HAL open science

Riemannian classification of EEG signals with missing values

Alexandre Hippert-Ferrer, Ammar Mian, Florent Bouchard, Frédéric Pascal

To cite this version:

Alexandre Hippert-Ferrer, Ammar Mian, Florent Bouchard, Frédéric Pascal. Riemannian classification of EEG signals with missing values. 30th European Signal Processing Conference (EUSIPCO 2022), Aug 2022, Belgrade, Serbia. 10.23919/eusipco55093.2022.9909703 . hal-03856816

HAL Id: hal-03856816

https://hal.science/hal-03856816v1

Submitted on 17 Nov 2022

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.

RIEMANNIAN CLASSIFICATION OF EEG SIGNALS WITH MISSING VALUES

A. Hippert-Ferrer[1], *A. Mian*[2], *F. Bouchard*[1], *F. Pascal*[1]

[1] Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France
[2] LISTIC, Université Savoie Mont Blanc

ABSTRACT

This paper proposes a strategy to handle missing data for the classification of electroencephalograms using covariance matrices. It relies on the observed-data likelihood within an expectation-maximization algorithm. This approach is compared to two existing state-of-the-art methods: *(i)* covariance matrices computed with imputed data; *(ii)* Riemannian averages of partially observed covariance matrix. All approaches are combined with the minimum distance to Riemannian mean classifier and applied to a classification task of two widely known paradigms of brain-computer interfaces. In addition to be applicable for a wider range of missing data scenarios, the proposed strategy generally performs better than other methods on the considered real EEG data.

***Index Terms*—** Missing data, Electroencephalography, Brain-computer interfaces, Covariance estimation, Riemannian geometry.

I. INTRODUCTION

Electroencephalography (EEG) is a non-invasive neuroimaging modality pioneered by Hans Berger [1], which consists in recording the electrical activity of the brain by placing electrodes on the scalp. The low cost, simplicity and high temporal resolution (it captures well the brain activity dynamics) of EEG have made its popularity and allowed for its use in various applications. EEG is indeed widely used in brain research for instance to study sleep or epilepsy. It is also central for brain-computer interfaces (BCI), where the subject interacts with a computer through brain signals. These can for example be employed to control exoskeleton [2] or help mechanical ventilation [3].

This paper focuses on BCI and more specifically on one of the various possible paradigms: event related potentials (ERP), which are specific time-locked responses of the brain to some stimuli, such as flashes [4]. Given the brain signals, the goal is to determine if the subject was exposed to the stimulus, *i.e.*, find out if the specific time-locked response is in the EEG recording. Thus, BCI boils down to a classification task of brain signals at hand. Various methods have specifically been designed to achieve this; see [5] for a recent review. In this work, we consider one of the most efficient approach, called minimum distance to Riemannian mean (MDRM) [6], which exploits covariance matrices of the data along with their natural Riemannian geometry.

BCI classification methods have shown good performance on a number of classical datasets; see *e.g.*, [7]. However, these datasets have been recorded in laboratories, which are well controlled environments, and data have been curated by hand to ensure the quality of the recordings. In real life scenarios, obtaining such idealistic data seems unrealistic. One can expect the data at hand to contain faulty measurements due for instance to impedence change, electrode disconnections [8] or artifacts coming from movements of the subject, blinks, *etc.* These can lead to missing data in EEG recordings as one may decide to discard it from the data. When it comes to classification tasks, missing data impedes the exploitation of complementary information in non-missing channels. Considering this, it is of primary importance to develop an appropriate procedure to classify incomplete EEG signals.

To achieve this, several strategies can be employed. A first possibility is to exploit some missing data imputation technique in order to retrieve a complete EEG recording. It has for instance been considered for EEG in [9] with an algorithm relying on spherical spline interpolation. Imputation can also be performed through the K-nearest neighbors (KNN) algorithm [10]. Another possibility is to directly work in the covariance space. When a channel is missing, an imputation of symmetric positive definite (SPD) matrices with uncorrelated white noise has been proposed in [11]. In a similar vein, the authors of [12] propose to estimate the Riemannian mean of partially observed covariance matrices with missing variables (channels) to later perform a classification task with the MDRM. When the data contains missing samples[1], it is possible to estimate the covariance matrix of incomplete data using an expectation-maximization (EM) algorithm [13]. More recently, a neural network approach was designed to interpolate missing electrodes in EEG [14].

