Estimating Positive Definite Matrices using Frechet Mean

Mehdi Jahromi, Kon Max Wong and Aleksandar Jeremic*
Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada

Keywords: Covariance Estimation, Constrained Optimization, Fréchet Mean.

Abstract: Estimation of covariance matrices is a common problem in signal processing applications. Commonly applied techniques based on the cost optimization (e.g. maximum likelihood estimation) result in an unconstrained estimation in which the positive definite nature of covariance matrices is ignored. Consequently this may result in accurate estimation of the covariance matrix which may affect overall performance of the system. In this paper we propose to estimate the covariance matrix using Fréchet mean which ensures that the estimate also has positive definite structure. We demonstrate the applicability of the proposed technique on both estimation and classification accuracy using numerical simulations. In addition we discuss some of the preliminary results we obtained by applying our techniques to high content cell imaging data set.

1 INTRODUCTION

The covariance matrix, or equivalently, the power spectral density (PSD) matrix, of the signals from a multi-sensor system is a feature useful for many purposes in statistical signal processing including detection, estimation, classification, and signal design. In a recent paper (Li and Wong, 2013) the importance of power spectral density matrix in classification of EEG signals was demonstrated.

In many applications of signal processing, the covariance matrix of the observed signal is utilized as a feature from which information is extracted. Often, for extraction of information, averaging and interpretation of these matrices are needed. To develop algorithms for such evaluations, one important fact has to be born in mind that the structural constraints, i.e., Hermitian symmetry and positive definiteness, on such matrices must be maintained (Li and Wong, 2013), (Jeuris et al., 2012). More specifically in high content cell images many of the features that are commonly used in analysis exhibit large degree of random variations from class to class (i.e well to well). In these cases it is expected that covariance structure may play significant role in correct classification as their distinguishing properties may be determined by type of randomness rather than the center of the mass of the cluster corresponding to particular well of cells used for screening.

In this paper, the focus of our attention is on the estimation of the Frechet mean of the covariance matrices on the manifold $\mathcal{M}$ using the different measures of Riemannian distances. The necessary frame work and algorithms to obtain the mean of covariance matrices from group of sample covariance matrices using Riemannian distances on manifold of positive definite matrices $\mathcal{M}$ will be developed and studied in this chapter. In Section 2 we introduce the Fréchet mean based on several Riemannian distances. In Section 3 we discuss computational algorithms for calculating the proposed distance means. In Section 4 we present numerical results that demonstrate applicability of our results.

*This work was supported by Natural Sciences and Engineering Research Council of Canada.

2 FRÉCHET MEAN

The history of defining mean goes back 2500 years when the ancient Greeks introduced ten types of different means. Among them only three of them survive and are still being used. These are the arithmetic, the geometric and the harmonic means.

We use the notion of Fréchet mean to unify the method of finding the mean of positive definite matrices. The Fréchet mean is given as the point which minimizes the sum of the squared distances (Barbaresco, 2008):

$$\hat{\mathcal{S}} = \text{argmin}_{\mathcal{S}\in\mathcal{M}} \sum_{i=1}^{n} d^2(\mathbf{S}_i, \mathcal{S}) \qquad (1)$$

where $\{\mathbf{S}_i\}_{i=1}^n$ represents the symmetric positive definite matrices and $d(.,.)$ denotes the metric being used respectively.

In fact if we have a closer look at the definition

Jahromi M., Wong K. and Jeremic A..
Estimating Positive Definite Matrices using Frechet Mean.
DOI: 10.5220/0005277902950299
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing* (BIOSIGNALS-2015), pages 295-299
ISBN: 978-989-758-069-7

of arithmetic mean of positive measurement $\{x_i\}_{i=1}^n$, which is denoted as $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$, and using the usual distance, we can see that it has the variational property. This means that it minimizes the sum of the squared distances to the points x_k :

$$\bar{x} = \text{argmin}_{x\geq 0} \sum_{i=1}^{n} |x - x_i|^2 \tag{2}$$

with respect to metric

$$d(x,y) = |x-y| \tag{3}$$

In fact if we form the quadratic cost function

$$f(x) = \sum_{i=1}^{n} (x - x_i)^2 \tag{4}$$

By taking the derivative of Eq.(4) with respect to the variable x and set it equal to zero one can obtain the $\bar{x}$ which is the arithmetic mean of positive scalars $\{x_i\}_{i=1}^n$.

