A Comparison of Two Algorithmic Recipes to Parametrize Rectangular Orthogonal Matrices

Simone Fiori*, Tetsuya Kaneko† and Toshihisa Tanaka†
* Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche,
Via Brecce Bianche, I-60131 Ancona, Italy, E-mail: s.fiori@univpm.it
† Dept. of Electrical and Electronic Engineering, Tokyo Univ. of Agriculture and Technology,
2-24-16, Nakacho, Koganei-shi, Tokyo 184-8588, Japan, E-mail: kaneko@sip.tuat.ac.jp, tanakat@cc.tuat.ac.jp

***Abstract*—The present contribution focuses on the parametrization of rectangular ('tall-skinny') orthogonal matrices, which play a fundamental role in signal processing and machine learning. Such matrices form a smooth curved space termed *compact Stiefel manifold*. The present contribution aims at illustrating a numerical comparison of two algorithmic recipes to parameterize Stiefel matrices in signal processing. A closed-form algorithmic recipe was recently presented in the paper S. Fiori, T. Kaneko and T. Tanaka, "Learning on the compact Stiefel manifold by a Cayley-transform-based pseudo-retraction map," in Proceedings of the** *2012 International Joint Conference on Neural Networks* **(WCCI-IJCNN 2012, Brisbane (Australia), June 10 - 15, 2012), pp. 3434 - 3441, 2012, while a closed-form recipe was presented in the paper G.-X. Huang, F. Yin and K. Guo, "An iterative method for the skew-symmetric solution and the optimal approximate solution of the matrix equation $AXB = C$,"** *Journal of Computational and Applied Mathematics*, **Vol. 212, pp. 231 – 244, 2008. The numerical comparison shows that closed-form solution is substantially lighter than the iterative solution in terms of computational runtime, although the computational complexity of the closed-form solution grows slightly faster than the computational complexity of the iterative solution.**

***Index Terms*—Compact Stiefel manifold; Cayley transform; Manifold pseudo-retraction and pseudo-lifting maps; Averaging on differentiable manifolds.**

I. Introduction

Signal-processing and machine-learning theories on smooth curved manifolds are based on appropriate local coordinate charts that allow parameterizing elements on curved manifolds via sets of variables living in flat spaces. Such coordinate charts permit to switch from curved spaces to flat coordinate spaces and vice-versa. A class of coordinate charts is suggested by the theory of *manifold retraction maps* (see, e.g., [1]). The fundamental idea behind manifold retraction is to exploit local maps between the neighborhood of a point on a manifold and the tangent space to the manifold at the same point. The tangent space at any point is a vector space of the same dimension of the manifold and may thus be used to parameterize the elements of the manifold in a neighborhood of such a point. In the case that the manifold possesses additionally the structure of a Lie group, the theory about manifold-retraction-based adaptive-system learning has been recently developed in [7]. The general case that the manifold of interest is not a Lie group is, however, *substantially more involved*. In particular, the computation of a map that sends a manifold-element belonging to a neighborhood of a point to a tangent space to the manifold at the base-point, termed *manifold lifting map*, appears difficult. The present paper focuses on the case of learning over the compact Stiefel manifold, that is the space of *tall-skinny* orthogonal matrices.

A theory for constructing a manifold retraction map was devised and illustrated in the earlier studies [3], [4]. Such a method combines the features of differentiable manifolds and of Lie groups, linked by the notion of *group action*. However, such a study did not focus on the computation of a lifting map associated to the constructed retraction map. In the particular case of the compact Stiefel manifold, the work [16] suggests how to construct a local parametrization based on group action that, in the present paper, will be referred to as *manifold pseudo-retraction*, and that is based on a Cayley transformation which makes the calculation of the associated *pseudo-lifting map* feasible in closed form under some fair existence conditions. The method proposed in [16] leads to a linear equation in a skew-symmetric unknown matrix to compute the pseudo-lifting map associated to the proposed pseudo-retraction map. The aim of the present paper is to compare the algorithmic recipe proposed in the contribution [16] to solve in closed-form such a matrix-type equation with the algorithmic recipe proposed in the paper [15]. The considered signal-processing application is the calculation of averages over the Stiefel manifold.

