

Parameters estimate of Riemannian Gaussian distribution in the manifold of covariance matrices

Paolo Zanini, Marco Congedo, Christian Jutten, Salem Said, Yannick Berthoumieu

► **To cite this version:**

Paolo Zanini, Marco Congedo, Christian Jutten, Salem Said, Yannick Berthoumieu. Parameters estimate of Riemannian Gaussian distribution in the manifold of covariance matrices. 9th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2016), Jul 2016, Rio de Janeiro, Brazil. <hal-01325055>

HAL Id: hal-01325055

<https://hal.inria.fr/hal-01325055>

Submitted on 1 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Parameters estimate of Riemannian Gaussian distribution in the manifold of covariance matrices

Paolo Zanini and Marco Congedo and Christian Jutten
 GIPSA-Lab, CNRS, University of Grenoble Alpes,
 Grenoble Institute of Technology
 Grenoble, France
 Email: paolo.zanini@gipsa-lab.grenoble-inp.fr

Salem Said and Yannick Berthoumieu
 IMS, CNRS, University of Bordeaux
 Bordeaux, France

Abstract—The study of \mathcal{P}_m , the manifold of $m \times m$ symmetric positive definite matrices, has recently become widely popular in many engineering applications, like radar signal processing, mechanics, computer vision, image processing, and medical imaging. A large body of literature is devoted to the barycentre of a set of points in \mathcal{P}_m and the concept of barycentre has become essential to many applications and procedures, for instance classification of SPD matrices. However this concept is often used alone in order to define and characterize a set of points. Less attention is paid to the characterization of the shape of samples in the manifold, or to the definition of a probabilistic model, to represent the statistical variability of data in \mathcal{P}_m . Here we consider Gaussian distributions and mixtures of Gaussian distributions on \mathcal{P}_m . In particular we deal with parameter estimation of such distributions. This problem, while it is simple in the manifold \mathcal{P}_2 , becomes harder for higher dimensions, since there are some quantities involved whose analytic expression is difficult to derive. In this paper we introduce a smooth estimate of these quantities using convex cubic spline, and we show that in this case the parameters estimate is coherent with theoretical results. We also present some simulations and a real EEG data analysis.

I. INTRODUCTION

The study of Symmetric Positive Definite (SPD) matrices has been recently become widely popular in many engineering applications, like radar signal processing [3], mechanics [16], computer vision [23][24], image processing [9], and medical imaging (especially diffusion tensor imaging [17] and brain-computer interface [4]). In particular, great improvements come from information geometry, since it has been known for some time that the space \mathcal{P}_m of $m \times m$ SPD matrices can be equipped with a Riemannian metric. This metric, usually called Rao-Fisher or affine-invariant metric, gives it the structure of a Riemannian manifold (specifically a homogeneous space of non-positive curvature). Given two elements $C_1, C_2 \in \mathcal{P}_m$, the Riemannian distance $d_R : \mathcal{P}_m \times \mathcal{P}_m \rightarrow \mathbb{R}_+$ induced by the Rao-fisher metric is defined as

$$d_R^2(C_1, C_2) = \text{tr}(\log^2(C_1^{-1/2}C_2C_1^{-1/2})) = \sum_{i=1}^m \log^2(\lambda_i), \quad (1)$$

where $\lambda_1, \dots, \lambda_m$ are the eigenvalues of the matrix $C_1^{-1/2}C_2C_1^{-1/2}$ or of a similar matrix $C_1^{-1}C_2$. An exhaustive description of the theoretic aspects of Rao-Fisher metric and Riemannian distance can be found in [6][22]. Here we remark only two useful properties. Affine-invariance: for any invertible

matrix B

$$d_R(BC_1B^T, BC_2B^T) = d_R(C_1, C_2), \quad (2)$$

and self-duality

$$d_R(C_1^{-1}, C_2^{-1}) = d_R(C_1, C_2). \quad (3)$$

Distance (1) is used to define the concept of Riemannian center of mass [1][15] (or barycentre, geometric mean, Karcher mean, etc.). Given a set $\{C_1, \dots, C_N\}$ of matrices in \mathcal{P}_m , the Riemannian center of mass \bar{C}_N is the unique global minimizer of:

$$\mathcal{E}_N(C) = \sum_{n=1}^N d_R^2(C_n, C) \quad (4)$$

and it is viewed as a representative of the points C_1, \dots, C_N in the manifold. A large literature is devoted to the problem of defining algorithms to find the minimizer of (4) in an efficient way (see [8] and [11] and references therein) or to make the concept of barycentre essential to many applications and procedures, like, for instance, classification of SPD matrices [4][13]. However this concept (or some robust alternatives, as the Riemannian median [2][3]) is often used alone in order to define and characterize a set of points. Less attention is paid to the characterization of the shape of samples in the manifold, or to the definition of a probabilistic model, able to represent the statistical variability of data in \mathcal{P}_m . The more common and natural statistical distribution is the Gaussian distribution. It can also be defined on the manifold \mathcal{P}_m and we call it Riemannian Gaussian distribution. It depends on two parameter $\bar{C} \in \mathcal{P}_m$ and $\sigma > 0$ and its probability density function is defined as

$$p(C|\bar{C}, \sigma) = \frac{1}{\zeta(\sigma)} \exp\left\{-\frac{d_R^2(C, \bar{C})}{2\sigma^2}\right\}, \quad (5)$$

where $\zeta(\sigma)$ is a normalization factor depending only on σ . A Riemannian Gaussian distribution can be obviously used to define mixtures of Riemannian Gaussian distributions, in order to describe a wide variety of probability distributions. First attempts to describe (5) can be found in [18], even if the treatment there remains incomplete and is based on asymptotic results. Reference [20] provides a complete description of (5). Gaussian distribution and mixtures of Gaussian distributions in the Riemannian framework have been used to improve the performances of SPD matrices classification in the context of texture image [20][21]. Besides the good results obtained in

[20], this analysis is limited to the manifold \mathcal{P}_2 . However in real applications, like the analysis of Electroencephalography (EEG) covariance matrices, we deal with matrices of high dimension. Increasing the dimension m of the manifold, the estimation of the parameters becomes problematic. Our goal in this paper is to overcome this drawback. In Section II we describe more in details the Riemannian Gaussian distribution and the possible problems which can arise in parameter estimation. Then, in Section III, we present a technique to overcome this drawback, through a smooth approximation of the normalization factor $\zeta(\sigma)$. Finally, in Section IV we show some results through numerical examples and a real application with EEG data.

II. RIEMANNIAN GAUSSIAN DISTRIBUTION

A first important issue for a complete description of (5) is the explicit definition of the normalization factor $\zeta(\sigma)$. In [20] it is shown that an analytic expression can be derived only for $m = 2$, while, for $m > 2$, $\zeta(\sigma)$ is defined through the formula

$$\zeta(\sigma) = k_m \int_{\mathbb{R}^m} e^{-\frac{r_1^2 + \dots + r_m^2}{2\sigma^2}} \prod_{i < j} \sinh\left(\frac{|r_i - r_j|}{2}\right) \prod_{i=1}^m dr_i, \quad (6)$$

where e^{r_1}, \dots, e^{r_m} are the eigenvalues of $C \in \mathcal{P}_m$, and

$$k_m = \frac{\pi^{\frac{m^2}{2}} 8^{\frac{m(m-1)}{4}}}{m! \Gamma_m(m/2)} \quad (7)$$

with Γ_m the multivariate Gamma function. Even if formula (6) seems quite complicated, it can be evaluated using Monte Carlo integration, allowing to build tables of $\zeta(\sigma)$ as a function of σ , for different values of m . The evaluation of $\zeta(\sigma)$ is necessary for parameters estimation of distribution (5), specifically for the estimate of σ . In [20] it is shown that \hat{C}_N , the Maximum Likelihood Estimate (MLE) of \bar{C} , corresponds to the well known Riemannian center of mass of C_1, \dots, C_N , while $\hat{\sigma}_N$, the MLE of σ , is the solution of equation

$$\sigma^3 \frac{d}{d\sigma} \log \zeta(\sigma) = \mathcal{E}_N(\hat{C}_N). \quad (8)$$

To solve equation (8) the tabulated values of $\zeta(\sigma)$ are considered and $\hat{\sigma}_N$ is taken as that σ which better solve (8). However, Monte Carlo integration provides unstable estimates of $\zeta(\sigma)$ (and, hence, of $\log \zeta(\sigma)$) as the dimension m and σ increase. This reflects in an unexpected behavior of $\hat{\sigma}_N$, shown in the top panels of Figure 1. Indeed, $\hat{\sigma}_N$ does not result to be continuous with respect to the mean square error \mathcal{E}_N , and this problem seems to be more prominent as m and σ increase.