2.1 Riemannian Metrics

So far we only considered the Euclidean distance which is valid on the space with zero sectional curvature.

To measure the distance between two $M \times M$ covariance matrices $\mathbf{A}$ and $\mathbf{B}$ on manifold of positive definite matrices $\mathcal{M}$, we consider the metrics which have been developed to measure distance between two points on the manifold itself. The following metrics will be considered throughout the remaining chapters.

The first metric is obtained when we lift the points $\mathbf{A},\mathbf{B}$ to the horizontal subspace $\mathcal{U} \subset \mathcal{H}$ using the fiber and measure the distance between them(Li and Wong, 2013):

$$d_{R_1}(\mathbf{A},\mathbf{B}) = \text{argmin}_{\tilde{\mathbf{U}}_1,\tilde{\mathbf{U}}_2 \in U(M)} \left\| \mathbf{A}^{\frac{1}{2}}\tilde{\mathbf{U}}_1 - \mathbf{B}^{\frac{1}{2}}\tilde{\mathbf{U}}_2 \right\|_2 \tag{5}$$

where $U(M)$ denotes the space of unitary matrices of size $M \times M$. Alternatively Eq.(5) can be rewritten as:

$$\sqrt{\text{Tr}(\mathbf{A}) + \text{Tr}(\mathbf{B}) - 2\,\text{Tr}(\mathbf{A}^{\frac{1}{2}}\mathbf{B}\mathbf{A}^{\frac{1}{2}})^{\frac{1}{2}}} \tag{6}$$

In general for any positive definite matrix $\mathbf{A}$ its square root is defined as $\mathbf{A}^{\frac{1}{2}} = \mathbf{S}\sqrt{\mathcal{L}}\mathbf{D}^H$; where $\mathbf{A} = \mathbf{S}\mathcal{L}\mathbf{D}^H$ is the eigenvalue value decomposition of matrix $\mathbf{A}$ with diagonal matrix $\mathcal{L}$ consisting of eigenvalues of $\mathbf{A}$.

In Eq(5), $\mathbf{U}_1$ and $\mathbf{U}_2$ are the left and right multiplicative of singular value decomposition of $\mathbf{B}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}$ (Mardia et al., 1979). Let the points $\mathbf{A},\mathbf{B} \in \mathcal{M}$ and let $\mathbf{X}$ be a the point on the manifold at which we construct a tangent plane (it is usually denoted as $T_{\mathcal{M}}\mathbf{X}$). According to the inner-product $\langle \mathbf{A},\mathbf{B}\rangle_{\mathbf{X}} = \text{Tr}(\mathbf{X}^{-1}\mathbf{A}\mathbf{X}^{-1}\mathbf{B})$ the log- Riemannian metric is given as (Moakher, 2005):

$$d_{R_3}(\mathbf{A},\mathbf{B}) = \left\| \log(\mathbf{A}^{-\frac{1}{2}}\mathbf{B}\mathbf{A}^{-\frac{1}{2}}) \right\|_2 = \sqrt{\sum_{i=1}^{M} \log^2(\mathcal{L}_i)} \tag{7}$$

where the $\mathcal{L}_i$'s are the eigenvalues of the matrix $\mathbf{A}^{-1}\mathbf{B}$ (Absil et al., 2009). (Metric d_{R3} has been developed in various ways and has, for a long time, been used in theoretical physics).