In this paper, we propose to adapt the state-of-the-art EEG classification pipeline [6] to handle missing data by exploiting an EM algorithm to compute covariance matri-

[1]To appreciate the different nature of works handling missing samples in the data and those treating on missing variables in the data and/or the covariance matrix, see Table 1 in [12].

ces. Compared to two state-of-the-art approaches – missing data imputation using KNN [10] and Riemannian averaging over incomplete covariance matrices [12] – our method is applicable in more missing data scenarios. Furthermore, experiments conducted onto real ERP and motor imagery (MI) EEG data tend to show that our method is competitive in terms of performance as compared to others.

II. EEG CLASSIFICATION FOR COMPLETE DATA

To classify EEG data, rather than directly using the raw data, one usually exploits their covariance matrices as features [6]. The classification of these covariance matrices is then performed by leveraging their Riemannian geometry: a well-adapted distance is used to compute centers of mass of the different classes and assigning the class to unknown recordings.

II-A. Covariance matrix estimation

A common hypothesis for EEG signals is to assume a multivariate normal distribution [15]. Let $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n] \in \mathbb{R}^{p\times n}$ be a recorded EEG trial containing p-dimensional real vectors, where p is the number of electrodes and n the time samples. $\boldsymbol{X}$ follows a zero-mean Gaussian distribution with covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, denoted $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$. When the data is complete, the Maximum Likehood Estimator (MLE) is the well-known Sample Covariance Matrix (SCM):

$$\boldsymbol{\Sigma} = \frac{1}{n}\boldsymbol{X}\boldsymbol{X}^\top \tag{1}$$

II-B. Minimum distance to Riemannian mean

The idea consists in leveraging a non-Euclidean distance based on the geometry of the space of SPD matrices (thus covariance matrices). In peculiar, theoretical studies have shown benefits to use the so-called affine invariant distance [16]. Given two covariances matrices $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$, its expression is given by:

$$\delta_{\mathcal{AF}}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \left\| \log\left(\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1/2}\right)\right\|_{\mathrm{F}}, \tag{2}$$

where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm and $\log$ is the matrix logarithm. Then, given a classification problem of Z classes, with training data $\{(\boldsymbol{\Sigma}_\ell, y_\ell) : 1 \leq \ell \leq L\}$, the training phase consists in computing the mean of each class $\{\overline{\boldsymbol{\Sigma}}^{(z)} : z \in \{1, \ldots, Z\}\}$ according to the affine-invariant distance by using the iterative algorithm of [17]. A predicted label y is then obtained on a new sample $\boldsymbol{\Sigma}$ by assigning it to the closest mean:

$$y = \underset{z\in\{1,\ldots,Z\}}{\arg\min}\ \delta_{\mathcal{AF}}(\boldsymbol{\Sigma}, \overline{\boldsymbol{\Sigma}}^{(z)}) \tag{3}$$

III. EXISTING STRATEGIES TO HANDLE MISSING DATA FOR EEG CLASSIFICATION

III-A. Data imputation: KNN algorithm

A first possibility to handle missing data is to impute them, which then allows to compute their covariance matrix according to (1) – and thus apply the MDRM classification method. We focus here on K-nearest neighbors (KNN) interpolation [10]. If the jth electrode of $\boldsymbol{x}_i$ is missing, KNN finds its K-closest neighbors in the training set according to an Euclidean distance $d\big(\boldsymbol{x}_i, \{\boldsymbol{x}_q\}_{q=1,q\neq i}^n\big)$ compatible with missing data, called the heterogeneous Euclidean-overlap metric (HEOM) [18]. If $\{\boldsymbol{v}_k\}_{k=1}^K$ represents the set of K-selected neighbors, then the imputed value $\hat{x}_{ij}$ is

$$\hat{x}_{ij} = \frac{1}{K}\sum_{k=1}^{K} w_k v_{kj}, \quad w_k = \frac{1}{d(\boldsymbol{x}_i, \boldsymbol{v}_k)^2}, \quad \sum_{k=1}^{K} w_k = 1 \tag{4}$$

where w_k denotes the corresponding weight of the k-th nearest neighbor $\boldsymbol{v}_k$ and $d(\boldsymbol{x}_i, \boldsymbol{v}_k)^2$ is the squared HEOM. The choice of K can be determined by cross-validation on observed data only, which can be challenging with high missing data ratio.

