A METHOD TO COMPUTE AVERAGES OVER THE COMPACT STIEFEL MANIFOLD

T. Kaneko, T. Tanaka*

Dept. of Electrical and Electronic Engineering Tokyo Univ. of Agriculture and Technology 2-24-16, Nakacho, Koganei-shi Tokyo 184-8588, Japan

ABSTRACT

The aim of the present contribution is to extend the algorithm introduced in the paper 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, to compute averages over the Stiefel manifold. The idea underlying the developed algorithms is that points on the Stiefel manifold are mapped onto a tangent space, where the average is taken, and then the average point on the tangent space is projected back to the Stiefel manifold. Based on this idea, a fixed-point algorithm is developed, and numerical examples are shown to support the analysis.

Index Terms—Matrix manifolds, Averaging on matrix manifolds, Manifold retraction, QR-decomposition.

1. INTRODUCTION

Representations involving structured matrices, such as orthogonal, symmetric and unitary matrices, arise frequently in signal processing. Well-known examples are Principal Component Analysis (PCA) and Independent Component Analysis (ICA) by signal pre-whitening [3]. Moreover, in statistical data processing, the data may appear under the form of random structured matrices (see, e.g., [4]). Random matrix theory is an important and active research area and it finds applications in wireless communication, compressed sensing and information theory. For example, in the last decades, a considerable amount of work has emerged in information theory on the fundamental limits of communication channels that makes use of results in random matrix theory [8]. A useful statistical characterization of a set of structured data-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.

S. Fiori[†]

Dip.to di Ingegneria dell'Informazione Università Politecnica delle Marche, Via Brecce Bianche I-60131 Ancona, Italy

In 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. However, in the case that constraints – such as orthogonality – are to be taken into account, arithmetic averaging does not produce a consistent result (this is due to the fact that, for example, the entry-by-entry addition of two orthogonal matrices is not orthogonal, in general). Therefore, to compute averages of structured matrices, it is necessary to build-up an averaging algorithm that takes into account the geometric features of the (generally curved) space that those matrices belong to.

Fiori and Tanaka [5] presented a general-purpose averaging algorithm that works for 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 \mathcal{G} is an algebraic group with a manifold structure compatible with the algebraic structure. The algebraic structure is made of a set equipped with the general structure of the algebraic groups (namely, a multiplication operation, an inversion operation and an identity element). The differential-geometric structure manifests itself through the Lie algebra g, which is a vector space given by the tangent space to the manifold at the identity of the group. The method presented in [5] exploits the relationship between a Lie group and its associate Lie algebra and can be described as follows, with reference to Figure 1:

- The first step is to perform a left-translation ℓ_x of each sample x_k ∈ G to a neighbourhood of the identity element e of the Lie group.
- The next step is to map all samples onto the Lie algebra g, via a function P_e^{-1} , and to compute their arithmetic mean denoted by \overline{u} .
- The final step consists into mapping the Lie-algebra element u
 onto G by a function P_e and into getting the mean element x by the inverse left-translation l⁻¹/_x.

Although there is a connection between the method proposed in [5] and the notion of Riemannian mean or Karcher mean

^{*}This work was supported in part by KAKENHI 23300069.

[†]This author performed the 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).



Fig. 1. Illustration of the averaging algorithm on the Lie group \mathcal{G} proposed in [5]. The dots (•) denote sample matrices and the box symbol (\Box) denotes the empirical average matrix.

[6], the substantial difference is that the Riemannian mean is defined on the basis of a least-mean dispersion criterion that involves the Riemannian (or geodesic) distance between two points, while the method proposed in [5] does not involve any metrics and is hence more general, in this regard.

Averaging on non-Lie-group-type manifolds is a *sub-stantially more involved problem*. 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, there appear to be no reports about the problem of averaging on the Stiefel manifold (the space of orthogonal rectangular 'tall-skinny' matrices), which is not a Lie group, although a number of signal-processing applications requires statistical computation over the Stiefel manifold, such as data clustering [2], image and video-based recognition [9] as well as Bayesian filtering [7].

The aim of the present paper is to extend the algorithm introduced in the paper [5] 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 projected back to the Stiefel manifold. Most of the effort concerns the individuation of an appropriate retraction map for the Stiefel manifold and in the computation of its inverse.