II. Manifold- retraction/lifting maps on the compact Stiefel manifold

The present section presents a manifold retraction/lifting-map theory on the Stiefel manifold. For a general reference on differential geometry, readers might see, e.g., [20].

A. Notation and problem description

A differentiable manifold M is a space which is locally parameterized by a subset of an Euclidean space. The tangent space at a point $x \in M$ to a differentiable manifold M is denoted by $T_x M$. Each tangent space is a vector space.

The present manuscript focuses on learning on the compact Stiefel manifold, which is defined by:

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

where symbol I_n denotes a $n \times n$ identity matrix and $n < p$. Its tangent space at a point $X \in \mathrm{St}(p,n)$ may be expressed as:

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

where 0_n denotes an all-zero $n \times n$ matrix.

A Lie group is an algebraic group with a differentiable manifold structure compatible with the algebraic structure. The algebraic structure is made of a set G equipped with a multiplication operation, an inversion operation and an identity element $e \in G$. The differential-geometric structure of a Lie group manifests through the Lie algebra $\mathfrak{g}$ associated to the Lie group, defined as $\mathfrak{g} \stackrel{\text{def}}{=} T_e G$.

Relevant to the present paper, the special orthogonal group of matrices, canonically denoted by $\mathrm{SO}(p)$, is defined as:

$$\mathrm{SO}(p) \stackrel{\text{def}}{=} \{R \in \mathbb{R}^{p\times p} | R^T R = I_p,\ \det(R) = 1\}. \quad (3)$$

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:

$$\mathfrak{so}(p) \stackrel{\text{def}}{=} \{\Omega \in \mathbb{R}^{p\times p} | \Omega^T = -\Omega\}, \quad (4)$$

namely, it is the set of $p \times p$ skew-symmetric matrices.

A *retraction* at a point $x \in M$ is a map $\mathcal{R}_x : T_x M \to M$ defined in some open ball B_x about $0 \in T_x M$ such that $\mathcal{R}_x(0) = x$ and $\frac{d}{dt}\mathcal{R}_x(tv)\big|_{t=0} = v$ for every $v \in B_x$. A retraction induces local coordinates on the manifold M. A map $\mathcal{L}_x : M \to T_x M$ such that $\mathcal{R}_x(\mathcal{L}_x(y)) = y$ for any y such that $\mathcal{L}_x(y)$ belongs to the neighborhood B_x is termed *lifting* map.

A well-known retraction map is the *exponential map* that may be associated to a Riemannian manifold. In some cases of interest (such as for the case of the compact Stiefel manifold), the explicit expression of the lifting map associated to the exponential map is not available, hence the need of alternative analytically-tractable retraction maps. For example, let us recall from [10], [11] that the expression of the exponential map $\exp_X : T_X\mathrm{St}(p,n) \to \mathrm{St}(p,n)$ under the *Euclidean metric* reads:

$$\begin{aligned} \exp_X(V) \quad = \quad & [X\ V] \exp\begin{pmatrix} X^T V & V^T V \\ I_m & X^T V \end{pmatrix} \times \\ & I_{2m,m} \exp\left(-X^T V\right), \end{aligned} \quad (5)$$

while the expression of the exponential map under the *canonical metric* reads:

$$\exp_X(V) = [X\ Q] \exp\left(\begin{bmatrix} X^T V & -R^T \\ R & 0_p \end{bmatrix}\right) \begin{bmatrix} I_p \\ 0_p \end{bmatrix}, \quad (6)$$

where $X \in \mathrm{St}(p,n)$, $V \in T_X\mathrm{St}(p,n)$, $I_{p,q}$ denotes a $p \times q$ pseudo-identity matrix, while Q and R denote the factors of the compact QR decomposition of the matrix $(I_n - XX^T)V$. The present authors are unaware of any closed-form expressions for the lifting maps associated to the above exponential maps.

A *group action* of a Lie group G on a differentiable manifold M is a map $\Lambda : G \times M \to M$. A group action is specified by $\Lambda_g(x)$ with $g \in G$, $x \in M$ and $\Lambda_g(x) \in M$.

