated lifting map, may be outlined as follows:

1) Compute N matrices Ω_k by means of the pseudo-lifting map $\hat{P}_X^{-1}(X_k)$.
2) Compute a linear combination of the obtained skew-symmetric matrices $\overline{\Omega} = \alpha \sum_{k=1}^N \Omega_k$, with $\alpha > 0$.
3) The empirical mean matrix must then satisfy the condition $X = \hat{P}_X(\overline{\Omega})$.

Note that the matrices Ω_k, and hence their linear combination $\overline{\Omega}$, depend on the matrix X. The above procedure is illustrated in Figure 5. The empirical mean matrix X is, thus, the solution of the non-linear, matrix-type equation:

$$X = \left(I_p + \alpha \sum_{k=1}^N \hat{P}_X^{-1}(X_k)\right)\left(I_p - \alpha \sum_{k=1}^N \hat{P}_X^{-1}(X_k)\right)^{-1} X. \tag{37}$$

Note that a matrix X can be a solution of the equation (37) only if there exists a matrix $\Omega \in \mathcal{W}$ or $\mathfrak{so}(p)$ such that $X = (I_p - \alpha\Omega)^{-1}(I_p + \alpha\Omega)X$, that is equivalent to $\Omega X = 0$.

In general, the equation (37) cannot be solved in closed form, hence, it is necessary to resort to an iterative algorithm to seek for its solution. In particular, in the present manuscript, a sequence $X^{(i)} \in \mathrm{St}(p,n)$ of increasingly refined estimates of the empirical mean matrix of a given set of sample is sought for via a fixed-point algorithm with initial guess $X^{(0)} \in \mathrm{St}(p,n)$:

$$X^{(i+1)} = \mathrm{Cay}\left(\alpha \sum_{k=1}^N \hat{P}_{X^{(i)}}^{-1}(X_k)\right) X^{(i)}. \tag{38}$$

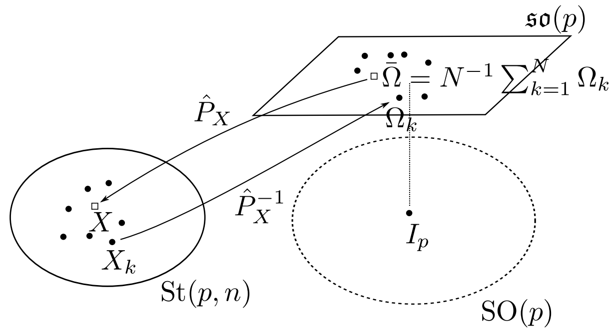

Fig. 5. Averaging over the Stiefel manifold via special-orthogonal-group action and the Cayley map. The dots (•) denote sample matrices and the box symbol (□) denotes their empirical average matrix. Symbol $\hat{P}_X$ denotes a pseudo-retraction map while symbol $\hat{P}_X^{-1}$ denotes its associated pseudo-lifting map.

Algorithm 4 Fixed-point averaging algorithm resulting from the pseudo-retraction map introduced in [23].

Input matrices $X_k \in \mathrm{St}(p,n)$, $k = 1, \ldots, N$ and $X^{(0)} \in \mathrm{St}(p,n)$ and number of iterations I
for $i = 0$ to I **do**
  Define the block-partition $X^{(i)} = \begin{bmatrix} X_u^{(i)} \\ X_l^{(i)} \end{bmatrix}$ with $X_u^{(i)} \in \mathbb{R}^{n\times n}$, $X_l^{(i)} \in \mathbb{R}^{(p-n)\times n}$
  for $k = 1$ to N **do**
    Define the block-partition $X_k = \begin{bmatrix} X_{u,k} \\ X_{l,k} \end{bmatrix}$ with $X_{u,k} \in \mathbb{R}^{n\times n}$, $X_{l,k} \in \mathbb{R}^{(p-n)\times n}$
    if $\det(X_u^{(i)} + X_{u,k}) = 0$ **then**
      Stop
    end if
    Compute matrix $L_k^{(i)} = (X_u^{(i)} + X_{u,k})^{-1}$
    Compute matrix $A_k^{(i)} = 2(L_k^{(i)})^T \mathrm{sk}(X_{u,k}^T X_u^{(i)} + (X_l^{(i)})^T X_{l,k}) L_k^{(i)}$
    Compute matrix $B_k^{(i)} = (X_{l,k} - X_l^{(i)}) L_k^{(i)}$
    Construct matrix $\Omega_k^{(i)} = \begin{bmatrix} A_k^{(i)} & -(B_k^{(i)})^T \\ B_k^{(i)} & 0 \end{bmatrix}$
  end for
  Compute matrix $\Gamma^{(i)} = \sum_{k=1}^N \Omega_k^{(i)}$
  Update $X^{(i+1)} = \left(I_p + \alpha\Gamma^{(i)}\right)\left(I_p - \alpha\Gamma^{(i)}\right)^{-1} X^{(i)}$
end for

The fixed-point algorithm resulting from the pseudo-retraction map introduced in the paper [23] is illustrated in the Algorithm 4, while the iterative algorithm resulting from the application of the full-Cayley setting is explained in the Algorithm 5. For the Algorithm 4, it is assumed that all the matrices X_k and $X^{(0)}$ lay in a sufficiently narrow neighbourhood of the point $[I_n\ 0]^T$, while such a limitation is not present in the Algorithm 5.

