A Closed-Form Unsupervised Geometry-Aware Dimensionality Reduction Method in the Riemannian Manifold of SPD Matrices

Marco Congedo, Pedro Luiz Coelho Rodrigues, Florent Bouchard, Alexandre Barachant, Christian Jutten

A Closed-Form Unsupervised Geometry-Aware Dimensionality Reduction Method in the Riemannian Manifold of SPD Matrices*

Congedo M., Rodrigues P.L.C., Bouchard F., Barachant A. Jutten C.

***Abstract*— Riemannian geometry has been found accurate and robust for classifying multidimensional data, for instance, in brain-computer interfaces based on electroencephalography. Given a number of data points on the manifold of symmetric positive-definite matrices, it is often of interest to embed these points in a manifold of smaller dimension. This is necessary for large dimensions in order to preserve accuracy and useful in general to speed up computations. Geometry-aware methods try to accomplish this task while respecting as much as possible the geometry of the original data points. We provide a closed-form solution for this problem in a fully unsupervised setting. Through the analysis of three brain-computer interface data bases we show that our method allows substantial dimensionality reduction without affecting the classification accuracy.**

I. Introduction

Classification methods based on Riemannian geometry have been recently shown to be robust and accurate, for instance in computer vision [1] and in brain-computer interface based on electroencephalographic data [2]. Given a set of K data point C_k, k:1,…,K, in the manifold of N×N symmetric positive-definite (SPD) matrices, we are interested in the problem of unsupervised, 'geometry-aware' embedding of these points on a SPD manifold of smaller dimension. Unsupervised refers to the fact that we do not assume knowledge of anything else but the input points, for instance, class labels, often available for a training data set in machine learning applications. Geometry-aware has been used recently in the literature to suggest that the embedding should respect as much as possible the *geometrical structure* of the data points [3-6]. The problem is relevant for at least two reasons:

- When observed SPD matrices have high-dimension, they may be badly conditioned with respect to inversion, that is, their smallest eigenvalues are close to zero, jeopardizing the numerical stability of all Riemannian geometry manipulations, which are based on spectral functions of the eigenvalues such as the logarithm, inverse, etc. [7].

- Working in the original high-dimensional space is computationally expensive or even impossible for very high dimensions; dimensionality reduction may be necessary or desirable for computational reasons. In the era of Big Data, this concern is increasingly relevant.

The problem is related to a long-lasting one in signal processing and machine learning, where in different contexts it can be found under the name of *dimensionality (*or *sub-space) reduction*, *model identification*, *data compression*, *multidimensional scaling* and *manifold learning*. Ideally, one expect that the data points lie on some hyperplanes of the N×N matrix manifold, therefore some dimensions can be eliminated without losing information. In practice this may be the case at best approximately because of noise and estimation errors, thus the identification of an appropriate sub-dimension is to be cast as an optimization problem by means of a functional representing the amount of lost information. In this work the information of interest is the *geometry of the observations*. Traditionally, in signal processing subspace reduction is performed by PCA (or whitening) applied on the arithmetic average of the matrix set, however PCA is not geometry-aware, nor it is robust to noise. Several improvements of the PCA (robust/kernel PCA) have been proposed and some recent ones try to exploit geometrical information, in line with recent advances on Riemannian geometry. This article aims at contributing to this latter framework.

Let hereafter $\mathcal{O}_{N,P}$ indicate the set of all N×P matrices with P<N orthonormal columns. The general problem is to find a matrix $Z \in \mathcal{O}_{N,P}$ so as to obtain the embedding in the sub-manifold via transformation $Z^T C_k Z$, for all k:1,…,K. The orthonormality of the columns of Z ensures that the resulting matrices are SPD [3-6]. In an unsupervised context it has been proposed to maximize the following functional [4-5]:

$$\arg\max_{Z \in \mathcal{O}_{N,P}} \sum_k \delta^2 \left(Z^T C_k Z,\ Z^T M Z \right), \tag{1}$$

where $\delta(\cdot,\cdot)$ is a suitable distance function in the SPD manifold for the two arguments, like the natural Riemannian (affine-invariant) distance and M is a center of mass of the points. Given a pair of N×N-dimensional matrices on the SPD manifold, the natural Riemannian distance is the length of the geodesic (shortest path) joining them. It is given by

$$\delta\left(C_i, C_j\right) = \sqrt{\sum_n \log^2 \lambda_n \left(C_i^{-\frac{1}{2}} C_j C_i^{-\frac{1}{2}} \right)}, \tag{2}$$

*Research partially supported by European project ERC-2012-AdG-320684-CHESS.