The notion of group action may be employed to construct a retraction map $\mathcal{R}_x$ for a manifold $M \ni x$ acted upon by a Lie group G. Such a result may be summarized as follows. Define a Lie-group action $\Lambda : G \times M \to M$, a coordinate map $\psi : \mathfrak{g} \to G$ such that $\psi(0) = e$, a linear map $\rho_x(u) \stackrel{\text{def}}{=} \frac{d}{dt}\Lambda_{\psi(tu)}(x)\big|_{t=0}$, with $u \in \mathfrak{g}$ and a linear map $a_x : T_x M \to \mathfrak{g}$ such that $\rho_x(a_x(v)) = v$ for any $v \in T_x M$. A retraction $\mathcal{R}_x$ which is a map from $T_x M$ to M is given by:

$$\mathcal{R}_x(v) \stackrel{\text{def}}{=} \Lambda_{\psi(a_x(v))}(x). \quad (7)$$

The special orthogonal group $\mathrm{SO}(p)$ *acts* upon the compact Stiefel manifold $\mathrm{St}(p,n)$ via a group action given by:

$$\Lambda_R(X) \stackrel{\text{def}}{=} RX \in \mathrm{St}(p,n), \quad (8)$$

where $R \in \mathrm{SO}(p)$ and $X \in \mathrm{St}(p,n)$. The action of the special orthogonal group upon the compact Stiefel manifold may be employed to design a retraction map for the compact Stiefel manifold, as it was suggested in [3] and applied to optimization problems with orthogonality constraints in [4]. By means of the retraction map constructed as in (7), with $M = \mathrm{St}(p,n)$, $G = \mathrm{SO}(p)$ and $\mathfrak{g} = \mathfrak{so}(p)$, the following retraction map is obtained:

$$\mathcal{R}_X(V) = \psi(a_X(V))X. \quad (9)$$

A difficulty related to the retraction (9) is that there are no known results about its associated lifting map. A result presented in the recent papers [13], [16] overcomes such a difficulty. It is concerned with the case that the coordinate map ψ is chosen as the Cayley map defined by:

$$\mathrm{cay}(\Omega) \stackrel{\text{def}}{=} (I_p + \Omega)(I_p - \Omega)^{-1} = (I_p - \Omega)^{-1}(I_p + \Omega). \quad (10)$$

For a recent account on the numerical properties of the Cayley map, see [14]. Indeed, the general retraction map $\mathcal{R}_X(V) = \psi(a_X(V))X$ was replaced with a *pseudo-retraction map* $\hat{\mathcal{R}}_X(\Omega) \stackrel{\text{def}}{=} \psi(\Omega)X$, so that the Stiefel matrices around the point X are parameterized by the skew-symmetric variable $\Omega = a_X(V)$ instead of the tangent vector V. Likewise, a *pseudo-lifting map* was defined as a map such that $\hat{\mathcal{R}}_X(\hat{\mathcal{L}}_X(Y)) = Y$, for any Y such that $\hat{\mathcal{L}}_X(Y) \in B_X$. The lifting-map was expressed in closed form under some fair existence conditions. In summary, the papers [13], [16] defined a pseudo-retraction map:

$$\hat{\mathcal{R}}_X(\Omega) \stackrel{\text{def}}{=} \mathrm{cay}(\Omega)X. \quad (11)$$

The calculation of the pseudo-lifting map $\hat{\mathcal{L}}_X$ boils down to solving of the equation $Y = \hat{\mathcal{R}}_X(\Omega)$ for $\Omega \in \mathfrak{so}(p)$, with $Y, X \in \mathrm{St}(p,n)$ given, namely, of the equation $(I_p - \Omega)^{-1}(I_p + \Omega)X = Y$. Rearranging terms, the latter equation may be rewritten as:

$$\Omega(Y + X) = Y - X. \quad (12)$$

Any solution Ω may be taken to be a lifting $\hat{\mathcal{L}}_X(Y)$. There appears to be no known closed-form solutions of equations of the type (12) for a general-dimensional case.