Algorithm 5 Fixed-point averaging algorithm on $\mathrm{St}(p,n)$, with n even, resulting from the full-Cayley pseudo-retraction map.

Input matrices $X_k \in \mathrm{St}(p,n)$, $k = 1,\ldots,N$ and $X^{(0)} \in \mathrm{St}(p,n)$ and number of iterations I
for $i = 0$ to I **do**
  for $k = 1$ to N **do**
    if $\det(\mathrm{sk}((X^{(i)})^T X_k)) = 0$ **then**
      Stop
    end if
    Compute $\Omega_k^{(i)} = \frac{1}{2}(X_k - X^{(i)})\mathrm{sk}^{-1}((X^{(i)})^T X_k)(X_k - X^{(i)})^T$
  end for
  Compute matrix $\Gamma^{(i)} = \sum_{k=1}^{N} \Omega_k^{(i)}$
  Update $X^{(i+1)} = \left(I_p + \alpha\Gamma^{(i)}\right)\left(I_p - \alpha\Gamma^{(i)}\right)^{-1} X^{(i)}$
end for

C. Relationships with other averaging methods

The fundamental equation (7) that defines an average matrix on the compact Stiefel manifold may be regarded as an extension of the Kolmogoroff-Nagumo averaging rule for real numbers (for a recent discussion see, e.g., [33]) and may be used to compute averages on other manifolds of interest. In particular, it may be applied to the manifold $\mathbb{R}^{p\times n}$. Given matrices $X, Q \in \mathbb{R}^{p\times n}$ and $V \in T_X\mathbb{R}^{p\times n} \cong \mathbb{R}^{p\times n}$, a retraction/lifting pair may be taken as $P_X(V) = X + V$ and $P_X^{-1}(Q) = Q - X$. Then, the fundamental equation would read:

$$X = X + \alpha \sum_{k=1}^{N} (X_k - X), \tag{39}$$

whose solution is the well-known Pythagoras' mean matrix X_{pa}:

$$X_{\mathrm{pa}} \stackrel{\mathrm{def}}{=} \frac{1}{N}\sum_{k=1}^{N} X_k. \tag{40}$$

It is worth noting that the solution does not depend on the value of the coefficient $\alpha \neq 0$. The fundamental equation (7) may be read as Pythagoras' mean rephrased in the language of manifolds.

Naive (or extrinsic) averaging consists in computing a linear combination of the available $\mathrm{St}(p,n)$ samples, thought of as elements of the linear matrix space $\mathbb{R}^{p\times n}$ followed by a projection of the result onto the manifold $\mathrm{St}(p,n)$. The linear combination of N samples $X_k \in \mathrm{St}(p,n)$ may be denoted as $\alpha \sum_k X_k$ (where $\alpha = 1/N$ would correspond to arithmetic average on the space $\mathbb{R}^{p\times n}$). A theoretical foundation of such an approach for another manifold of interest (namely, the manifold of symplectic matrices) was explained in the paper [20]. To what concerns the Stiefel manifold, it should be noted that the method described in the section II-B.3 essentially implements such a naive approach. From the equations (29), that describe in details the orthographic retraction/lifting pair-based averaging method, it is readily seen that the matrix A is computed as a linear combination of the available $\mathrm{St}(p,n)$-samples and that the method implements the required projection.

It is also worth discussing the relationship between the averaging method proposed in the present contribution with *Maximum Likelihood* estimation of the parameters of statistical distributions defined on the Stiefel manifold, as described in detail in the book [14]. The approach followed in the present paper is different from an approach based on maximum likelihood estimation. Maximum likelihood estimation is based on hypothesizing a probability model for the data, which includes some parameters, and in finding the optimal values of the parameters in a maximum-likelihood sense. In the present paper, no hypotheses are made about the distributional properties of the samples (except that the distribution is concentrated around a center of mass).

A well-known averaging theory is that of *Karcher mean* [29]. Karcher mean is based on the optimization on a criterion function, which in turn is written in terms of geodesic distance between the samples and the sought-for mean. Such a construction basically extends in a geometrically-sound way the definition of sample average over a flat space as the point that is as close as possible to the samples altogether. Most of the works on the subject available in the literature are concerned with the minimization of such a criterion function by a gradient-steepest-descent methods and on the individuation of the conditions under which such methods converge. Some recent works also consider alternative methods, such as Newton optimization, conjugate gradient or stochastic optimization. The method proposed in this paper does not rely on any criterion function and on any gradient-based-like optimization method but is completely different in nature. It is based on a characterization of the average in terms of arithmetic average on a tangent space, which leads in a natural way to a fixed-point implementation.