III-B. Masked Minimum to Riemannian mean

In [12], authors propose an implementation of the MDM algorithm for incomplete data. Such design requires to construct a mask to extract[2] the observed data from an incomplete input $\boldsymbol{X} \in \mathbb{R}^{p\times n}$. A mask $\boldsymbol{M} \in \mathbb{R}^{p\times(p-r)}$ is defined as the identity matrix with columns of the r missing variables removed. If $\boldsymbol{X}$ is incomplete, a complete $(p-r)\times n$ submatrix $\check{\boldsymbol{X}}$ can be extracted using $\boldsymbol{M}^\top\boldsymbol{X}$. The covariance matrix becomes

$$\check{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{M}^\top\boldsymbol{\Sigma}\boldsymbol{M} \tag{5}$$

where $\boldsymbol{\Sigma}$ is the complete covariance matrix as defined in (1). The next step is to consider the set of $\{\boldsymbol{\Sigma}_i\}_{i=1}^n$ incomplete SPD matrices, where each $\boldsymbol{\Sigma}_i$ has r_i missing variables and an associated mask $\boldsymbol{M}_i$. For each class $z \in \{1, \ldots, Z\}$, the masked Riemannian mean $\bar{\boldsymbol{\Sigma}} \in \mathcal{S}_{++}^p$ is defined as:

$$\overline{\boldsymbol{\Sigma}}^{(z)} = \underset{\boldsymbol{\Sigma}\in\mathcal{S}_{++}^p}{\arg\min}\ \frac{1}{2}\sum_{i=1}^{n} w_i \delta_{\mathcal{AF}}^2(\boldsymbol{M}_i^\top\boldsymbol{\Sigma}_i\boldsymbol{M}_i, \boldsymbol{M}_i^\top\boldsymbol{\Sigma}\boldsymbol{M}_i) \tag{6}$$

Finally, the label prediction is performed as in (3) with the suited mask added.

IV. AN EM ALGORITHM FOR COVARIANCE ESTIMATION WITH MISSING DATA

A powerful strategy to deal with missing values is to directly estimate the covariance feature without imputation. In this case, one can rely on maximum likelihood estimation using the EM algorithm developed in [19]. Here, the unknown parameter to estimate is simply $\boldsymbol{\theta} = \{\boldsymbol{\Sigma}\}$. In the following EM algorithm, $\boldsymbol{x}_i$ is transformed into $\widetilde{\boldsymbol{x}}_i$ such that $\widetilde{\boldsymbol{x}}_i = \boldsymbol{P}_i\boldsymbol{x}_i = \begin{pmatrix}\boldsymbol{x}_i^{\mathrm{o}} & \boldsymbol{x}_i^{\mathrm{m}}\end{pmatrix}^\top$, where $\boldsymbol{P}_i \in \mathbb{R}^{p\times p}$ is a permutation matrix, $\boldsymbol{x}_i^{\mathrm{o}}$ and $\boldsymbol{x}_i^{\mathrm{m}}$ denote the observed and

[2]Such matrix is often called selection matrix.

missing elements of $\boldsymbol{x}_i$, respectively. The covariance matrix of $\widetilde{\boldsymbol{x}}_i$ is

$$\widetilde{\boldsymbol{\Sigma}}_i = \begin{pmatrix} \widetilde{\boldsymbol{\Sigma}}_{i,\text{oo}} & \widetilde{\boldsymbol{\Sigma}}_{i,\text{mo}} \\ \widetilde{\boldsymbol{\Sigma}}_{i,\text{om}} & \widetilde{\boldsymbol{\Sigma}}_{i,\text{mm}} \end{pmatrix} = \boldsymbol{P}_i \boldsymbol{\Sigma} \boldsymbol{P}_i^\top \tag{7}$$

where $\widetilde{\boldsymbol{\Sigma}}_{i,\text{mm}}$, $\widetilde{\boldsymbol{\Sigma}}_{i,\text{mo}}$, $\widetilde{\boldsymbol{\Sigma}}_{i,\text{oo}}$ are the block CM of $\boldsymbol{x}_i^\text{m}$, of $\boldsymbol{x}_i^\text{m}$ and $\boldsymbol{x}_i^\text{o}$, and of $\boldsymbol{x}_i^\text{o}$. Replacing $\boldsymbol{x}_i$ by its permuted version $\widetilde{\boldsymbol{x}}_i$, the *complete* log-likelihood for the Gaussian distribution is expressed as:

$$\mathcal{L}_c(\boldsymbol{\theta}|\boldsymbol{x}) \propto -n \log |\boldsymbol{\Sigma}| - \sum_{i=1}^n \begin{pmatrix} \boldsymbol{x}_i^\text{o} \\ \boldsymbol{x}_i^\text{m} \end{pmatrix}^\top \widetilde{\boldsymbol{\Sigma}}_i^{-1} \begin{pmatrix} \boldsymbol{x}_i^\text{o} \\ \boldsymbol{x}_i^\text{m} \end{pmatrix} \tag{8}$$