B. Closed-form solution of the equation (12)

The papers [13], [16] contribute to the present matter by a closed-form solution to the problem (12) in the case that the dimension index n of the manifold $\mathrm{St}(p,n)$ is an *even number*. A possible closed-form solution to the problem (12) was found as:

$$\begin{cases} \Omega = 0 \text{ for } Y = X, \\ \Omega = (Y-X)(Y^TX - X^TY)^{-1}(Y-X)^T \\ \qquad \text{for } X^TY - Y^TX \text{ nonsingular.} \end{cases} \tag{13}$$

Recall the following properties about skew-symmetric matrices: (a) Any skew-symmetric matrix of odd size is not invertible; (b) If a skew-symmetric matrix Ω is invertible, its inverse is also skew-symmetric. It is straightforward to verify that the expression (13) returns a skew-symmetric matrix as long as the dimension n is even. It is worth remarking that: (a) In practice, the constraint on the size of the matrices in a manifold $\mathrm{St}(p,n)$ to posses an even number of columns is not restrictive in signal-processing applications as it would cost as much as an extra column computation whenever the intended number of columns is odd; (b) The size of the matrix whose inverse needs to be computed in expression (13), namely $X^TY - Y^TX$ is generally small compared to the size of the matrices in the manifold $\mathrm{St}(p,n)$, because in the real-world applications it is generally assumed that $n \ll p$.

C. Iterative solution of the equation (12)

An iterative algorithm has been proposed and studied in the contribution [15] to find the skew-symmetric solution Ω of equations of the type:

$$A\Omega B = C \tag{14}$$

with matrices A, B, C given. There is no claim of existence of a solution to the equation (12), in general, nor of its uniqueness. The main features of the algorithmic recipe proposed in the contribution [15] to solve the equation (14) may be summarized as: (a) If the matrix A is of size $m \times p$ and the matrix B is of size $p \times n$, the iterative algorithm may converge in no more than $\min\{mn, p^2\}$ iterations in the absence of roundoff errors and if the problem (14) is consistent; (b) The iterative algorithm proposed in [15] may converge to a minimal-norm solution.

The equation (12) is a special case of the equation (14) with $A = I_p$, $B = Y + X$ and $C = Y - X$. As $n < p$, the iterative algorithm can converge in no more than np iterations. A pseudocode to implement the calculation of the map $\hat{\mathcal{L}}_X(Y)$ by means of the iterative solution of the equation (14) adapted from [15] is suggested in the Algorithm 1. Note that, because of the influence of calculation errors, the theoretical exiting condition $R_i = 0$ has been replaced by the practical exiting condition $\|R_i\|_{\mathrm{F}} < \varepsilon$, with the threshold ε being chosen as an appropriate small value.

Algorithm 1 Calculation of the pseudo-lifting map $\hat{\mathcal{L}}_X(Y)$ by the iterative solution of the equation (14) adapted from [15].

```
▷ Input matrices X, Y ∈ St(p, n)
Compute B = Y + X and C = Y − X
Choose an arbitrary matrix Ω_0 ∈ so(p)
Set R_0 = C − Ω_0 B
Set P_0 = R_0 B^T
Set Q_0 = 1/2 (P_0 − P_0^T)
Set i = 0
while ‖R_i‖_F ≥ ε and i ≤ np do
    Update Ω_{i+1} = Ω_i + (‖R_i‖^2 / ‖Q_i‖^2) Q_i
    Update R_{i+1} = C − Ω_{i+1} B
    Update P_{i+1} = R_{i+1} B^T
    Update Q_{i+1} = 1/2 (P_{i+1} − P_{i+1}^T) + (tr(P_{i+1} Q_i) / ‖Q_i‖^2) Q_i
    Update i = i + 1
end while
▷ Output matrix Ω_i
```

III. Application to averaging over the compact Stiefel manifold

A statistical characterization of a set of structured sample-matrices is their empirical mean, which appears as an average matrix carrying on the same structure of the matrices to average.

Averaging over a data-set is a good method to smooth-out data and to alleviate measurement errors and random fluctuations. In the case of unconstrained data, arithmetic averaging produces the desired result. However, in the case that constraints, such as orthogonality, are to be taken into account, it is necessary to build-up averaging algorithms that take into account the geometric structure of the space that the sample-matrices belong to. This is the case, for example, of the computation of the mean shift on Riemannian matrix manifolds [19], of the computation of the centroid in the space of symmetric positive-definite matrices [18], of the averaging of measured Jones matrices in radio interferometry [23] and of the recently-emerged problem of averaging over the manifold of symplectic matrices [8], [9].