III. Numerical results

The present section illustrates the numerical behavior of the discussed retraction/lifting map-pairs in the context of averaging over the Stiefel manifold. The discussed retraction/lifting map-pairs are summarized in the Table I, along with their principal features.

In the numerical experiments, the center of the distribution $C \in \mathrm{St}(p,n)$ is generated by computing the Q-factor of a thin-QR decomposition of a matrix randomly generated in $\mathbb{R}^{p\times n}$ with normally-distributed entries. The N samples to average are generated by the rule

$$X_k = \exp(a\Omega_k)C, \tag{41}$$

with $\Omega_k \stackrel{\mathrm{def}}{=} \mathrm{sk}(A_k)$, with A_k being a matrix randomly generated in $\mathbb{R}^{p\times p}$ with normally-distributed entries, $a > 0$ controls the spread of the distribution around the center and $k = 1,\ldots,N$.

The initial guess $X^{(0)}$ may be chosen near one of the available samples. In the present simulations, the initial guess was chosen by slightly rotating the sample X_1 via a quasi-unit random rotation.

In order to inspect the behavior of the proposed algorithms, the following measure of discrepancy $\delta : \mathrm{St}(p,n)\times\mathrm{St}(p,n) \to$

TABLE I
RETRACTION/LIFTING MAP-PAIRS, DISCUSSED IN SECTION II, ALONG WITH THEIR PRINCIPAL FEATURES.

Retraction/lifting on $\mathrm{St}(p,n)$	**Features**
QR-decomposition	The lifting map may be computed efficiently by solving linear systems in an appropriate order.
Polar decomposition	The retraction map may be expressed in closed form. The lifting map may be computed by solving a CARE.
Orthographic	The retraction map may be computed by solving a CARE. The lifting map may be expressed in closed form. The resulting averaging algorithm is an instance of 'naive' averaging.
Cayley-type	Both the pseudo-retraction map and the pseudo-lifting map may be expressed in closed form. The resulting averaging algorithm is well-defined only in a neighborhood of the matrix $[I_n\ 0]^T$.
Full-Cayley-type	Both the pseudo-retraction map and the pseudo-lifting map may be expressed in closed form. The resulting averaging algorithm is well-defined only for n even.

$\mathbb{R}_0^+$ between two Stiefel-manifold matrices is made use of:

$$\delta(X,Y) \stackrel{\text{def}}{=} \|I_n - X^T Y\|_{\mathrm{F}}, \tag{42}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. Note that for every $X,Y \in \mathrm{St}(p,n)$, it holds that $\delta(X,X)=0$ and $\delta(X,Y)=\delta(Y,X)$. It may be used to measure the discrepancy between two successive steps of the algorithm, computed as $\delta(X^{(i)},X^{(i+1)})$, and the discrepancy between the current estimate and the actual center of the distribution, namely $\delta(X^{(i)},C)$.

A. Tests on learning stepsize

In order to determine a learning stepsize α for the learning algorithms, a number of numerical tests was performed. In particular, the following values were tested: $\frac{1}{N}$, $\frac{1}{2N}$ and $\frac{1}{4N}$. The result of extensive numerical simulations is that there are no real advantages in choosing a value different from the theoretically-optimal one $\alpha=\frac{1}{N}$. Such a value corresponds to arithmetic averaging over tangent spaces and makes the learning equations take the simplest form.

Although, in principle, variable stepsize schedules might be devised, in the present paper such a possibility was not exploited in order to keep the computational burden limited.

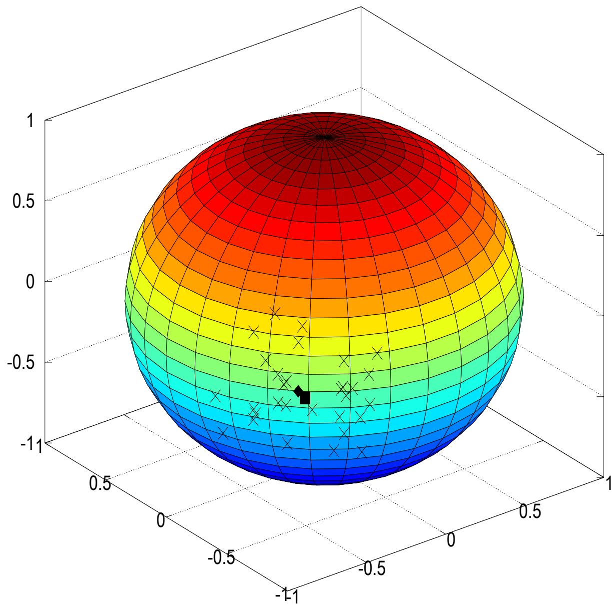

Fig. 6. Experiment about averaging on the sphere $\mathrm{St}(3,1)$. The samples to average are denoted by cross marks ($\times$), the actual center of the distribution is denoted by a diamond mark ($\diamond$) and the computed empirical mean is denoted by a box symbol ($\square$).