If we consider the parameter $\eta = -\frac{1}{\sigma^2}$, it can be shown that the function $\psi(\eta) = \log \zeta$ is a strictly convex function. Then, $\psi'(\eta)$ is an increasing function. Since problem (8) can be reformulated to find $\hat{\eta}_N$, the MLE of η , as the solution of

$$\psi'(\eta) = \frac{\mathcal{E}_N(\hat{C}_N)}{2}, \quad (9)$$

it is clear that $\hat{\eta}_N$, and then $\hat{\sigma}_N$, need to be a continuous function of $\mathcal{E}_N(\hat{C}_N)$. In the next Section we present the main contribution of this paper, i.e., exploiting the properties of $\psi(\eta)$ in order to avoid this continuity problem. Our goal is thus to provide a tractable instrument to describe and analyze Riemannian distributions even for relatively high dimension of the manifold \mathcal{P}_m , thus for a wide variety of real applications.

III. PARAMETERS ESTIMATE

The procedure we propose here is to consider the convexity property of $\psi(\eta)$. We evaluate $\log \zeta$ for different values of η using the Monte Carlo integration cited above. Then, we use these points as starting estimates to fit a curve, constrained to be convex, which will represent $\psi(\eta)$. In order to build a convex approximation we consider spline functions.

A spline function $f(x)$ is a piecewise polynomial function defined on the interval $[a, b]$. It depends on the number of knots $a = x_0 < x_1 < \dots < x_N = b$. A spline of degree K is composed by a polynomial of order $K - 1$ on each interval, and the connections at the knots are made such that the spline is globally continuous up to the derivative of order $K - 2$. The most common spline used in the application are the spline of order 4, named cubic spline. Cubic spline are very important in the approximation of a set of data points through a smooth function (where in general smooth means that the second derivative is quite regular). A strong result says that, given a sample of points $(x_i, y_i), \dots, (x_N, y_N)$ the function f that minimizes the quantity

$$\sum_{i=1}^N (f(x_i) - y_i)^2 + \lambda \int f''(x)^2 dx$$

is a cubic spline [5].

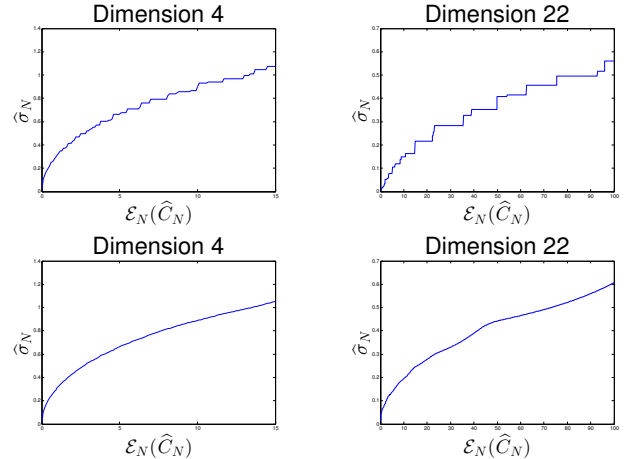


Fig. 1. $\hat{\sigma}_N$ as a function of $\mathcal{E}_N(\hat{C}_N)$ for dimension 4 (left panels) and 22 (right panels). On the top panels $\hat{\sigma}_N$ is found using the tabulated value of $\log \zeta$ provided by the Monte Carlo estimates, on the bottom panels the smooth spline approximation of $\log \zeta$ is exploited.