Obtaining the Fréchet mean of set of positive definite Hermitian matrices $\{\mathbf{S}_i\}_{i=1}^n$ with respect to metric d_{R1} results in:

$$\text{argmin}_{\mathcal{S}\in\mathcal{M}} \sum_{i=1}^{n} \left\| \mathbf{S}_i^{\frac{1}{2}}\mathbf{U}_i - \mathcal{S}^{\frac{1}{2}}\mathbf{U} \right\|_2^2 \tag{8}$$

In the next section we discuss numerical/analytical methods for calculating corresponding means.

3 COMPUTATIONAL ALGORITHMS

Following the approach of (Crosilla and Beinat, 2002) we first define function g as

$$g(\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,...,\mathbf{A}_n) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\geq i} \left\| \mathbf{A}_i\mathbf{U}_i - \mathbf{A}_j\mathbf{U}_j \right\|_2^2. \tag{9}$$

where in Eq.(9) $\mathbf{U}_i$'s are unitary operators.

The next algorithm simultaneously find the set of unitary matrices$\{\mathbf{U}_i\}_{i=1}^n$ in order to minimize function $g(\mathbf{A}_1,\mathbf{A}_2,...,\mathbf{A}_n)$ in (9) and as a consequent finding the Fréchet mean with respect to metric d_{R1}.

Algorithm for computing the Fréchet mean of metric d_{R_1}:

Algorithm 1: Fréchet mean for metric d_{R_1}.

1. Initialize the positive threshold value ε. For the set $\{\mathbf{S}_i\}_{i=1}^n$ of positive definite matrices on manifold $\mathcal{M}$ find the square root of each element: $\mathbf{A}_i = \mathbf{S}_i^{\frac{1}{2}}$; $i = 1,2,...,n$.
2. For each $i = 1,2,...,n$ consider $\hat{\mathbf{A}}_i := \frac{1}{n-1}\sum_{j\neq i}^{n}\mathbf{A}_j$ and find $\hat{\mathbf{U}}_i$ which minimizes $\left\|\hat{\mathbf{A}}_i - \mathbf{A}_i\mathbf{U}_i\right\|_2$; then consider $\hat{\mathbf{A}}_{\text{inew}} := \mathbf{A}_i\hat{\mathbf{U}}_i$
3. At iteration $(k+1)$ set $\mathbf{A}_i = \hat{\mathbf{A}}_{\text{inew}}; i = 1,2,...,n$ and Evaluate g_{k+1} using Eq.(9)
4. Repeat step 2 until: $|g_k(\mathbf{A}_1,\mathbf{A}_2,...,\mathbf{A}_n) - g_{k+1}(\mathbf{A}_1,\mathbf{A}_2,...,\mathbf{A}_n)| \leq \varepsilon$.
5. Calculate $\hat{\mathbf{C}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\mathbf{A}}_i\mathbf{U}_i$.
6. The resulting Fréchet mean on manifold $\mathcal{M}$ is then obtained as $\hat{\mathcal{S}} = \hat{\mathbf{C}}\hat{\mathbf{C}}^H$

The optimization problem in Eq.(1) with respect to the metric d_{R2} and given positive definite hermitian matrices $\{\mathbf{S}_i\}_{i=1}^n$, is expressed as:

$$\hat{\mathcal{S}} = \text{argmin}_{\mathcal{S}\in\mathcal{M}} \sum_{i=1}^{n} \left\| \mathbf{S}_i^{\frac{1}{2}} - \mathcal{S}^{\frac{1}{2}} \right\|_2^2 \qquad (10)$$

The optimization problem (10) has closed form solution on manifold of positive definite matrices. It can be obtained through the following lemma:

For the set of positive definite Hermitian matrices $\{\mathbf{S}_i\}_{i=1}^n$ on manifold $\mathcal{M}$ we consider $\mathbf{L}_i = (\mathbf{S}_i)^{\frac{1}{2}}$; $i = 1,2,...,n$. Then we have:

$$\hat{\mathbf{L}} = \text{argmin}_{\mathbf{L}} \sum_{i=1}^{n} \left\| \mathbf{L}_i - \mathbf{L}^{\frac{1}{2}} \right\|_2^2 = \frac{1}{n} \sum_{i=1}^{n} \mathbf{L}_i \qquad (11)$$

The Fréchet mean with respect to metric d_{R3} for the set of positive definite matrices $\{\mathbf{S}_i\}_{i=1}^n$ in manifold $\mathcal{M}$ using Eq.(1) can be formulated as:

$$\hat{\mathcal{S}} = \text{argmin}_{\mathcal{S}\in\mathcal{M}} \sum_{i=1}^{n} d_{R3}^2(\mathbf{S}_i, \mathcal{S}) \qquad (12)$$

Before going through the algorithm for finding the optimum solution of Eq.(12) we show that this optimization problem has a unique solution. It has been demonstrated that the directional derivative of the function $f(\mathbf{X}) = \sum_{i=1}^n \left\| \log(\mathbf{A}_i^{-\frac{1}{2}} \mathbf{X} \mathbf{A}_i^{-\frac{1}{2}}) \right\|_2^2$,where $\{\mathbf{A}_i\}_{i=1}^n \in \mathcal{M}$, is given by:

$$D_{\mathbf{Y}} f(\mathbf{X}) = 2 \sum_{i=1}^{n} \left\langle \mathbf{X}^{-1} \log(\mathbf{X}\mathbf{A}_i^{-1}), \mathbf{Y} \right\rangle \qquad (13)$$

At this point we can use gradient descent algorithm to find the minimizer of $f(\mathbf{X})$.

4 EVALUATION OF FRÉCHET MEAN OF SYMMETRIC POSITIVE DEFINITE MATRICES

We have introduced different estimators corresponding to the different distance measures to find the Fréchet mean of set of symmetric positive definite matrices $\{\mathbf{S}_i\}_{i=1}^n$ on manifold $\mathcal{M}$. In order to compare the performance of each estimator we consider a population of $M \times M$ covariance matrices and find the mean of them using each estimator. We will consider different models having the same true means so that a comparison of the closeness of the different estimates to this true mean is possible.

4.1 Model Description

To come up with the first model we consider the known symmetric positive definite matrix $\mathcal{S}$ as the nominal value. Then we apply the Cholesky decomposition to it. By definition the Cholesky factor of a symmetric positive definite matrix $\mathcal{S}$ is a lower triangular matrix $\mathcal{W}$ with positive diagonal elements such that $\mathcal{S} = \mathcal{W}\mathcal{W}^H$.
We denote the Cholesky factor of $\mathcal{S}$ in the model with $\mathcal{W}$ and set $\mathcal{W}$=Chol($\mathcal{S}$); where *Chol* represents the Cholesky factor of $\mathcal{S}$. We also consider set of matrices $\{\mathbf{X}_i\}_{i=1}^n$ with the entries $\left\{x_{jk}^i\right\}_{j,k}$ drawn from a normal distribution with zero mean and prescribed variance $\mathcal{S}^2$. Now to form the new population of covariance matrices $\{\mathbf{S}_i\}_{i=1}^n$ with respect to the nominal covariance matrix $\mathcal{S}$ we consider the following model:

$$\mathbf{S}_i = (\mathcal{W} + \mathbf{X}_i)(\mathcal{W} + \mathbf{X}_i)^H \qquad (14)$$