A recent account of the general problem about the computation of Riemannian means over a Riemannian manifold may be found in the contribution [2] based on the early study [17]. The computation of Riemannian means is based on casting the problem as an optimization one, on the basis of a mean spread function that depends on the geodesic distance of points on the manifold. A practical question that arises in the application of the above definition of average element is that, for a given manifold of interest, the explicit expression of the geodesic distance may not be available, as it is the case for the compact Stiefel manifold.

The recent contribution [12] presents a general-purpose averaging algorithm that is suitable for Lie groups. Averaging on non-Lie-group-type manifolds is *a substantially more involved problem*. In particular, there appear to be no solutions to the problem of averaging on the compact Stiefel manifold,

Algorithm 2 General fixed-point averaging algorithm.

$\triangleright$ 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**
    Compute $\Omega_k^{(i)} = \hat{\mathcal{L}}_{X^{(i)}}(X_k)$
  end for
  Compute matrix $\overline{\Omega}^{(i)} = \frac{1}{N}\sum_{k=1}^{N} \Omega_k^{(i)}$
  Compute $X^{(i+1)} = \left(I_p + \overline{\Omega}^{(i)}\right)\left(I_p - \overline{\Omega}^{(i)}\right)^{-1} X^{(i)}$
end for
$\triangleright$ Output matrix $X^{(I)}$

although a number of signal-processing applications requires statistical computation over the Stiefel manifold, such as data clustering [5], Bayesian filtering [21] and image and video-based recognition [22].

A. Averaging algorithm

On the basis of the two pseudo-retraction/lifting map pairs recalled in the Section II, the algorithm introduced in the paper [12] may be extended 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, by means of a pseudo-lifting map, and then the average point on the tangent space is brought back to the Stiefel manifold by a pseudo-retraction map. The N sample-matrices to average are denoted by X_k, $k \in \{1, \ldots, N\}$. The averaging algorithm related to the use of a pseudo-retraction map, and of its associated pseudo-lifting map, generates a sequence $X^{(i)} \in \mathrm{St}(p,n)$ of estimates of the empirical mean matrix of a given set of samples is sought for via a fixed-point algorithm with initial guess $X^{(0)} \in \mathrm{St}(p,n)$ and $i = 0, 1, \ldots, I$, which reads:

$$X^{(i+1)} = \hat{\mathcal{R}}_{X^{(i)}}\left(\frac{1}{N}\sum_{k=1}^{N} \hat{\mathcal{L}}_{X^{(i)}}(X_k)\right). \tag{15}$$

The general iterative algorithm that implements the iteration (15) is rendered as a pseudocode in the Algorithm 2.

In the case that the lifting map is chosen as the closed-form algorithmic recipe, its valued is computed by the equation (13). In the case that the lifting map is chosen as the result of the iterative algorithmic recipe, its value may be computed on the basis of the Algorithm 1.

B. Numerical results

The present subsection aims at comparing the numerical features of the discussed manifold-lifting evaluation method. The numerical tests were performed by running a MATLAB® 7 (64 bit) code on platform featuring an Intel® Xeon® (2.93 GHz) with 8 cores and 12GB RAM.

The first numerical test aims at comparing numerical error measures corresponding to the closed-form solution of the equation (12) explained in the subsection II-B and to the iterative solution explained in the subsection II-C. The simulation conditions are as follows: Base manifold $\mathrm{St}(10,4)$, initial value in the Algorithm 1 generated as $\Omega_0 = \frac{1}{2}(H - H^T)$, with H being a 10×10 matrix with random entries drawn from a normal distribution. In this experiment, a matrix $X \in \mathrm{St}(10,4)$ was generated randomly and a matrix $Y = \exp(\sigma\Omega)X$ was generated with $\Omega \in \mathfrak{so}(10)$ random with spread parameter $\sigma = 0.1$. A measure of discrepancy between two Stiefel matrices is defined as $\delta(X,Y) \stackrel{\mathrm{def}}{=} \|I_n - X^T Y\|_{\mathrm{F}}$, where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. In this experiment, the threshold value for the iterative Algorithm 1 was set to $\varepsilon = 0$ in order to let the iteration extend to the maximal value $np = 40$ predicted by the theory. The Figure 1 shows the error $\|\Omega(Y+X) - (Y-X)\|_{\mathrm{F}}$ during iteration of the Algorithm 1 versus the error pertaining to the closed-form solution (13). The Figure 2 shows the discrepancy between the matrix Y and the matrix $\mathrm{cay}(\sigma\Omega)X$ during iteration of the Algorithm 1 versus the discrepancy pertaining to the closed-form solution (13). The obtained numerical results show that the iterative solution needs, in general, several iterations to reach the same error level that the closed-form solution achieves in a 'single shot'. The Figure 2 also shows that the iterative algorithm reaches a discrepancy value that is exactly the same of the closed-form method, therefore, the iterative algorithm cannot do better than the closed-form method. Moreover, both Figures 1 and 2 show that the iterative algorithm is unstable: After reaching the minimum discrepancy value, the discrepancy grows.