Cubic spline are not necessarily globally convex. Nevertheless it is possible to define, imposing some constraints, convex cubic spline [7][19]. To implement convex cubic spline we use the Shape Language Modeling (SLM) Matlab toolbox. In Figure 2 the spline approximation of $\psi(\eta)$ in the case of \mathcal{P}_{22} is depicted¹. The knots (green vertical line in the figures) used to build spline function have been chosen to be equally spaced in σ , hence they are not in η .

¹We chose $m = 22$, since it corresponds to the dimension of the real data example described in Section IV.

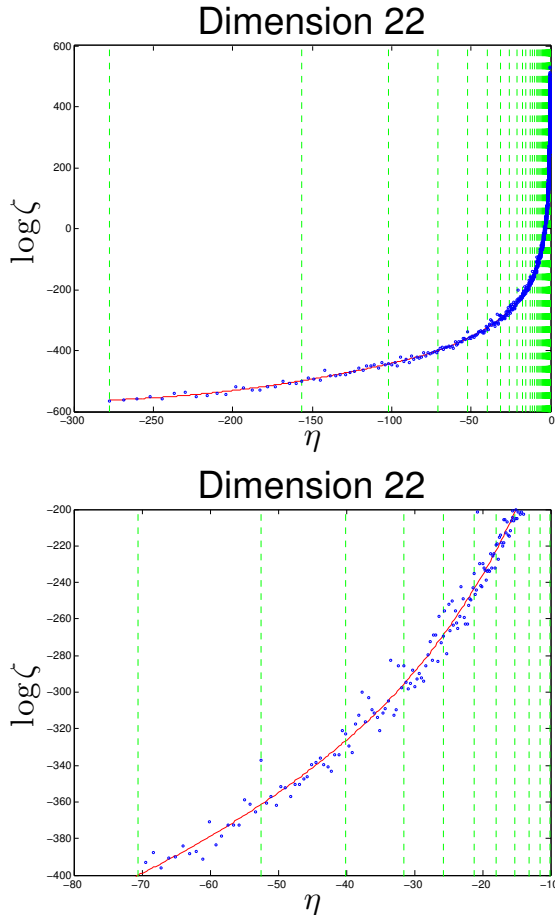


Fig. 2. Fitting of $\log \zeta$ as a function of η . Convex cubic spline (red line) fitted on the Monte Carlo estimates (blue points) for \mathcal{P}_{22} . Vertical green lines represent the knots used to build the spline function. Figure on the bottom is a zoom of the figure on the top for η varying between -70 and -15.

Now we can use this approximation of $\log \zeta$ in order to obtain the $\hat{\sigma}_N$. As can be appreciated in the bottom panels of Figure 1, $\hat{\sigma}_N$ is now a continuous function of $\mathcal{E}_N(\hat{C}_N)$, as expected. We now can use Riemannian distributions to analyze several statistical problems, and we might consider any dimension of \mathcal{P}_m . In the next Section we focus on a numerical example and a real EEG data analysis regarding the classification problem of covariance matrices.

IV. RESULTS

Riemannian Gaussian distributions can be used, for instance, for the classification of data lying in \mathcal{P}_m . Classification in \mathcal{P}_m is a challenging problem in various applications based on signal and image analysis ranging from medical applications [4][14], computer vision [23][10] to remote sensing [9][12]. In this paper, we are interested in a motor imagery experiment for which a spatial mapping of the brain is analyzed from recorded EEG signals, with electrodes placed along the scalp [4]. The work proposed in [4] has shown that the structure of the covariance matrix between electrodes is an excellent feature to recognize mental actions. Due to the fact that, for all these previous works, a probabilistic model is not available, authors proposed a classification procedure based on Minimum Distance to Mean (MDM) classifier. From [4], MDM is defined as follows. Given K groups and a training

phase where the barycentres $\hat{C}(k)$ of the groups ($k = 1, \dots, K$) are estimated, a new observation C_i is assigned to the \hat{k} group according to the classification rule

$$\hat{k} = \arg \min_{k \in \{1, \dots, K\}} \{d_R(C_i, \hat{C}(k))\}. \quad (10)$$