B. Single trials and comparisons

The Figure 6 is about averaging on the sphere $\mathrm{St}(3,1)$, for which is it possible to provide a graphical rendering of the result. Such a numerical simulation was performed with $N=30$ samples, generated with a spread value $a=0.3$, and was performed by using the QR-retraction-based averaging algorithm. The Figure 6 gives a quick picture of the meaning of the developed learning theory on curved manifolds.

The following experiment is about averaging over the manifold $\mathrm{St}(20,4)$. For this experiment, a number $N=30$ of samples were generated with a spread parameter $a=0.01$. The Figure 7 shows the values of the index $\delta(X^{(i)},C)$, while the Figure 8 shows the values of the index $\delta(X^{(i)},X^{(i+1)})$. The two pictures compare the behavior of the QR-decomposition-based-retraction algorithm, the Polar-decomposition-based-retraction averaging algorithm, the Cayley- and the full-Cayley-pseudo-retraction-based algorithms and the Orthographic-retraction-based algorithm. In the present experiment, all the algorithms behave satisfactorily and converge to solution-matrices with similar discrepancies with the actual center of the distribution. It is to be noted that the Cayley-pseudo-retraction-based algorithm is the slowest to converge, due to the limitations explained in the section II-B. The QR-decomposition, Orthographic and Polar-decomposition retraction/lifting based algorithms converge the fastest.

The Figure 9 shows a result of averaging real-world samples over the manifold $\mathrm{St}(5,2)$. The $N=50$ samples to average were obtained by running a fastICA algorithm [15], which separates 2 independent source signals out of 5 mixtures, on 50 independent trials on the same separation problem. The Figure 9 illustrates the obtained results, expressed in terms of separation performance index (PI) [15]. Again the QR-decomposition-based-retraction algorithm, the Polar-decomposition-based-retraction averaging algorithm, the Cayley- and the full-Cayley-pseudo-retraction algorithms and the Orthographic-retraction-based algorithm were tested. As they behave similarly about final performance after iteration, the average value of the 5 separation indices was retained as a good representative of their collective behavior. The figure

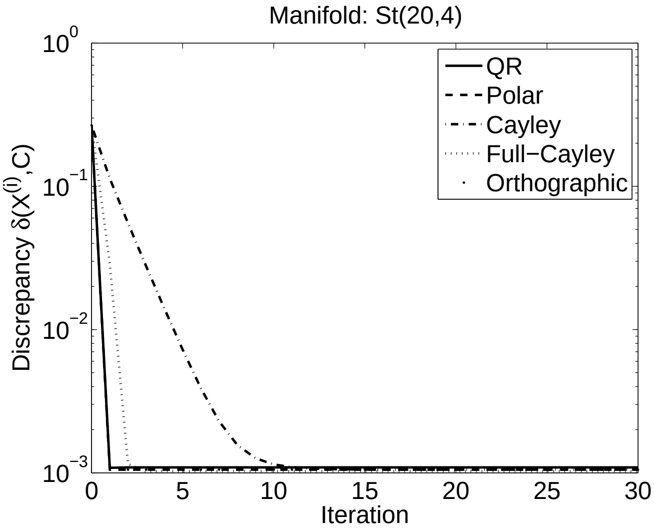

Fig. 7. Experiment about averaging on the manifold $\mathrm{St}(20, 4)$. Index $\delta(X^{(i)}, C)$ during iteration.

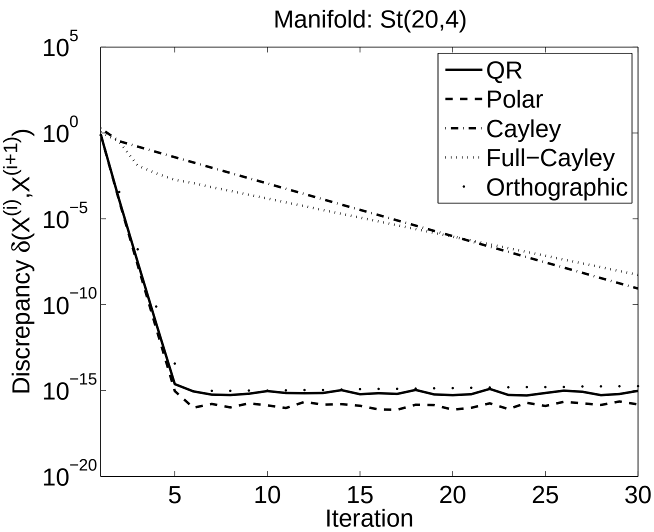

Fig. 8. Experiment about averaging on the manifold $\mathrm{St}(20, 4)$. Index $\delta(X^{(i)}, X^{(i+1)})$ during iteration.