In order to take to account the signature of randomness in producing samples using model (14) we use Eq.(15) to measure the discrepancy in several simulation runs; This approach is known as Monte Carlo simulation (MacKay, 1998). In this method for the fixed covariance matrix $\mathcal{S}$ one can generate the population set $\{\mathbf{S}_i\}_{i=1}^n$ for N times. Each time the Fréchet mean of the population will be evaluated, $\tilde{n} = 1,2,...,N$. Finally, the criterion which is known as *Root Mean Square Error* or (RMSE) is formed as follows :

$$RMSE_{d_F} = \sqrt{\frac{1}{N} \sum_{\tilde{n}=1}^{N} {d_F}^2 \left(\mathcal{S}, \hat{\mathcal{S}}_{\tilde{n}}\right)} \qquad (15)$$

Where metric d_F is defined as :

$$d_F\left(\mathcal{S}, \hat{\mathcal{S}}\right) = \left\| \mathcal{S} - \hat{\mathcal{S}} \right\|_2 \qquad (16)$$

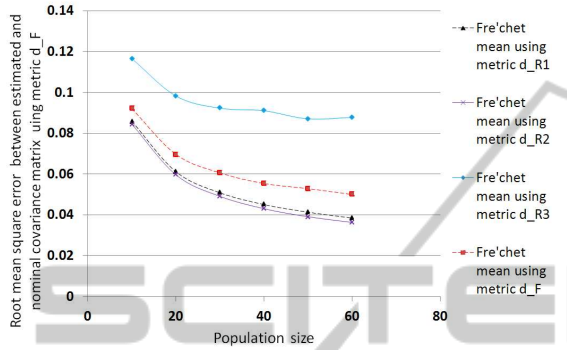
in which for matrix $\mathbf{A}$, $\|\mathbf{A}\|_2 = \sqrt{\text{Tr}\mathbf{A}\mathbf{A}^H}$.

First we consider the model (14) to demonstrate the performance of the Fréchet mean of Riemannian distances. For this reason we consider a Covariance matrix $\mathcal{S}_{3\times3}$. The eigenvalues of the covariance matrix is $\mathcal{L} = \text{diag}\,[1, 0.3573, 0.065]$. As far as the model (14) is concerned the Cholesky factor of the covariance matrix is considered. The additive random noise matrix $\{\mathbf{X}_i\}_{i=1}^n$ has independent and identically distributed (i.i.d) entries come from Gaussian distribution with zero mean:

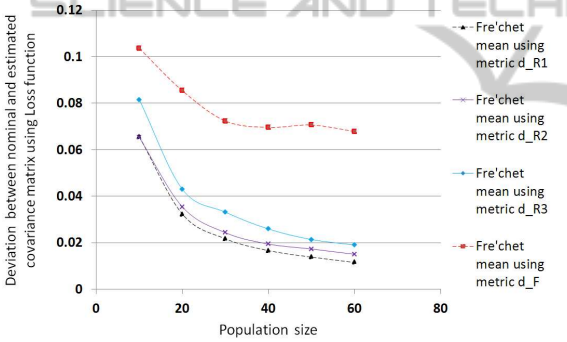
$$E\left(x_{j,k}^i\right) = 0 \quad j,k = 1,2,3 \ , i = 1,2,..,n \qquad (17)$$

where E denotes the expected value of the random variable. The standard deviation of the entries of random noise is 0.09 in this experiment.

The population size of the covariance matrices, $\{\mathbf{S}_i\}_{i=1}^{n}$, varies between 10 to 60 in step size 10. In order to take to account the signature of randomness of the additive Gaussian noise matrix $\mathbf{X}_i$'s in model (14), for each population we perform the Monte-Carlo simulation 2000 times and obtain the resulting error between $\hat{\mathcal{S}}$ and the nominal covariance matrix using loss function, $RMSE_{d_F}$. The results are shown in Figure 1.



(a)



(b)

Figure 1: (a): Error is measured using metric d_F.(b): Error is measured using Loss function.

4.2 Classification Based on the Distance to the Center of Mass

So far we have mathematically developed the concept of mean for group of positive definite Hermitian matrices on manifold $\mathcal{M}$ from the distance point of view. Moreover, we have seen that depending on model and the criterion of measuring the closeness of each estimator to the nominal covariance matrix, Fréchet means of Riemannian distances are better estimators.