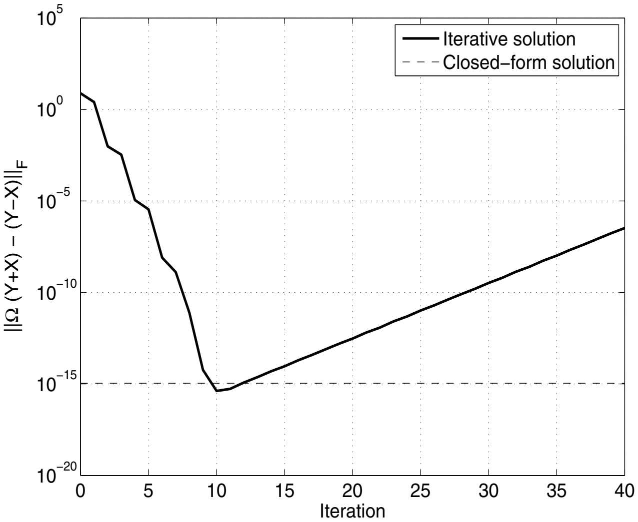

Fig. 1. First numerical test: Error measure corresponding to the closed form solution of the equation (12) explained in the subsection II-B and the iterative solution explained in the subsection II-C.

The second experiment is about averaging real-world samples over the manifold $\mathrm{St}(10,4)$. The $N = 50$ samples $X_k \in \mathrm{St}(10,4)$ to average were obtained by running a fastICA algorithm [6], which separates out 4 independent source signals from 10 mixtures, on 50 independent trials on the same separation problem. The Figure 3 illustrates the obtained results, expressed in terms of separation performance

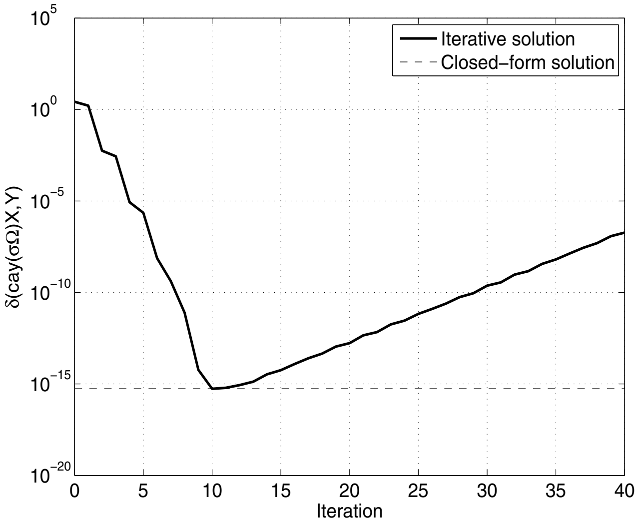

Fig. 2. First numerical test: Discrepancy measure corresponding to the closed form solution of the equation (12) explained in the subsection II-B and the iterative solution explained in the subsection II-C.