The EM algorithm, which is given in Algorithm 1, can now be formulated.

Initialization. At step $t = 0$ of the algorithm, initialize the covariance matrix with the SCM computed from fully observed $\boldsymbol{x}_i$ (*i.e.*, no missing values), denoted $\boldsymbol{\Sigma}_{\text{SCM-obs}}$.

E-step. Compute the expectation of the missing data conditioned by the observed data and the estimated covariance at the t-th iteration:

$$\begin{aligned} Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= \mathbb{E}_{\boldsymbol{x}_i^\text{m}|\boldsymbol{x}_i^\text{o},\boldsymbol{\theta}^{(t)}}\left[\mathcal{L}_c(\boldsymbol{\theta}|\boldsymbol{x}_i^\text{o}, \boldsymbol{x}_i^\text{m})\right] \\ &= \text{tr}(\boldsymbol{B}_i^{(t)} \widetilde{\boldsymbol{\Sigma}}_i^{-1}) \end{aligned} \tag{9}$$

with $\boldsymbol{B}_i^{(t)} = \begin{pmatrix} \boldsymbol{x}_i^\text{o} \boldsymbol{x}_i^{\text{o}^\top} & \mathbb{E}_{\boldsymbol{x}_i^\text{m}|\boldsymbol{x}_i^\text{o},\boldsymbol{\theta}^{(t)}}[\boldsymbol{x}_i^\text{o} \boldsymbol{x}_i^{\text{m}^\top}] \\ \mathbb{E}_{\boldsymbol{x}_i^\text{m}|\boldsymbol{x}_i^\text{o},\boldsymbol{\theta}^{(t)}}[\boldsymbol{x}_i^\text{m} \boldsymbol{x}_i^{\text{o}^\top}] & \mathbb{E}_{\boldsymbol{x}_i^\text{m}|\boldsymbol{x}_i^\text{o},\boldsymbol{\theta}^{(t)}}[\boldsymbol{x}_i^\text{m} \boldsymbol{x}_i^{\text{m}^\top}] \end{pmatrix}$. The only expectation to compute are $\mathbb{E}_{\boldsymbol{x}_i^\text{m}|\boldsymbol{x}_i^\text{o},\boldsymbol{\theta}^{(t)}}[\boldsymbol{x}_i^\text{m}]$ and $\mathbb{E}_{\boldsymbol{x}_i^\text{m}|\boldsymbol{x}_i^\text{o},\boldsymbol{\theta}^{(t)}}[\boldsymbol{x}_i^\text{m} \boldsymbol{x}_i^{\text{m}^\top}]$, which are the expectation of the sufficient statistics $\boldsymbol{x}_i^\text{m}$ and $\boldsymbol{x}_i^\text{m} \boldsymbol{x}_i^{\text{m}^\top}$ (denoted $\mathbb{E}\big[\boldsymbol{x}_i^\text{m}\big]$ and $\mathbb{E}\big[\boldsymbol{x}_i^\text{m} \boldsymbol{x}_i^{\text{m}^\top}\big]$ hereafter). A classical result in conditional distributions [20] gives the following expressions:

$$\mathbb{E}\big[\boldsymbol{x}_i^\text{m}\big] = \widetilde{\boldsymbol{\Sigma}}_{i,\text{mo}} \widetilde{\boldsymbol{\Sigma}}_{i,\text{oo}}^{-1} \widetilde{\boldsymbol{x}}_i^\text{o} \tag{10}$$

$$\mathbb{E}\big[\boldsymbol{x}_i^\text{m} \boldsymbol{x}_i^{\text{m}^\top}\big] = \widetilde{\boldsymbol{\Sigma}}_{i,\text{mm}} - \widetilde{\boldsymbol{\Sigma}}_{i,\text{mo}} \widetilde{\boldsymbol{\Sigma}}_{i,\text{oo}}^{-1} \widetilde{\boldsymbol{\Sigma}}_{i,\text{om}} + \mathbb{E}\big[\boldsymbol{x}_i^\text{m}\big] \mathbb{E}\big[\boldsymbol{x}_i^\text{m}\big]^\top \tag{11}$$