This rule takes into consideration the distance of the new observation to the barycentres, ignoring information on the variability of the groups, encoded by the parameter σ in the Riemannian Gaussian distribution. Thus, the principle of Bayesian classification can be used exploiting such a distribution. In this case, the classification rule based on the a posteriori distribution reads

$$\hat{k} = \arg \min_{k \in \{1, \dots, K\}} \left\{ \log \zeta(\hat{\sigma}(k)) + \frac{d_R^2(C_i, \hat{C}(k))}{2\hat{\sigma}^2(k)} \right\}. \quad (11)$$

In order to allow more general distributions, in [20] the authors, starting from Gaussian distribution (5), define mixtures of Gaussian distributions in a straightforward way such as

$$p(C) = \sum_{j=1}^M \omega_j p(C | \bar{C}_j, \sigma_j), \quad (12)$$

where ω_j are positive weight adding up to 1. To estimate the parameters of mixtures model, in [20] an extension of EM algorithm for data lying in \mathcal{P}_m is proposed. Here this algorithm is integrated with the spline approximation of the different normalization constants. Using this class of distributions, the Bayesian classification rule is updated accordingly.

We consider here a simulated example where two classes (i.e., $K = 2$) are generated according to the Gaussian law (5) in \mathcal{P}_{22} . To generate random samples from (5) we follow the procedure described in [20]. The two groups are characterized by different parameters σ , specifically $\sigma(1) = 0.15$ and $\sigma(2) = 0.17$. It is worth to point out that for Riemannian Gaussian distribution σ is a dispersion parameter, and these two values allow here a different dispersion between the two groups. $\bar{C}(1)$ corresponds to the identity matrix and then we consider four different frameworks where we gradually increase $d_R(\bar{C}(2), \bar{C}(1))$. We generate 60 SPD matrices for each sample and then we repeat the same experiment five times for each framework. In Table I we display the mean classification accuracy over the five runs for two classifiers, the MDM and the Bayes classifier using simple Gaussian distribution. In the following we consider also Gaussian Mixtures (GM) with M modes. The case of simple Gaussian can be seen as a Gaussian mixture with $M = 1$. Classification accuracy is evaluated through a 30-fold cross-validation procedure. Looking at the results in Table I we can observe that, when the two barycentres are very close, the different variability of the classes is the only way to classify correctly. As the distance increases, the performances of the two classifiers become closer, until the limit situation of framework 4, where the barycentres are very far and the variability is no longer essential to obtain good classification performances.

| $d_R(\overline{C}(2), \overline{C}(1))$ | MDM | GM ($M = 1$) |
|---|------|----------------|
| 0.26 | 59.5 | 75.2 |
| 0.36 | 67.5 | 75.3 |
| 0.6 | 84.5 | 95.5 |
| 0.91 | 97.2 | 98.2 |

TABLE I. SIMULATED EXAMPLE - CLASSIFICATION ACCURACY IN 30-FOLD CROSS-VALIDATION VARYING THE DISTANCE OF THE BARYCENTRES OF THE TWO CLASSES.

To conclude, we analyze a real EEG motor imagery problem. The analyzed dataset is the one from BCI competition analyzed in [4]. It contains EEG data from nine subjects who perform 4 kinds of motor imagery (right hand, left hand, foot, and tongue imagined movements). Hence in this experiment we have $K = 4$. A total of 576 trials per subject are available, each trial corresponding to a movement (balanced experiment, i.e., 144 trials per class). EEG signals are recorded using 22 electrodes, hence covariance matrices here belong to \mathcal{P}_{22} . In Table II the classification accuracies obtained through a 30-fold cross-validation procedure are displayed. In this case we considered also Bayesian classification using Gaussian mixtures, varying the number of modes from 1 to 4.

| Subject | MDM | $M = 1$ | $M = 2$ | $M = 3$ | $M = 4$ | M best |
|---------|--------------|--------------|--------------|--------------|--------------|----------|
| 1 | 79.17 | 78.99 | 74.48 | 76.74 | 77.26 | 78.99 |
| 2 | 47.74 | 48.96 | 50.69 | 50.17 | 45.49 | 50.69 |
| 3 | 76.91 | 76.91 | 78.30 | 78.30 | 80.38 | 80.38 |
| 4 | 59.72 | 59.90 | 59.90 | 57.29 | 55.03 | 59.90 |
| 5 | 40.62 | 40.62 | 39.41 | 38.37 | 38.54 | 40.62 |
| 6 | 46.88 | 46.35 | 38.89 | 37.15 | 42.01 | 46.35 |
| 7 | 75.52 | 75.00 | 79.51 | 79.51 | 77.08 | 79.51 |
| 8 | 77.26 | 77.78 | 80.38 | 83.33 | 84.20 | 84.20 |
| 9 | 77.08 | 77.26 | 78.99 | 80.90 | 80.90 | 80.90 |
| Tot | 64.54 | 64.64 | 64.51 | 64.64 | 64.54 | 66.84 |