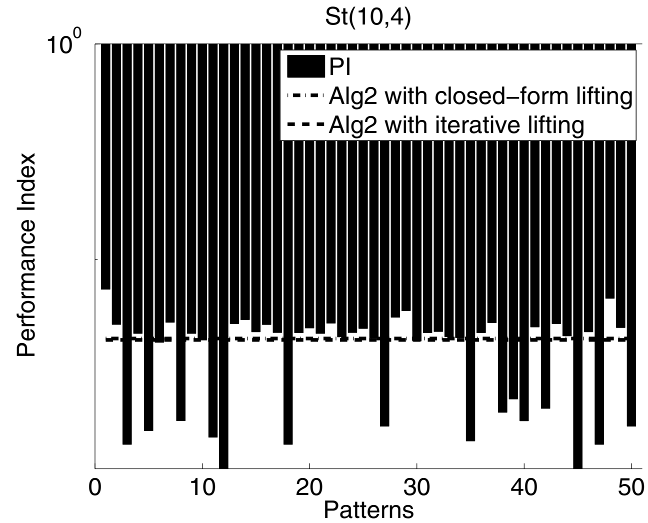

Fig. 3. Second numerical test: Results of averaging over the manifold $\mathrm{St}(10, 4)$ on real-world samples. The bars show the values of the separation performance index (PI) pertaining to each pattern $X_k \in \mathrm{St}(10, 4)$, while the horizontal linen indicate the PI corresponding to the empirical average matrix $X \in \mathrm{St}(10, 4)$ computed by the Algorithm 2 where the lifting map was computed by the closed-form expression (12) explained in the subsection II-B and by the iterative solution explained in the subsection II-C.

index (PI) [6]. The Figure 3 shows that the value of the PI corresponding to the empirical average matrix $X \in \mathrm{St}(10, 4)$ collocates in an average position with respect to the PI values of the single patterns $X_k \in \mathrm{St}(10, 4)$. The initial value in the Algorithm 1 was generated again as $\Omega_0 = \frac{1}{2}(H - H^T)$, with H being a 10×10 matrix with random entries drawn from a normal distribution, and the threshold value for the iterative Algorithm 1 was set to $\varepsilon = 10^{-10}$. The obtained results show that the Algorithm 2 endowed with a manifold-lifting operator computed by the closed-form expression (12) explained in the subsection II-B and by the iterative solution explained in the subsection II-C performs similarly on real-world data and is able to compute a meaningful average matrix on the compact Stiefel manifold.

The third numerical test aims at comparing the computational complexity of the averaging Algorithm 2 endowed with a manifold-lifting operator computed by the closed-form expression (12) explained in the subsection II-B and by the iterative solution explained in the subsection II-C on a base manifold $\mathrm{St}(100, n)$ with the size n ranging from 4 to 100 with step 4. The set of Stiefel matrices to average were generated numerically as a cloud of random points around a given center of the distribution. The center of the distribution $C \in \mathrm{St}(100, n)$ was generated by computing the Q-factor of a thin-QR decomposition of a matrix randomly generated in $\mathbb{R}^{100\times n}$ with normally-distributed entries. The 50 samples to average were generated by the rule $X_k = \exp(\sigma S_k)C$, with $S_k = \frac{1}{2}(H_k - H_k^T)$, with H_k being a matrix randomly generated in $\mathbb{R}^{100\times 100}$ with normally-distributed entries, and $\sigma > 0$ controls the spread of the distribution. The spread parameter was set to $\sigma = 0.1$. The Figure 4 shows the runtime figures pertaining to the closed-form case and to the iterative-solution case evaluated in the case that the threshold is set to $\varepsilon = 10^{-10}$ and the case that the threshold is set to $\varepsilon = 10^{-3}$. The

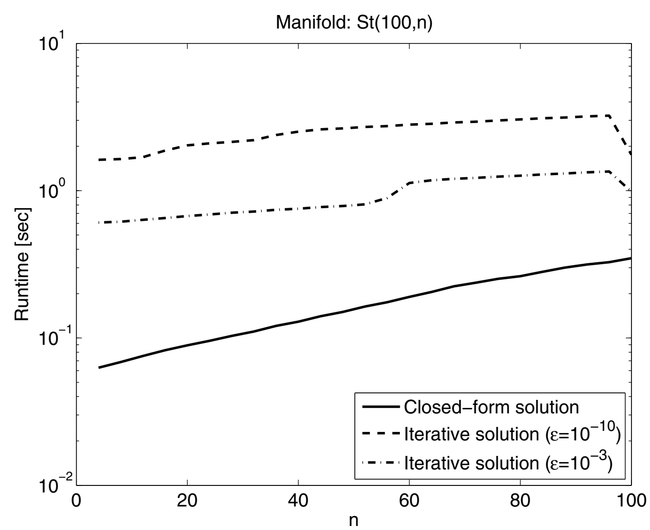

Fig. 4. Third numerical test: Computational complexity of averaging over the manifold $\mathrm{St}(100, n)$ by the Algorithm 2 where the lifting map was computed by the closed-form expression (12) explained in the subsection II-B and by the iterative solution explained in the subsection II-C.