2. AN AVERAGING ALGORITHM ON THE COMPACT STIEFEL MANIFOLD

The compact Stiefel manifold defined by:

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

where I_n is a $n \times n$ identity matrix and n < p, namely, the manifold $\operatorname{St}(p,n)$ is the space of the 'tall-skinny' orthogonal matrices. Its tangent space at a point $X \in \operatorname{St}(p,n)$ may be expressed as $T_X \operatorname{St}(p,n) = \{V \in \mathbb{R}^{p \times n} | X^T V + V^T X = 0\}$. A *retraction* map at a point $X \in \operatorname{St}(p,n)$ is a map $P_X :$ $T_X \operatorname{St}(p,n) \to \operatorname{St}(p,n)$. An inverse map of a retraction is termed *lifting* map and is denoted by $P_X^{-1} : \operatorname{St}(p,n) \to$ $T_X \operatorname{St}(p,n)$. An inverse retraction is defined only locally, in general. Note that the inverse of a retraction map is not unique, in general. In this section, an averaging method on the Stiefel manifold based on the notion of retraction is presented. In particular, the proposed method is based on a fixed-point algorithm.

2.1. A retraction map and its associated lifting map on the compact Stiefel manifold

In [1], it is shown that one of the retractions P_X that map a point of $T_X \operatorname{St}(p, n)$ onto $\operatorname{St}(p, n)$ is given by:

$$P_X(V) \stackrel{\text{def}}{=} qf(X+V) \tag{2}$$

where the quantity qf(X + V) denotes the Q-factor of the *thin QR decomposition* of the matrix $X + V \in \mathbb{R}^{p \times n}$ and the R-factor is a upper-triangular matrix with strictly positive elements on its main diagonal, so that the decomposition is unique. The lifting map P_X^{-1} can be represented by:

$$P_X^{-1}(Q) = QR - X,$$
 (3)

where $X, Q \in St(p, n)$ are given and R is an upper-triangular matrix with strictly positive elements on its main diagonal. Given matrices $X, Q \in St(p, n)$ if there exists an uppertriangular matrix with strictly positive elements on its main diagonal R such that $QR - X \in T_X St(p, n)$, then the lifting map (3) exists. The matrix R must satisfy the condition:

$$X^{T}(QR - X) + (QR - X)^{T}X = 0.$$
 (4)

Namely, the matrix R may be calculated by solving the linear system of $\frac{n(n+1)}{2}$ equations $(X^TQ)R + R^T(X^TQ)^T = 2I_n$.

2.2. A fixed-point averaging algorithm on the compact Stiefel manifold

Denote the sample matrices to average as $X_k \in \text{St}(p, n)$, with $k \in \{1, \ldots, N\}$, and assume that the samples X_k are distributed in a neighbourhood of a center of mass $C \in \text{St}(p, n)$. The following considerations lead to an equation characterizing the empirical mean matrix:

 The first step is to map the points X_k ∈ St(p, n) in a neighbourhood of the sought-for mean-matrix X ∈ St(p, n) onto T_XSt(p, n) by a lifting map. Such points are denoted as V_k^{def}=X_kR_k − X.



Fig. 2. Getting an average matrix by utilizing a Stiefelmanifold retraction. The dots (•) denote sample matrices and the box symbol (\Box) denotes their empirical mean-Stiefelmatrix.

- The second step is to compute the mean vector $\overline{V} = N^{-1} \sum_{k=1}^{N} V_k$, according to an arithmetic-average rule.
- The last step is to bring back the mean vector \overline{V} to $\operatorname{St}(p,n)$ by the retraction map and to get a mean matrix $X = \operatorname{qf}(X + \overline{V})$.

Such a procedure is illustrated in Figure 2. Summarizing the above procedure, a mean matrix $X \in \text{St}(p, n)$ is the solution of the matrix-type equation:

$$X = qf\left(X + \frac{1}{N}\sum_{k=1}^{N}\left(X_kR_k - X\right)\right)$$
(5)

in the variable X. (Note that the matrices R_k depend on the matrix X via the condition $X_k R_k - X \in T_X \operatorname{St}(p, n)$.) In general, however, the equation (5) cannot be solved in closed form. It may be solved by means of a fixed-point iteration algorithm, that generates a sequence $X^{(i)} \in \operatorname{St}(p, n)$ of estimates converging to the sought-for empirical mean matrix X, and that may be written as:

$$X^{(i+1)} = qf\left(\frac{1}{N}\sum_{k=1}^{N} X_k R_k(X^{(i)})\right), \ i \ge 0,$$
 (6)

where the matrix $X^{(0)} \in \text{St}(p, n)$ denotes an initial guess and the notation $R_k(X^{(i)})$ emphasizes the fact that the uppertriangular matrix R_k depends on the current estimate $X^{(i)}$ via the condition (4).