M.C., P.L.C.R., F.B and C.J. authors are with the GIPSA-lab (Grenoble Images Parole Signal Automatique), CNRS, University of Grenoble Alpes and Grenoble Institute of Technology, Grenoble, France (phone: +33476826252; e-mail: marco.congedo@gmail.com).

A.B. author is an independent researcher, Grenoble, France (e-alexandre.barachant@gmail.com).

where $\lambda_n(\cdot)$, with n:1,…,N, denotes the n^{th} eigenvalue of the argument. Let us remind here that such distance is invariant by congruence, that is,

$$\delta\left(C_i, C_j\right) = \delta\left(B^T C_i B, B^T C_j B\right):\ B \text{ non-singular.} \tag{3}$$

Optimization (1) amounts to maximizing the variance (dispersion) around the mean, analogously to a classical compression by a principal component analysis (PCA). While the idea of emulating PCA in a Riemannian context appears relevant, optimizing (1) does not guarantee that the geometrical structure of the points is preserved. For instance, switching the position of any two points on the manifold leaves the variance unchanged, but disrupts dramatically the geometrical structure. Furthermore, the variance is sensitive to outliers and noisy points, as it is well known. In order to remediate to the last drawback, [6] has proposed to estimate M for each C_k as the center of mass of a number of points in its neighborhood. Such procedure may indeed be resistant to outliers, still, there is no guarantee to preserve the original geometrical structure, in that distinct neighborhoods of points may be projected in arbitrary positions of the embedding manifold. Importantly, all methods discussed so far try to optimize the sought matrix Z given a center of mass M in the embedding manifold (or several local centers of mass), but in the search the actual value of M depends on Z.

II. Method

Let us consider the set of K²-K index pair $\{i,j\}$ for i,j=1,…,K and $i\neq j$. so that asymmetric divergences can also be used instead of (2), and the set of non-negative weights

$$\sum_{i\neq j} w_{ij} = 1,\ \ w_{ij} = w_{ji}. \tag{4}$$

In this work we analyze the following functional

$$\underset{Z\in\mathcal{O}_{N,P}}{\arg\min}\ \sum_{i\neq j} w_{ij}\left[\delta^2\left(C_i,\, C_j\right) - \delta^2\left(Z^T C_i Z,\, Z^T C_j Z\right)\right], \tag{5}$$

which consider all pair-wise distances between points in the original and embedded manifold and try to minimize their mismatch. Any two points C_i and C_j in the Riemannian SPD manifold verify inequality $\delta(C_i, C_j)\geq\delta(Z^T C_i Z, Z^T C_j Z)$, thus (5) is nullified only when the geometric structure of the data points in the embedded manifold is *globally identical* to the original, out of rotations and/or translation of axes, despite Q=N-P axes have been eliminated. We may of course write (5) more simply as

$$\underset{Z\in\mathcal{O}_{N,P}}{\arg\max}\ \sum_{i\neq j}\left[w_{ij}\,\delta^2\left(Z^T C_i Z,\, Z^T C_j Z\right)\right], \tag{6}$$

showing that criterion (5) maximizes the dispersion of the points in the embedding manifold, like the criteria considered previously such as (1), albeit unlike them it does not need the approximation of one or several centers of mass on the embedding manifold. Previous works treating problem (1) have proceeded iteratively by Riemannian optimization methods on the Grassmann manifold [3-5] or other optimization methods on the orthonormal group [6]. In contrast to these previous geometry-aware attempts, here we show that the solution to problem (6) can be approximated by a symmetric eigenvalues problem in closed form. Our goal is to offer hereby a simple general framework for Riemannian sub-manifold embedding.

Given K input points functional (6) involves the evaluation of (K²-K)/2 distances, which is problematic for high-K problems. Also, like (1), it is sensitive to outliers because it considers distances among all single points. In this study we also investigate the idea of applying functional (6) not on the K input points C_k directly, but on a random subset of them or on L<K centers of mass bootstrapped from them. This way the complexity of the problem decreases from K² to L², and typically one can find a suitable L for a given data set, no matter how large K is. Furthermore, the influence of outliers is considerable lessened evaluating (6) on the L means instead that on all single points because the Riemannian geometric mean is resistant to outliers [2]. In general, we may prefer to set conveniently the number of bootstrapped means L and compute the mean on only a few points, so at to guarantee that the bootstrapped means are sufficiently dispersed in the manifold, providing enough geometrical information.