TABLE II. REAL EEG MOTOR IMAGERY PROBLEM - CLASSIFICATION ACCURACY IN 30-FOLD CROSS-VALIDATION. THE BEST ACCURACY FOR EACH SUBJECT IS PRINTED IN BOLD

A first thing we can observe from Table II is that classification accuracies of the MDM and the simple Gaussian ($M = 1$) classifiers provide comparable results. This is due to the fact that in this specific example the four classes present a similar variability. In this case, as already mentioned, rules (10) and (11) coincide. Secondly, classification accuracy of the best Bayesian classifier is, for all subjects, at least as performant as the MDM. The best classifier has not always the same number of modes. This implies that, even if the experiment is common for every subject, the shapes of data clouds is highly variable between individuals. This aspect is not usually studied in the literature of Riemannian geometry applied to EEG data, and can be useful and interesting in this domain. The example described here is a first attempt of an analysis in this direction.

It is worth to point out a comment on computational costs of the illustrated procedures. First of all, Monte Carlo integration and spline approximation done to obtain a table of $\zeta(\sigma)$ are offline operations which have to be performed only once. Then, when a table for $\zeta(\sigma)$ is available, classification algorithms can be implemented without worrying about the computational time needed to build the table. Secondly, we can observe in Table III the computational time needed to perform the 30-fold cross-validation for subject 1 for the the classification algorithms analyzed. It is interesting to observe that the Bayes classifier with Gaussian distributions requires the same time as the reference Minimum Distance to Mean

classifier. On the contrary, considering mixtures of Gaussian distributions makes the classification algorithm more time consuming.

| Method | MDM | $M = 1$ | $M = 2$ | $M = 3$ | $M = 4$ |
|----------|------|---------|---------|---------|---------|
| Time (s) | 40.8 | 44 | 361.9 | 427 | 412.7 |

TABLE III. COMPUTATIONAL TIME NEEDED TO PERFORM 30-FOLD CROSS-VALIDATION FOR SUBJECT 1.

To conclude, in this paper we analyzed Gaussian distribution and mixtures of Gaussian distribution on the manifold \mathcal{P}_m of SPD matrices. In particular we introduced an approximation technique to derive the normalization constant of the distribution, based on convex cubic spline. This allows to solve a problem in the parameter estimation, making the estimate coherent with the theoretical results presented in [20]. In this way Riemannian distribution can be used for applications in any dimension m , for classification analysis, introducing Bayesian classifiers, or for the observation of the shape of data clouds in the Riemannian manifold.

ACKNOWLEDGMENT

This work have been partly supported by the ERC project CHESS: 2012-ERC-AdG-320684.