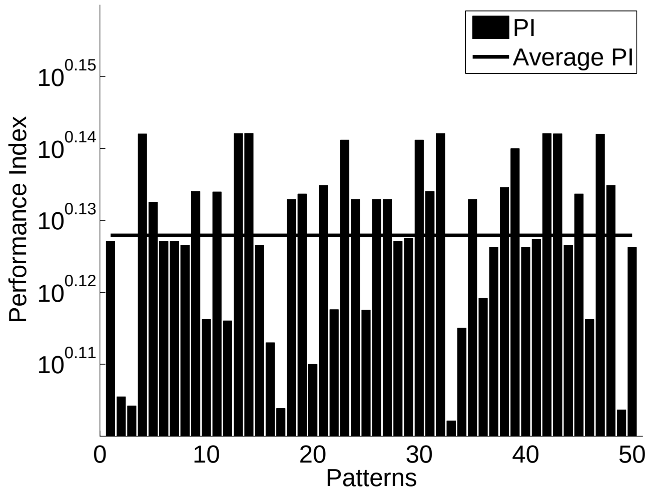

Fig. 9. Result of averaging over the manifold $\mathrm{St}(5, 2)$ on real-world fastICA samples. The bars show the values of the separation performance index (PI) pertaining to each sample, while the horizontal line indicates the average PI corresponding to the average separation matrix computed by the 5 tested algorithms on the same data-set.

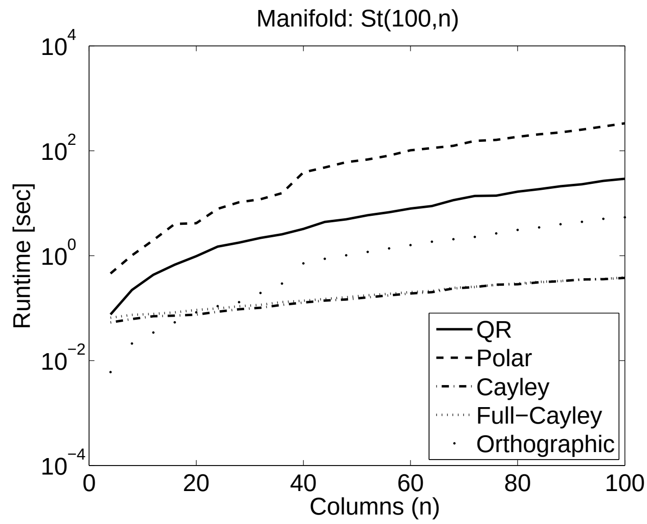

Fig. 10. Figure about runtimes for the manifold $\mathrm{St}(100, n)$ with n varying.

shows that the value of the PI corresponding to the empirical average matrix collocates in an average position with respect to the PI values of the single samples.

C. Computational-complexity evaluation

The Figure 10 shows the runtimes corresponding to the five tested algorithms run on the manifold $\mathrm{St}(100, n)$ with n varying. Such a numerical simulation was performed with $N = 50$ samples generated with a spread-parameter value $a = 0.01$. Each averaging experiment for each value of the index n was repeated 100 times to get rid of random fluctuations in the evaluation of runtimes. The obtained results allow the conclusion that for low-dimensional problems, i.e., for $n \leq 20$, the Orthographic retraction based method is the lightest in terms of computational burden, while for larger-sized problems, the Cayley and full-Cayley pseudo-retraction based algorithms are preferable.

D. Consistency of the estimation

In order to get some insights into the consistency of the devised averaging method, some numerical experiments were performed by varying the number of available samples and by varying the spread of the distribution of the samples to average around a given center of mass.

The Figure 11 shows the results of averaging obtained by varying the number of available samples, while the spread parameter was fixed to $a = 0.03$. The curves shown in the Figure 11 are the average result of 5 independent trials over the manifold $\mathrm{St}(20, 4)$ obtained by each method corresponding to the different retraction/lifting pairs discussed in the Section II.

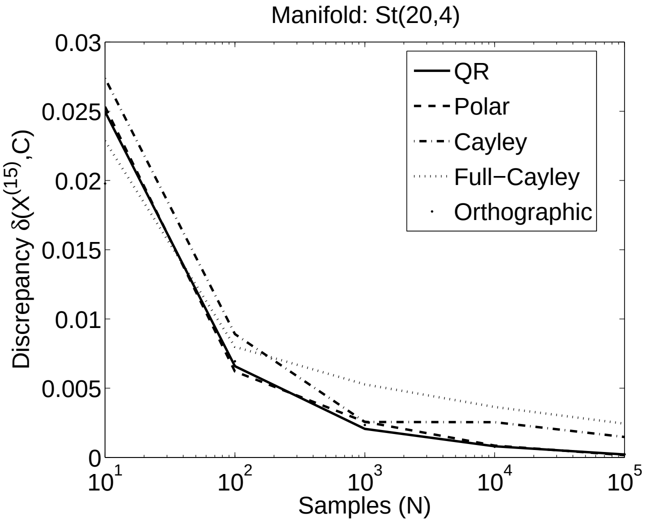

Fig. 11. Results of averaging obtained by varying the number N of available samples from 10 to 10^5 with log-step 10. The number of iterations for each algorithm was set to 15.