M-step. Resolve the following maximization problem:

$$\boldsymbol{\Sigma}^{(t+1)} = \underset{\boldsymbol{\Sigma} \in \mathcal{S}_{++}^p}{\arg\max} \sum_{i=1}^n Q_i(\boldsymbol{\Sigma}|\boldsymbol{\Sigma}^{(t)}) \tag{12}$$

where $\mathcal{S}_{++}^p$ is the set of symmetric positive definite (SPD) matrices. Solving (12) requires to solve $\frac{\partial Q_i}{\partial \boldsymbol{\Sigma}} = 0$. Fortunately in this case, the following closed-form expression can be found with simple derivation calculus:

$$\boldsymbol{\Sigma}^{(t+1)} = \frac{1}{n} \sum_{i=1}^n \boldsymbol{P}_i \boldsymbol{B}_i^{(t)\top} \boldsymbol{P}_i^\top \tag{13}$$

Algorithm 1 Estimation of $\boldsymbol{\Sigma}$ with the EM algorithm

Require: $\{\widetilde{\boldsymbol{x}}_i\}_{i=1}^n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}), \{\boldsymbol{P}_i\}_{i=1}^n$
Ensure: $\widehat{\boldsymbol{\Sigma}}$
1: Initialization: $\boldsymbol{\Sigma}^{(0)} = \widehat{\boldsymbol{\Sigma}}_{\text{SCM-obs}}$
2: **repeat**
3: Compute $\boldsymbol{B}_i^{(t)} = \begin{pmatrix} \boldsymbol{x}_i^\text{o} \boldsymbol{x}_i^{o\top} & \mathbb{E}[\boldsymbol{x}_i^\text{o} \boldsymbol{x}_i^{\text{m}^\top}] \\ \mathbb{E}[\boldsymbol{x}_i^\text{m} \boldsymbol{x}_i^{\text{o}^\top}] & \mathbb{E}[\boldsymbol{x}_i^\text{m} \boldsymbol{x}_i^{\text{m}\top}] \end{pmatrix}$
4: Compute $\boldsymbol{\Sigma}^{(t+1)} = \frac{1}{n} \sum_{i=1}^n \boldsymbol{P}_i \boldsymbol{B}_i^{(t)\top} \boldsymbol{P}_i^\top$
5: $t \leftarrow t + 1$
6: **until** $||\boldsymbol{\Sigma}^{(t+1)} - \boldsymbol{\Sigma}^{(t)}||_F^2$ converges

V. NUMERICAL EXPERIMENTS

V-A. Data

The proposed procedures are assessed through a detection task of EEG recordings of two distinct BCI experiments[3]:

i) *P300 Event Related Potentials* triggered from target stimuli consisting of flashes [4]. This dataset consists in 3 recording sessions of EEG potentials on 10 subjects using 16 electrodes[4]. The signals are band-pass filtered between 1 and 20Hz, downsampled to 128Hz and epoched into 1728 trials: one trial out of six corresponds to the flash stimulus (class 1) while the rest correspond to the absence of target (class 0). Before computing covariance matrices, raw data are augmented by concatenating the prototyped ERP response as prescribed in [22]. This leads for each subjects to a dataset of $L = 1728$ covariance matrices as in (1) with $p = 16$ and $n = 103$.

ii) *Motor Imagery*: this dataset [23] consists of EEG data from 9 subjects recorded using 22 electrodes. This BCI paradigm consists of four different motor imagery tasks, namely the imagination of movement of the left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4). Two sessions were recorded for each subject, with a total of 288 trials per session. This leads for each subjects to a dataset of $L = 576$ covariance matrices with $p = 22$ and $n = 1001$.

V-B. Missing data scenarios

We consider two realistic scenarios of incomplete EEG signals: *electrode popping* (s1) and *eye blinking* (s2). As illustrated in Fig.1, electrode popping yields one or multiple missing electrodes during signal acquisition. Eye blinking creates an undesired artifact over a short time period (typically ∼200 ms) over a group of electrodes. Such artifacts are then removed, which creates missing data. As shown in Table I, compared to the other methods considered, our proposed strategy is the only one working in both cases.