REFERENCES

- [1] T. Ando, C.K. Li, R. Mathias, "Geometric means", *Linear Algebra Appl.*, vol. 385, 2004, pp. 305-334.
- [2] M. Arnaudon, L. Yang, F. Barbaresco, "Stochastic algorithms for computing p-means of probability measures, geometry of Radar Toeplitz covariance matrices and applications to HR Doppler processing", *International Radar Symposium (IRS)*, 2011, pp. 651-656.
- [3] M. Arnaudon, F. Barbaresco, L. Yang, "Riemannian medians and means with applications to Radar signal processing", *IEEE J. Sel. Topics Signal Process.*, vol 7(4), 2013, pp. 595-604.
- [4] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, "Multiclass brain-computer interface classification by Riemannian geometry", *IEEE Trans. Biomed. Eng.*, vol. 59(4), 2012, pp. 920-928.
- [5] R.H. Bartels, J.C. Beatty, B.A. Barsky, "Hermite and Cubic Spline Interpolation", Ch. 3 in *An Introduction to Splines for Use in Computer Graphics and Geometric Modelling*. San Francisco, CA: Morgan Kaufmann, 1998, pp. 9-17.
- [6] R. Bhatia, *Positive Definite Matrices*. New Jersey: Princeton University Press, 2007.
- [7] H. Bouchard, "Extremal positive spline with applications to interpolation and approximation by generalized convex functions", *Bull. Amer. Math. Soc.*, vol. 79, 1973, pp. 959-963.
- [8] M. Congedo, B. Asfari, A. Barachant, M. Moakher, "Approximate Joint Diagonalization and Geometric Mean of Symmetric Positive Definite Matrices", *PLoS ONE*, vol. 10(4), 2015.
- [9] G. Dong, G. Kuang, "Target recognition in SAR images via classification on Riemannian manifolds", *IEEE Geosci. Remote Sens. Lett.*, vol 21(1), 2015, pp. 199-203.
- [10] S. Hidot, C. Saint-Jean, "n expectation-maximization algorithm for the Wishart mixture model: application to movement clustering", *Pattern Recogn. Lett.*, vol. 31(14), 2010, pp. 2318-2324.
- [11] B. Jeuris, R. Vanderbril, B. Vandereycken, "A survey and comparison of contemporary algorithms for computing the matrix gepmetric mean", *Elec. Trans. Numer. Anal.*, vol. 39, 2012, pp. 379-402.
- [12] J. S. Lee, M. R. Grunes, T. L. Ainsworth, L. J. Du, D. L. Schuler, S. R. Cloude, "Unsupervised classification using polarimetric decomposition and the complex Wishart classifier", *IEEE Trans. Geosci. Remote Sens.*, vol. 37(5), 1999, pp. 2249-2258.

- [13] Y. Li, K. Wong, H. deBruin, "EEG signal classification based on a riemannian distance measure", Proc. IEEE Toronto Int. Conf. Sci. Technol. Humanity (TIC-STH), 2009, pp. 268-273.
- [14] R. Luis-Garcia, C.F. Westin, C. Alberola-Lopez, "Gaussian mixtures on tensor fields for segmentation: application to medical imaging", *Comp. Med. Imag. and Grap.*, vol. 35(1), 2011, pp. 16-30.
- [15] M. Moakher, "A differential geometric approach to the geometric mean of symmetric positive-definite matrices", *SIAM J. Matrix Anal. and Appl.*, vol 26(3), 2005, pp. 735-747.
- [16] M. Moakher, "On the averaging of symmetric positive-definite tensors", *Journal of Elasticity*, vol. 82(3), 2006, pp. 273-296.
- [17] X. Pennec, P. Fillard, N. Ayache, "A Riemannian framework for tensor computing", *Int. J. Comput. Vision*, vol. 66(1), 2006, pp. 41-66.
- [18] X. Pennec, "Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements", *J. Math. Imaging Vis.*, vol. 25(1), 2006, pp. 127-154.
- [19] D. L. Russell, "Computing Convex Spline Approximation", *International Journal of Information and Systems Sciences*, vol. 5, 2009, 83-97.
- [20] S. Said, L. Bombrun, Y. Berthoumieu, J. H. Manton, "Riemannian Gaussian Distributions on the Space of Symmetric Positive Definite Matrices", arXiv, 2015.
- [21] S. Said, L. Bombrun, Y. Berthoumieu, "Texture classification using Rao's distance on the space of covariance matrices", *Conference on Geometric Science of Information (GSI)*, 2015.
- [22] A. Terras, *Harmonic analysis on symmetric spaces and applications*, Vol. II. New York: Springer-Verlag, 1988.
- [23] O. Tuzel, F. Porikli, P. Meer, "Pedestrian detection via classification on Riemannian manifolds", *IEEE Trans. Pattern Anal. and Match. Intell.*, vol. 30(10), 2008, pp. 1713-1727.
- [24] L. Zheng, G. Qiu, J. Huang, J. Duan, "Fast and accurate nearest neighbor search in the manifolds of symmetric positive definite matrices", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014, pp. 3804-3808.