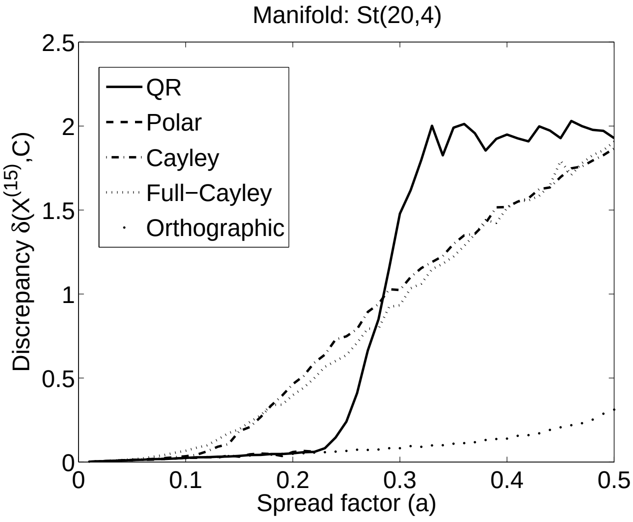

Fig. 12. Results of averaging obtained by varying the spread a of the distribution of the samples from 0.01 to 0.5 with step 0.01. The number of iterations for each algorithm was set to 15.

The obtained numerical results show that the discrepancy between the computed empirical arithmetic average and the actual center of the distribution decreases when the number of available samples increases, which confirms numerically the consistency of the proposed averaging method.

The Figure 12 shows the results of averaging obtained by varying the spread of the distribution of the samples, while the number of available samples was fixed to $N = 100$. The curves shown in the Figure 12 are the average result of 100 independent trials over the manifold $\mathrm{St}(20, 4)$ by each method corresponding to the retraction/lifting pairs discussed in the Section II. Note that the Polar-decomposition method and the QR-decomposition method (whose curves look superimposed in the Figure 12) eventually become unstable (around $a = 0.2$). The Orthographic-based method proves to be the most robust with respect to the spread of the distribution.

IV. Conclusions

The present research work extends the algorithm introduced in [21] to compute averages over Lie groups to the compact Stiefel manifold. The present method inherits the main advantage of the previous method, namely, it does not involve any metrics and is hence more general than the Riemannian mean method.

The idea underlying the developed algorithms is that points on the Stiefel manifold get mapped onto a specific tangent space, where the average is taken, and then the average point on the tangent space is brought back to the Stiefel manifold. The switching of points from/to the manifold to/from the tangent bundle is performed by the help of a retraction/lifting maps pair (or pseudo-retraction/pseudo-lifting maps pairs), customized to the case of the Stiefel manifold. In particular, four different retraction (or pseudo-retraction) maps were recalled from the literature, namely, the retraction map based on the thin-QR-decomposition, the retraction map based on the polar decomposition, the retraction map based on orthographic projection and the pseudo-retraction map based on the action of the special orthogonal group on the Stiefel manifold, when the Cayley map is used as a retraction on the special orthogonal group of matrices. The problem of calculating the associated lifting (or pseudo-lifting) maps was addressed both from a theoretical viewpoint and from an implementation viewpoint. The main aim of the submitted paper was to find ways to calculate the lifting maps associated to the mentioned retraction maps. Such calculations were the major source of difficulty and of necessary research work.

The Stiefel matrix calculated by means of the proposed method may be defined as 'average matrix' because it corresponds, via appropriate applications of retraction/lifting maps, to the Pythagoras' arithmetic average calculated over a flat space, which is a tangent space at a specific point to the Stiefel manifold. In other terms, the average over the tangent space corresponds to the average over the manifold upon non-linear transforms, which makes the found matrix be an average Stiefel matrix.

Numerical experimental results were shown and commented on in order to illustrate the numerical behavior of the proposed procedure. The obtained results confirm that the developed algorithms converge steadily and in a few iterations and that they are able to cope with relatively large-size problems with no significant numerical errors. In particular, a comparison based on convergence features and computational complexity figures allows concluding that the full-Cayley pseudo-retraction based algorithm and the Orthographic-projection-based method offer the best trade-off between convergence speed, computational burden and robustness.

Some topics related to the present research emerged that should be pursued in the future:

- Study analytically the convergence properties of the fixed-point averaging algorithm, possibly in conjunction with a variable stepsize schedule selection rule. Such an analysis could benefit of some recent works on the convergence

of fixed-point algorithms on certain manifolds as summarized, e.g., in the paper [32].
- Seek for a more general solution to the equation (34) to compute the lifting map associated to the full-Cayley-transform.
- Take into account the 'Mostow decomposition' [6] to design a further retraction/lifting pair for the Stiefel manifold.
- The proposed averaging method is based on arithmetic averaging a tangent space, which is a linear space. The proposed method may be generalized by invoking different kinds of averaging over a linear space on the basis of different distance measures (or, more generally, divergences) over linear spaces. A noteworthy class of divergences is given by Bregman theory, that was used to define averages over a hypersphere in [18] and used to define averages on subsets of $\mathbb{R}^n$ in [35].