curves represent the results obtained by averaging the runtime figures over 100 independent trials. The obtained numerical results suggest that the closed-form solution is substantially lighter than the iterative solution in terms of computational complexity, although the computational complexity of the closed-form solution grows slightly faster than that of the iterative solution (by a linear approximation, it was estimated that the slopes of the curves are 0.0175, 0.0081 and 0.0029 from bottom to top).

IV. Conclusions

The present research focuses on Stiefel-matrix parametrization by manifold retraction with application to averaging 'tall-skinny' matrices and compares two contributions on such a topic that appeared recently in the scientific literature. In particular, the present paper compares a closed-form solution by the present authors with an iterative solution published in a 2008 mathematics paper which appears to be, up to date and to the best of authors' knowledge, the only available solution to the problem.

Several numerical experiments were performed and illustrated, which aimed at comparing the numerical features of the two algorithmic methods in terms of numerical precision and of computational burden. The obtained numerical results show that the iterative solution needs, in general, several iterations to reach the same solution achieved by the closed-form solution in a 'single shot' the closed-form solution is substantially lighter than the iterative solution in terms of computational runtime, although the computational complexity of the closed-form solution grows slightly faster than the computational complexity of the iterative solution.

It is worth underlying that, while a rigorous computational burden analysis about the closed-form solution would certainly be feasible, it appears that the same analysis about the iterative algorithm would be difficult to carry out because the exiting condition based on the approximation error (see Algorithm 1 as a reference) makes the number of iterations depend on the actual problem at hand. Such phenomenon is very well evidenced by the shape of the runtime-curves of Figure 4.

Acknowledgment

S. Fiori performed this work while at the Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology, supported by a JSPS fellowship (Grant S11724 – 2011). T. Kaneko and T. Tanaka were supported in part by the grant KAKENHI 23300069.

References

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

[2] 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.

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

[4] 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.

[5] 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.

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

[7] S. Fiori, "Lie-group-type neural system learning by manifold retractions," *Neural Networks*, Vol. 21, No. 10, pp. 1524 – 1529, December 2008.

[8] 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.

[9] 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.

[10] S. Fiori, "Extended Hamiltonian learning on Riemannian manifolds: Theoretical Aspects," *IEEE Transactions on Neural Networks*, Vol. 22, No. 5, pp. 687 – 700, May 2011.

[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] 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.

[13] S. Fiori, T. Kaneko and T. Tanaka, "Learning on the compact Stiefel manifold by a Cayley-transform-based pseudo-retraction map," in Proceedings of the *2012 International Joint Conference on Neural Networks* (WCCI-IJCNN 2012, Brisbane (Australia), June 10 - 15, 2012), pp. 3434 - 3441, 2012.

[14] G. Hori and T. Tanaka, "Pivoting in Cayley tranform-based optimization on orthogonal groups," in *Proceedings of the Second APSIPA Annual Summit and Conference*, pp. 181 – 184 (Biopolis, Singapore, 14-17 December 2010), 2010.

[15] G.-X. Huang, F. Yin and K. Guo, "An iterative method for the skew-symmetric solution and the optimal approximate solution of the matrix equation $AXB = C$," *Journal of Computational and Applied Mathematics*, Vol. 212, pp. 231 – 244, 2008.

[16] T. Kaneko, T. Tanaka and S. Fiori, "Averaging matrices over the Stifel manifold," in *Proceeding of the 26th Signal Processing Symposium in Japan*, pp. 216 – 221, Sapporo, Japan, November 2011 (in Japanese).

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

[18] 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.

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

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

[21] 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.

[22] 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, November 2011.

[23] S. Yatawatta, "On the interpolation of calibration solutions obtained in radio interferometry," Monthly Notices of the Royal Astronomical Society, submitted (2012).