The natural Riemannian distance (2) is a non-linear function of the eigenvalue of $C_i^{-½}C_j\ C_i^{-½}$. For any pair of matrices $\{C_i, C_j\}$ let us define matrix

$$S_{ij} = \mathrm{Log}^2\left(C_i^{-½} C_j C_i^{-½}\right), \tag{7}$$

where Log(·) denotes the matrix logarithm of the argument and notice that for input data transformed by any orthogonal matrix B such as $\tilde{C}_k = B^T C_k B$, the matrices S_{ij} are transformed accordingly such as

$$\begin{aligned}
&\mathrm{Log}^2\left(\tilde{C}_i^{-½}\tilde{C}_j\tilde{C}_i^{-½}\right) = \left[\mathrm{Log}\left(B^T C_i^{-½} B B^T C_j B B^T C_i^{-½} B\right)\right]^2 \\
&= B^T \mathrm{Log}\left(C_i^{-½} C_j C_i^{-½}\right) B B^T \mathrm{Log}\left(C_i^{-½} C_j C_i^{-½}\right) B \\
&= B^T \mathrm{Log}^2\left(C_i^{-½} C_j C_i^{-½}\right) B = B^T S_{ij} B.
\end{aligned} \tag{8}$$

Eq. (8) shows that we can work directly on matrices S_{ij}, i.e., rotations with optimal properties on matrices S_{ij} can be applied to the original points C_k. When square matrix B is replaced by rectangular matrix Z, (8) will hold only approximatively, which is why our method yields an approximate solution. The squared natural Riemannian distance (2) $\delta^2(C_i, C_j)$ is the *trace* of S_{ij}. The total (weighted) squared distance across all i,j pairs is the trace of

$$S = \sum_{i\neq j}\left(w_{ij} S_{ij}\right), \tag{9}$$

thus, using well-known extremal properties of eigenvalues, the maximization of the total squared distance (6) is obtained by the P eigenvectors of S associated to its P largest eigenvalues; the problem is the distance-based analogous to a

dimensionality reduction by PCA, where instead of retaining the maximum of the energy of the points we retain the maximum of their dispersion on the manifold. The algorithm, named RME (Riemannian Manifold Embedding), follows:

***Algorithm* RME** (Riemannian Manifold Embedding)
INPUT
- K symmetric N×N positive definite matrices $\{C_1,\ldots,C_K\}$
- K²-K associated non-negative weights w_{ij} summing up to 1, with $1 \le i \ne j \le K$

OUTPUT
- A matrix $Z \in \mathcal{O}_{N,P}$

BEGIN
- Compute $S = \sum_{1 \le i \ne j \le K} \left[w_{ij} \operatorname{Log}^2 \left(C_i^{-½} C_j C_i^{-½} \right) \right]$

RETURN Z, the matrix holding in columns the P normalized eigenvectors of S associated to its P largest eigenvalues.

III. Result

Fig. 1 shows the typical inter-distance matrix obtained by applying dimensionality reduction via PCA, RME and RME applied on bootstrapped means (bmRME). In all simulations we have performed, as the one pictured in Fig. 1, the RME method is the best in preserving the inter-distances among all points, followed by bmRME and PCA.

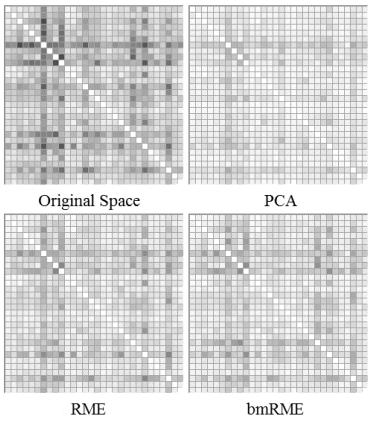

Figure 1. Typical result of applying dimensionality reduction. 30 random SPD matrices of dimension 8x8 were created with the model proposed in [8], using therein SNR=0.1 and 6 as signal dimension (signal subspace). Each plot is a *heat map* of an *inter-distance matrix*, i.e., the matrix holding all pair-wise Riemannian distances (2) among the 30 original matrices of dimension 8x8 (leftmost plot) and the 6x6 reduced matrix by, in the order, PCA, RME and RME applied on 15 means bootsrapped using 4 matrices at each bootstrap (bmRME). The darker the color, the higher the distance (zero on diagonal).

We compared the classification accuracy obtained by dimensionality reduction methods PCA, RME, bmRME and HRD (acronym of the author) [4], which are all unsupervised, to the classification accuracy obtained in the original manifold. We used three brain-computer interface data sets based on electroencephalography (EEG). As classifier we used the Riemannian minimum distance to mean (MDM) [9].