The concept of Fréchet mean can be utilized in distance based detection and classification on manifold $\mathcal{M}$ (Pigoli et al., 2014),(Barachant et al., 2010). For this purpose suppose that we have a set of covariance matrices $\{\mathbf{S}_{ik}\}_{i=1}^{n_k}$ where k represents the label of each class and n_k denotes the number of covariance matrices within k^{th} class. For each class k the Fréchet mean of the class, depending on type of metric, can be obtained as representative of each class. For the unknown observation its covariance matrix is formed and considered as the unknown feature. The observation is assigned to the class which has minimum distance to the Fréchet mean of the class. This method can be recapitulated in form of the following algorithm.

Algorithm 2: Distance to the center of mass algorithm.

1. Input: the given known classes $1,2,3,,...,k$ and set of covariance matrices $\{\mathbf{S}_{ik}\}_{i=1}^{n_k}$ within each class.
2. For each class k compute $\hat{\mathcal{S}}_{ik}$ as the Fréchet mean of $\{\mathbf{S}_{ik}\}_{i=1}^{n_k}$.
3. For the covariance matrix $\mathbf{S}$ of unknown observation compute
$$\hat{k} = \arg\min_{k} d(\mathbf{S}, \mathcal{S}_k). \quad (18)$$
4. The covariance matrix $\mathbf{S}$ corresponding to the unknown observation in step 3 will be assigned to class $\hat{k}$.

In order to inspect and evaluate the Algorithm 1 we perform it on the simulated data set. For this purpose we consider three classes $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$ consisting of samples $\{\mathbf{x}_1(i)\}_{i=1}^{10000}$, $\{\mathbf{x}_2(i)\}_{i=1}^{10000}$ and $\{\mathbf{x}_3(i)\}_{i=1}^{10000}$ drawn from the normal distribution with zero mean and covariance matrices $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ respectively. At the same time Gaussian random noise with mean zero and standard deviation $\mathcal{S}$ is added to the samples of both classes. Then we split each class to the half for train and test purpose and perform two fold cross validation .

At training step we consider training sets $\mathcal{C}_{1train}$, $\mathcal{C}_{2train}$ and $\mathcal{C}_{3train}$. From each training set $\mathcal{C}_{j_{train}}$, $j = 1,2,3$; we form a sequence of $\{\mathbf{X}_{k,j}\}$ of observations $k = 1,2,..,20$. Each observation $\{\mathbf{X}_{k,j}\}$ has 40 samples which can be shown as $\left[\mathbf{x}_{jk}(1),\mathbf{x}_{jk}(2),...,\mathbf{x}_{jk}(40)\right]^T$.

The Frechét mean of the covariance matrices $\{\mathbf{S}_{k,j}\}$ of the observation $\{\mathbf{X}_{k,j}\}$ are obtained using metrics d_{R_1},d_{R_2}, d_{R_3} and d_F respectively. The method of distance to the center of mass is performed to classify the new observation $\mathbf{X}_{test}$ according to its observed covariance matrix $\mathbf{S}_{test}$.

From (Johnson and Wichern, 2002) it has been known that when we have sample of observations from a p-variate normal distribution with zero mean and covariance matrix $\mathcal{S}$ then $N\bar{\mathbf{x}}\mathcal{S}^{-1}\bar{\mathbf{x}}^T$ has chi-square distribution with p degrees of freedom where $\bar{\mathbf{x}}$ is the sample mean vector of size $1 \times p$ for the observation $\mathbf{X}$ of size $N \times p$; which is taken over columns

of $\mathbf{X}$ and N is the sample size. When the sample size is fairly large we can replace $\mathcal{S}$ with $\hat{\mathcal{S}}$ (Anderson, 1954). As a result, a new observation $\mathbf{X}_{test}$ is classified to class j whenever:

$$\chi^2_p(1-\alpha/2) \leq N\bar{\mathbf{x}}_{test}\hat{\mathcal{S}}_j^{-1}\bar{\mathbf{x}}_{test}^T \leq \chi^2_p(\alpha/2) \qquad (19)$$

where $\chi^2_p(\alpha)$ is given by:

$$P\left(\chi^2_p > \chi^2_p(\alpha)\right) = \alpha$$

In Eq.(19) the significant level is set to be $\alpha = 0.05$. At the same time we also compare the result of distance to the center of mass in classification with the result of Eq.(19).