3. NUMERICAL RESULTS

The first experiment refers to the case of averaging over the manifold St(4,3). In such an experiment, three dif-



Fig. 3. Results of averaging over the manifold St(4,3). The three curves refer to three different values of the sample-set cardinality.

ferent sets of samples of different cardinality N were generated. The data-sets were generated around a common center of mass $C \in \text{St}(4,3)$ and the averaging algorithm (6) is always initialized to the same point $X^{(0)}$. The datasets count N = 10, N = 50 and N = 200 samples, respectively. In the following, the measure of discrepancy $\delta : \text{St}(p,n) \times \text{St}(p,n) \to \mathbb{R}^{+}_{0}$ between two Stiefel-manifold matrices, defined as $\delta(X,Y) \stackrel{\text{def}}{=} ||I_n - X^T Y||_{\text{F}}$, where $|| \cdot ||_{\text{F}}$ denotes Frobenius norm, is made use of. The obtained curves, in term of discrepancy $\delta(C, X^{(n)})$, are shown in the Figure 3. The figure shows that the algorithm converges steadily and in a few iterations and, as the available information becomes richer, the accuracy of the estimate improves.

The second set of experimental results concerns the problem of averaging over the manifolds St(4,3), St(30,3), St(20,5) and St(20,8) by keeping fixed the cardinality of the data-set. The Figure 4 illustrates the obtained results, again in terms of discrepancy $\delta(C, X^{(n)})$, which show that by keeping p fixed and increasing n, the convergence becomes more difficult, and the same happens when n is kept fixed and p increases. Yet the developed algorithm can cope with relatively large-size problems.

The third experimental result refers to the case of averaging real-world samples over the manifold St(5,3). The N = 50 samples $X_k \in St(5,3)$ to average were obtained by running a fastICA algorithm [3], which separates out 3 independent source signals from 5 mixtures, on 50 independent trials on the same separation problem. The Figure 5 illustrates the obtained results, expressed in terms of separation performance index (PI) [3]. The figure shows that the value of the PI corresponding to the empirical average matrix $X \in St(5,3)$



Fig. 4. Results of averaging over four manifolds St(p, n) of different sizes.

collocates in an average position with respect to the PI values of the single patterns $X_k \in St(5,3)$.

4. CONCLUSIONS

The present paper extends the algorithm introduced in [5] to compute averages over Lie groups to the 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 numerical results show that the algorithm converges steadily and in a few iterations and can cope with relatively large-size problems.

A thorough analysis of the possible retraction/lifting pairs that may be associated to the Stiefel manifold, along with their numerical implementation features, is being conducted and will be presented in a forthcoming longer report.

5. REFERENCES

- P.-A. Absil, R. Mahony and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. New Jersey: Princeton University Press, 2008.
- [2] H.E. Çetingül and R. Vidal, "Intrinsic mean shift for clustering on Stiefel and Grassmann manifolds," in Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR'09, Miami FL (USA), June 2009), pp. 1896 – 1902, 2009.
- [3] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. Wiley, New York, 2002.



Fig. 5. Results of averaging over the manifold St(5,3) on real-world samples. The bars show the values of the separation performance index (PI) pertaining to each pattern $X_k \in St(5,3)$, while the horizontal line indicates the PI corresponding to the empirical average matrix $X \in St(5,3)$ computed by means of the QR-decomposition-based algorithm (6).

- [4] 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.
- [5] 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.
- [6] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Communications on Pure and Applied Mathematics*, Vol. 30, No. 5, pp. 509 – 541, September 1977.
- [7] F. Tompkins and P.J. Wolfe, "Bayesian filtering on the Stiefel manifold," in Proceedings of the 2nd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP, St. Thomas, U.S. Virgin Islands, 12-14 December, 2007), pp. 261 – 264, 2007.
- [8] G.H. Tucci, "A note on averages over Gaussian random matrix ensembles," submitted. Preprint available from http://arxiv.org/abs/0910.0575,2011.
- [9] P. Turaga, A. Veeraraghavan, A. Srivastava and R. Chellappa, "Statistical computations on special manifolds for image and video-based recognition," *IEEE Transactions* on Pattern Analysis and Machine Intelligence, pp. 1911 – 1918, March 2011.