[3]Both data sets are accessed through the `moabb` database [7].

[4]Fz, FCz, Cz, CPz, Pz, Oz, F3, F4, C3, C4, CP3, CP4, P3, P4, PO7, PO8. Please refer to [21] for an extensive description of the dataset.

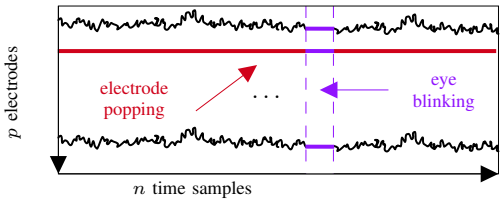

Fig. 1: Illustration of electrode popping (red) and eye blinking (purple).

	KNN [10]	Masked mean [12]	EM [19]
Electrode popping (s1)	×	✓	✓
Eye blinking (s2)	✓	×	✓

Table I: Applicability (✓: applicable; ×: not applicable) of the proposed methods for different missingness scenarios.

For scenario s1, two electrodes are entirely removed. For scenario s2, three groups of electrodes (11 for P300 and 12 for MI) during three time intervals are removed. In both cases, removals are done on an increasing number of trials to appreciate the effect on classification accuracy.

V-C. Results

To compare the accuracy of the classification of P300 and Motor Imagery datasets, the MRDM classifier as described in section II is used. The following strategies are considered to estimate the covariance matrices: **SCM**: the SCM of *complete* data, *i.e.*, without missing values; **KNN + SCM**: KNN imputer implemented from the `scikit-learn` package [24] applied to incomplete data with a number of $k = 5$ neighbors. **Masked-SCM**: MDRM classifier using a masked version of the Riemannian mean [12]. **EM-SCM**: hybrid strategy where the covariance matrix is estimated through Algorithm 1 and fed into the MDRM classifier.

Performance is evaluated using the classical training-test paradigm with stratified K-folding ($K = 5$). Algorithms are calibrated on data recorded during the training phase, and applied on data recorded during the test phase.

Electrode popping experiment. Accuracy results for this experiment are presented in Figure 2. The best performance is achieved on complete data (SCM), which serves as a baseline. EM-based classification achieves equal or higher scores than the masked MDRM for both datasets. On these data, our proposed strategy thus appears more advantageous than the masked version of the Riemannian mean.

Eye blinking experiment. Results for this experiment are presented in Figure 3. For the motor imagery dataset, KNN imputation does not perform as well as the EM-based classification. For the P300 dataset, KNN imputation shows higher accuracies as compared to the EM algorithm when the amount of missing data increases.

VI. CONCLUSION

In this paper, a new strategy to handle missing data in the context of EEG classification is proposed. It relies on an EM algorithm to directly estimate covariance matrices in the presence of missing data. As compared to two other state-of-the-art strategies – based on KNN imputation and masked Riemannian means – our proposed method works with a wider range of missing data scenarios. In addition, experiments conducted on real EEG data tends to show that our method is also competitve in terms of accuracy results.

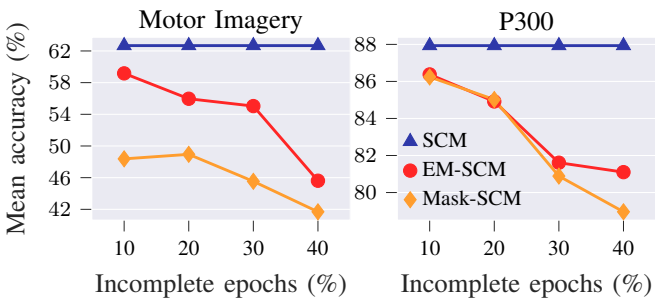

Fig. 2: Electrode popping experiment: classification mean accuracy as function of the incomplete epochs ratio.

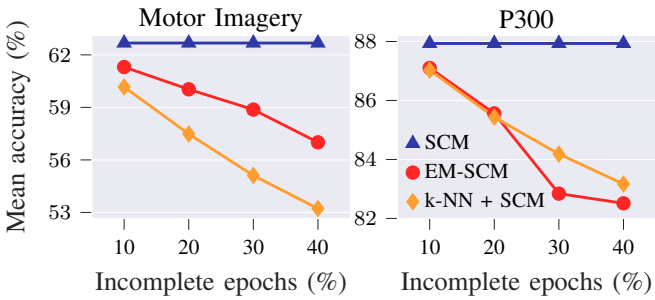

Fig. 3: Eye blinking experiment: classification mean accuracy as function of the incomplete epochs ratio.