Further efforts will be directed along the line of seeking more general schemes to compute averages over the Grassmann manifold and the 'flag manifold'.

ACKNOWLEDGMENT

The authors wish to gratefully thank the anonymous reviewers and the associate editor for their detailed comments and stimulating criticisms that helped improving the technical content and the clarity of the present manuscript.

This work was supported in part by the grant KAKENHI 23300069.

S. Fiori performed this work while at the Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology, supported by a fellowship of the Japan Society for the Promotion of Science (JSPS grant ID S11724).

REFERENCES

[1] P.-A. Absil, R. Mahony and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds.* New Jersey: Princeton University Press, 2008.

[2] P.-A. Absil and J. Malick, "Projection-like retractions on matrix manifolds," *SIAM Journal on Optimization*, Vol. 22. No. 1, pp. 135 – 158, 2012.

[3] M. Arnaudon, F. Barbaresco and L. Yang, "Medians and means in Riemannian geometry: Existence, uniqueness and computation," In *Matrix Information Geometries* (R. Bhatia and F. Nielsen, Eds.), Springer, 2012.

[4] M. Arnaudon, C. Dombry, A. Phan and L. Yang, "Stochastic algorithms for computing means of probability measures," *Stochastic Processes and their Applications*, Vol. 122, No. 4, pp. 1437 – 1455, April 2012.

[5] W.F. Arnold III and A.J. Laub, "Generalized eigenproblem algorithms and software for algebraic Riccati equations," *Proceedings of the IEEE*, Vol. 72, pp. 1746 – 1754, 1984.

[6] F. Barbaresco, "Robust statistical radar processing in Fréchet metric space: OS-HDR-CFAR and OS-STAP processing in Siegel homogeneous bounded domains," *Proceedings of the 2011 International Radar Symposium*, (IRS, September 7-9, 2011, Advanced Developments Department, Thales Air Systems, Limours, France), pp. 639 – 644, 2011.

[7] F. Barbaresco, "Information Geometry of Covariance Matrix: Cartan-Siegel Homogenous Bounded Domains, Mostow/Berger Fibration and Frechet Median," In *Matrix Information Geometries* (R. Bhatia and F. Nielsen, Eds.), Springer, 2012.

[8] B. Berkels, G. Linkmann and M. Rumpf, "An SL(2) invariant shape median," *Journal of Mathematical Imaging and Vision*, Vol. 37, No. 2, pp. 85 – 97, June 2010.

[9] 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.

[10] S. Bonnabel and R. Sepulchre, "Riemannian Metric and Geometric Mean for Positive Semidefinite Matrices of Fixed Rank," *SIAM Journal on Matrix Analysis and Applications*, Vol. 31, No. 3, pp. 1055 – 1070, August 2009.

[11] E. Celledoni and B. Owren, "On the implementation of Lie group methods on the Stiefel manifold," *Numerical Algorithms*, Vol. 32, pp. 163 – 183, 2003.

[12] E. Celledoni and S. Fiori, "Descent methods for optimization on homogeneous manifolds", *Journal of Mathematics and Computers in Simulation* (Special issue on "Structural Dynamical Systems: Computational Aspects", Guest Editors: N. Del Buono, L. Lopez and T. Politi), Vol. 79, No. 4, pp. 1298 – 1323, December 2008.

[13] H.E. Çetingül and R. Vidal, "Intrinsic mean shift for clustering on Stiefel and Grassmann manifolds," in Proceedings of the *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (June 20-25, 2009, Miami Beach, FL, USA), pp. 1896 – 1902, 2009.

[14] Y. Chikuse, *Statistics on Special Manifolds*, New York: Springer, 2003.

[15] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing.* Wiley, New York, 2002.

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

[17] H. Eves, *Elementary Matrix Theory.* Dover Publications, 1980.

[18] S. Fiori, "On vector averaging over the unit hyphersphere," Digital Signal Processing (Elsevier), Vol. 19, No. 4, pp. 715 – 725, July 2009.

[19] S. Fiori, "Averaging over the Lie group of optical systems transference matrices," *Frontiers of Electrical and Electronic Engineering in China* (Springer), Special issue of the Sino foreign-interchange Workshop on Intelligence Science and Intelligent Data Engineering 2010 - Part A, Vol. 6, No. 1, pp. 137 – 145, March 2011.

[20] S. Fiori, "Solving minimal-distance problems over the manifold of real symplectic matrices," *SIAM Journal on Matrix Analysis and Applications*, Vol. 32, No. 3, pp. 938 – 968, 2011.

[21] 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.