The first data set is about motor imagery. It is available at the Physionet website and comprises EEG recordings on 64 electrodes from 109 subjects [10]. We only used the data from tasks of imagined hands and feet movement (2-class problem), which yielded approximately 44 trials per subject (22 for each class). As only pre-processing, we filtered the EEG signals in the 8-30 Hz frequency band-pass by a 5-th order butterworth IIR filter and considered each trial as a segment from 0.5 to 2.5s after trial onset. From these segments, sample covariance matrices were estimated and used as input points. The accuracy obtained by MDM in the original manifold (N=64) vs. the accuracy obtained by MDM in the embedded manifold (P=24) according to methods HRD, PCA, RME and bmRME, is shown for each subject as dots in Fig. 2.

The second data set comes from EEG experiments performed in out laboratory on the P300-based game *Brain Invaders* [11]. We used data on 32 electrodes from 48 subjects. A total of 720 trials were available (120 target and 600 non-target). Of those, a random subset of 180 trials were used (30 target and 150 non-target). Data was filtered in the band-pass region 1-20 Hz by a 5-th order butterworth IIR filter. Each trial had a duration of one second starting right after a flash. We used the approach described in [12] to estimate a special form of covariance matrix, suiting signals containing event-related potentials and we performed the same comparison as for the previous database (with P=24). This analysis (Fig. 3) and the analysis obtained for all avilable trials (720 trials, data not shown) suggest the following conclusions: i.) when the sample size is small, sub-space reduction methods can overall improve the performance and ii.) there is a trend for such improvement to concern subjects with medium to high performance, while the opposite trend emerges for low-performance subjects.

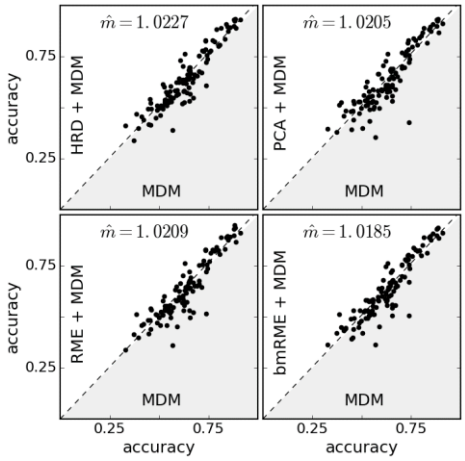

Figure 2. Scatter plot of single-subject accuracy (109 subjects, motor imagery) obtained with no dimensionality reduction (MDM) vs. HRD+MDM, PCA+MDM, RME+MDM, bmRME+MDM (all with P=24). For bmRME, 50 means of 5 points were bootstrapped. A linear regression line has been fitted to the obtained points forcing the intercept to be zero (dotted line). As reference, the area below a line with slope 1.0 is shaded. The value of the slopes obtained is printed on the plots. All slopes are higher then 1.0, but none in a statistically significant way (p>0.05).

We also analyzed a third data set (Motor Imagery), the one of the BCI competition IV (2a) [13]. It comprises EEG recordings from 9 subjects on only 22 electrodes sampling only the superior part of the head, thus reducing the dimension

for this data may not be appropriate. We used only tasks corresponding to imagination of right hand and feet movements (2-class problem). The dataset is composed of two recording sessions of 144 trials per subjects (72 per classes). The first session was used as training, and the second session was used for results evaluation (cross-session learning). As only pre-processing, we filtered the data in the 8-35Hz frequency band-pass using a 5-th order butterworth bandpass IIR and a segmentation from 3.5 to 5.5 second after the cue onset. The results in term of mean accuracy are shown in Fig. 4: as soon as P is lowered from 22 to 18 the accuracy degrades, but less so using HRD and RME as compare to using bmRME and, even more, PCA.

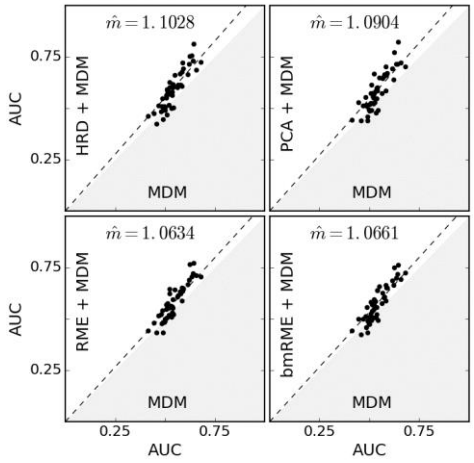

Figure 3. Scatter plot of the single-subject area under the ROC curve (AUC, 48 subjects, P300) obtained with no dimensionality reduction (MDM) vs. HRD+MDM, PCA+MDM, RME+MDM, bmRME+MDM (P=24). For bmRME, 50 means of 5 points were bootstrapped. See caption of Fig.2. All observed slopes are significantly higher then 1.0 ($p<0.001$), showing that all reduction methods perform better as compared to MDM in the original space. This effect disappears if we consider all 720 available trials (data not shown), i.e., all slopes in this case are not different from 1.0.