Table 1: Probability of correct classification within three classes C_1, C_2 and C_3 in comparison to the resulting classifier using Eq.(19).

	Accuracy Class 1 (%)	Accuracy of Class 2 (%)	Accuracy for Class 3 (%)
d_{R3}	0.92	0.86	0.95
d_F	0.83	0.39	0.62
d_{R1}	0.91	0.51	0.78
d_{R2}	0.92	0.51	0.82
Eq.(19)	0.80	0.60	0.72

In addition we applied our technique to the classification of human breast cancer cells undergoing treatment of different drugs. As explained before our technique is based on classification in clusters based on the covariance estimate distances rather than center of the class (corresponding to the mean of the data point cloud). The original data set consisted of 11 different labels corresponding to 11 different treatments. Each label consisted of 382 wells which were imaged using Perkin Elmar high content imaging system. Our preliminary results indicate that our average classification error is approximately 13% when the half of the cells are used for training. The preliminary comparison with commonly used clustering techniques based on the sample average (mean) indicate that our performance is significantly better (5%) however it may be due to the large training set.

5 CONCLUSIONS

In this paper we proposed a new technique to for estimating positive definite matrices in the presence of uncertainty. Unlike commonly used techniques our method uses Frechet mean which implicitly accounts for the positive definite structure of the covariance matrix which is ignored in commonly used estimators which do not exploit geometric constraint given by positive definite property. We demonstrate the calculation of the proposed mean using three different distance measures which may be better choice in different applications. We demonstrated the applicability and performance of our techniques on a simulated data set and established that in the preliminary analysis the results look promising for high content cell imaging classification problem.

REFERENCES

Absil, P.-A., Mahony, R., and Sepulchre, R. (2009). *Optimization algorithms on matrix manifolds*. Princeton University Press.

Anderson, T. W. (1954). *An Introduction To Multivariate Statistical Analysis*. Wiley Eastern Private Limited; New Delhi.

Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2010). Riemannian geometry applied to BCI classification. In *Latent Variable Analysis and Signal Separation*, pages 629–636. Springer.

Barbaresco, F. (2008). Innovative tools for radar signal processing based on Cartans geometry of SPD matrices & information geometry. *Radar Conference, 2008. RADAR'08. IEEE*, pages 1–6.

Crosilla, F. and Beinat, A. (2002). Use of generalised procrustes analysis for the photogrammetric block adjustment by independent models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 56(3):195–209.

Jeuris, B., Vandebril, R., and Vandereycken, B. (2012). A survey and comparison of contemporary algorithms for computing the matrix geometric mean. *Electronic Transactions on Numerical Analysis*, 39:379–402.

Johnson, R. A. and Wichern, D. W. (2002). *Applied multivariate statistical analysis*, volume 5. Prentice Hall Upper Saddle River, NJ.

Li, Y. and Wong, K. M. (2013). Riemannian distances for EEG signal classification by power spectral density. *IEEE journal of selected selected topics in signal processing*.

MacKay, D. J. (1998). Introduction to Monte Carlo methods. In *Learning in graphical models*, pages 175–204. Springer.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate analysis*. Academic Press.

Moakher, M. (2005). A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3):735–747.

Pigoli, D., Aston, J. A., Dryden, I. L., and Secchi, P. (2014). Distances and inference for covariance operators. *Biometrika*.