[22] 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

[23] C. Fraikin, K. Hüper and P. Van Dooren, "Optimization over the Stiefel manifold," in *Proceedings in Applied Mathematics and Mechanics*, Vol. 7, December 2007.

[24] H. Gong, C. Pan, Q. Yang, H. Lu and S. Ma, "A semi-supervised framework for mapping data to the intrinsic manifold," in Proceedings of the *Tenth IEEE International Conference on Computer Vision*, (October 17-20, 2005, Beijing, China), 2005.

[25] J.P. Hamaker, "Understanding Radio Polarimetry IV: The full-coherency analogue of scalar self calibration," Astronomy and Astrophysics Supplement Series, Vol. 143, No. 3, pp. 515 – 534, May 2000.

[26] N.J. Higham, "Computing the polar decomposition – With applications," *SIAM Journal on Scientific and Statistical Computing*, Vol. 7, No. 4, pp. 1160 – 1174, October 1986.

[27] T. Kaneko and T. Tanaka, "Averaging matrices on Stiefel manifold," *IEICE Technical Report*, SIP2010-178, Vol. 110, No. 439, pp. 353 – 358, March 2011 (in Japanese).

[28] T. Kaneko, T. Tanaka and S. Fiori, "A method to compute averages over the compact Stiefel manifold," in *Proceeding of the International Conference on Acoustics, Speech and Signal Processing*, (ICASSP 2012, Kyoto International Conference Center (Japan), March 25 - 30, 2012), pp. 3829 – 3832, 2012.

[29] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Communications on Pure and Applied Mathematics*, Vol. 30, No. 5, pp. 509 – 541, September 1977.

[30] A. Laub, "A Schur method for solving algebraic Riccati equations," *IEEE Transactions on Automatic Control*, Vol. 24, No. 6, pp. 913 – 921, December 1979.

[31] 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.

[32] C. Li, G. López and V. Martín-Márquez, "Iterative algorithms for nonexpansive mappings on Hadamard manifolds," *Taiwanese Journal of Mathematics*, Vol. 14, No. 2, pp. 541 – 559, April 2010.

[33] J.-L. Marichal, "On an axiomatization of the quasi-arithmetic mean values without the symmetry axiom," *Aequationes Mathematicae*, Vol. 59, No. 1-2, pp. 74 – 83, 2000.

[34] S. Mittal and P. Meer, "Conjugate gradient on Grassmann manifolds for robust subspace estimation," *Image and Vision Computing*, Vol. 30, No.s 67, pp. 417 – 427, June 2012.

[35] F. Nielsen and R. Nock, "Sided and symmetrized Bregman centroids," *IEEE Transactions on Information Theory*, Vol. 55, No. 6, pp. 2882 – 2904, 2009.

[36] M.C. de Oliveira and J.W. Helton, "Solvability of linear matrix equations in a symmetric matrix variable," in *Proceedings of the* 49^{th} *IEEE Conference on Decision and Control* (Atlanta, GA, USA), pp. 804 – 809, 2010.

[37] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, 2^{nd} Edition, Berkeley, CA: Publish or Perish Press, 1979.

[38] R. Subbarao and P. Meer, "Nonlinear mean shift over Riemannian manifolds," *International Journal on Computer Vision*, Vol. 63, No. 84, pp. 1 – 20, 2009.

[39] F. Tompkins and P.J. Wolfe, "Bayesian filtering on the Stiefel manifold," Proceedings of the 2^{nd} *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing* (December 12-14, 2007, St. Thomas, U.S. Virgin Islands), pp. 261 – 264, 2007.

[40] 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.

PLACE PHOTO HERE

Tetsuya Kaneko received the B.E. from the Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology, in 2011. He is currently a graduate student at the same University.

PLACE PHOTO HERE

Simone Fiori received the Italian Laurea (Dr. Eng.) *cum laude* 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, where he is currently an Adjunct Professor. 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 155 refereed journal and conference 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) and the Bioinformatics Institute (Biopolis, Singapore).

PLACE PHOTO HERE

Toshihisa Tanaka received the B.E., the M.E., and the Ph.D. degrees from the Tokyo Institute of Technology in 1997, 2000, and 2002, respectively. From 2000 to 2002, he was a JSPS Research Fellow. From October 2002 to March 2004, he was a Research Scientist at RIKEN Brain Science Institute. In April 2004, he joined Department of Electrical and Electronic Engineering, the Tokyo University of Agriculture and Technology, 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 Department of Electrical Engineering, the University of Hawaii at Manoa. His research interests include image and signal processing, multirate systems, statistical signal processing and machine learning, brain and biomedical signal processing, and adaptive systems. He is a co-editor of Signal Processing Techniques for Knowledge Extraction and Information Fusion (with Mandic, Splinger), 2008. He served as a guest editor of special issues in journals including Neurocomputing. He has served as an associate editor of IEICE Transactions on Fundamentals. He is a chair of the Technical Committee on Biomedical Signal Processing, APSIPA. He is a senior member of IEEE, and a member of IEICE and APSIPA.