VII. REFERENCES

[1] H. Berger, "Über das elektrenkephalogramm des menschen," *Archiv für psychiatrie und nervenkrankheiten*, vol. 87, no. 1, pp. 527–570, 1929.

[2] E. K. Kalunga, S. Chevallier, Q. Barthélemy, K. Djouani, E. Monacelli, and Y. Hamam, "Online ssvep-based bci using riemannian geometry," *Neurocomputing*, vol. 191, pp. 55–68, 2016.

[3] S. Chevallier, G. Bao, M. Hammami, F. Marlats, L. Mayaud, D. Annane, F. Lofaso, and E. Azabou, "Brain-machine interface for mechanical ventilation using respiratory-related evoked potential," in *International Conference on Artificial Neural Networks*, 2018, pp. 662–671.

[4] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer

interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[5] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update," *Journal of neural engineering*, vol. 15, no. 3, 2018.

[6] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain–computer interface classification by Riemannian geometry," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 920–928, 2011.

[7] V. Jayaram and A. Barachant, "MOABB: trustworthy algorithm benchmarking for BCIs," *Journal of Neural Engineering*, vol. 15, no. 6, p. 066011, 2018.

[8] N. Bahador, K. Erikson, J. Laurila, J. Koskenkari, T. Ala-Kokko, and J. Kortelainen, "A correlation-driven mapping for deep learning application in detecting artifacts within the EEG," *Journal of Neural Engineering*, vol. 17, no. 5, p. 056018, 2020.

[9] F. Perrin, J. Pernier, O. Bertrand, and J. Echallier, "Spherical splines for scalp potential and current density mapping," *Electroencephalography and Clinical Neurophysiology*, vol. 72, no. 2, pp. 184–187, 1989.

[10] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. B. Altman, "Missing value estimation methods for DNA microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, 2001.

[11] P. L. C. Rodrigues, M. Congedo, and C. Jutten, "A data imputation method for matrices in the symmetric positive definite manifold," in *XXVIIème colloque GRETSI (GRETSI 2019)*, 2019.

[12] F. Yger, S. Chevallier, Q. Barthélemy, and S. Sra, "Geodesically-convex optimization for averaging partially observed covariance matrices," in *Proceedings of The 12th Asian Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 129, 2020, pp. 417–432.

[13] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[14] S. Saba-Sadiya, T. Alhanai, T. Liu, and M. M. Ghassemi, "EEG channel interpolation using deep encoder-decoder networks," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2020, pp. 2432–2439.

[15] M. Congedo, C. Gouy-Pailler, and C. Jutten, "On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics," *Clinical Neurophysiology*, vol. 119, no. 12, pp. 2677–2686, 2008.

[16] X. Pennec, "3 - manifold-valued image processing with SPD matrices," in *Riemannian Geometric Statistics in Medical Image Analysis*. Academic Press, 2020, pp. 75–134.

[17] B. Jeuris, R. Vandebril, and B. Vandereycken, "A survey and comparison of contemporary algorithms for computing the matrix geometric mean," *Electronic Transactions on Numerical Analysis*, vol. 39, pp. 379–402, 2012.

[18] P. J. García-Laencina, J.-L. Sancho-Gómez, A. R. Figueiras-Vidal, and M. Verleysen, "K nearest neighbours with mutual information for simultaneous classification and missing data imputation," *Neurocomputing*, vol. 72, no. 7, pp. 1483–1493, 2009.

[19] A. Hippert-Ferrer, M. El Korso, A. Breloy, and G. Ginolhac, "Robust low-rank covariance matrix estimation with a general pattern of missing values," *Signal Processing*, vol. 195, p. 108460, 2022.

[20] T. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 1965.

[21] P. Aricò, F. Aloise, F. Schettini, S. Salinari, D. Mattia, and F. Cincotti, "Influence of P300 latency jitter on event related potential-based brain–computer interface performance," *Journal of Neural Engineering*, vol. 11, no. 3, p. 035008, 2014.

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

[23] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. Miller, G. Mueller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the bci competition iv," *Frontiers in Neuroscience*, vol. 6, 2012.

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.