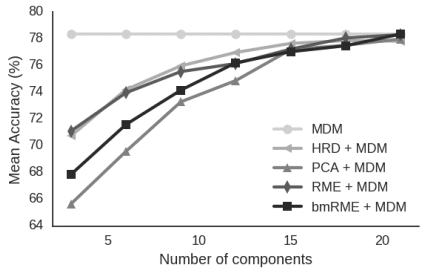

Figure 4. Mean accuracy across the 9 subject as a function of number of components (P), the sub-space dimension, for the 2-class classification problem of data set BCI competition IV 2a. The straight line indicates the accuracy with no dimensionality reduction (constant across values of P) and the curves the mean accuracy obtained with P=3, 6,…,21, for data reduced by HRD [4], PCA, RME and bmRME. For bmRME, 20 means of 3 points were bootstrapped.

IV. Conclusion

We have proposed a closed-form solution for the problem of unsupervised geometry-aware sub-manifold embedding for the manifold of symmetric positive-definite matrices. The resulting method, RME (Riemannian Manifold Embedding) and the extension based on bootstrapped means allow dimensionality reduction with no loss of classification accuracy in the three brain-computer interface data bases we have analyzed, as long as the sub-dimension is not underestimated. The advantage of these and the previously published geometry-aware method [4] over a principal component analysis does not appear dramatic, but it is appreciable for small dimensions. Further research should establish under what conditions geometry-aware methods are advantageous, both theoretically and in practice.

Geometry-aware Riemannian manifold learning is a recent research direction. The general method we have proposed is by far the most simple currently available. It is also general, in that it is unsupervised, thus it can be used on any kind of data and not only for classification purposes.

References

[1] A. Cherian, S. Sra S, "Positive Definite Matrices: Symmetric positive definite (SPD) matrices Data Representation and Applications to Computer Vision", in *Algorithmic Advances in Riemannian Geometry and Applications*, H.D. Minh and V. Murino (Eds.), Spinger, 93-114, 2016.

[2] M. Congedo, A. Barachant, R. Bhatia, "Riemannian Geometry for EEG-based Brain-Computer Interfaces; a Primer and a Review", submitted for publication.

[3] M. Harandi, M. Salzmann, R. Hartley, "From manifold to manifold: Geometry-aware dimensionality reduction for SPD matrices", in *Proc. European Conf. on Computer Vision*, 17-32, 2014.

[4] M. Harandi, M. Salzmann, R. Hartley, "Dimensionality Reduction on SPD Manifolds: The Emergence of Geometry-Aware Methods", arXiv: 1605.06182, 2016.

[5] I. Horev, F. Yger, M. Sugiyama, "Geometry-aware principal component analysis for symmetric positive definite matrices", *Machine Learning, doi:10.1007/s10994-016-5605-5*, 2016.

[6] A. Davoudi, S.S Ghidary, K. Sadatnejad, "Dimensionality reduction based on Distance Preservation to Local Mean (DPLM) for SPD matrices and its application in BCI", arXiv:1608.00514, 2016.

[7] R. Bhatia, *Positive Definite Matrices*, Prin. Univ. Press, 2015.

[8] M. Congedo, A. Barachant, E. Kharati Koopaei, "Fixed Point Algorithms for Estimating Power Means of Positive Definite Matrices", *IEEE Transactions on Signal Processing, in press*, 2017.

[9] A. Barachant, S. Bonnet, M. Congedo, C. Jutten. "Multi-Class Brain Computer Interface Classification by Riemannian Geometry", *IEEE Transactions on Biomedical Engineering*, 59(4), 920-928, 2012.

[10] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, J.R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system", *IEEE Transactions on Biomedical Engineering*, 51(6), 1034-1043, 2004.

[11] M. Congedo, M. Goyat, N. Tarrin, L. Varnet, B. Rivet, G. Ionescu, et al. ""Brain Invaders": a prototype of an open-source P300-based video game working with the OpenViBE platform", in *Proc. 5th International BCI Conference, Graz, Austria, 280-283*, 2011.

[12] A. Barachant, M. Congedo, "A plug & play P300 BCI using information geometry", *arXiv: 1409.0107*, 2014.

[13] M. Tangermann, K.R., Müller, A. Aertsen, N., Birbaumer, C. Braun, C. Brunner, et al, "Review of the BCI Competition IV", *Frontiers In Neurosciences*, 6:55. doi: 10.3389/fnins.2012.00055.