



**HAL**  
open science

# A Separability Marker Based on High-Dimensional Statistics for Classification Confidence Assessment

Nathalie T H Gayraud, Nathanael Foy, Maureen Clerc

► **To cite this version:**

Nathalie T H Gayraud, Nathanael Foy, Maureen Clerc. A Separability Marker Based on High-Dimensional Statistics for Classification Confidence Assessment. IEEE International Conference on Systems, Man, and Cybernetics , Oct 2016, Budapest, Hungary. hal-01407759

**HAL Id: hal-01407759**

**<https://inria.hal.science/hal-01407759>**

Submitted on 2 Dec 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.

# A Separability Marker Based on High-Dimensional Statistics for Classification Confidence Assessment

Nathalie T.H. Gayraud

Université Côte d’Azur, Inria, France

Email: nathalie.gayraud@inria.fr

Nathanael Foy

Université Côte d’Azur, Inria, France

Email: nathanael.foy@inria.fr

Maureen Clerc

Université Côte d’Azur, Inria, France

Email: maureen.clerc@inria.fr

**Abstract**—This work provides a theoretical analysis framework for features that belong to the high dimensional Riemannian manifold of symmetric positive definite matrices. In non-invasive EEG-based Brain Computer Interfaces, such as the P300 speller, these are sample covariance matrices of the epoched EEG signal that are classified into two classes. An analysis of the class shape on the manifold is performed, and the separability level of the two classes is evaluated. The main contribution is the Separability Marker (SM)-confidence method, a method that appends a confidence marker to the prediction of a binary classifier whose decision function is based on the comparison of Riemannian distances.

## I. INTRODUCTION

Non-invasive EEG-based Brain-Computer Interfaces (BCI) are devices that provide a direct communication channel between a human brain and a machine. Many potential BCI applications exist [1]; one of them is the P300 speller, an application that uses the P300 Event-Related Potential and enables the user to spell words by observing a screen of flashing letters.

The main motivation behind this research is the development of new machine learning tools that in the future will lead to calibration-free BCIs. The major drawback of EEG-based BCI applications is the low signal to noise ratio of the signal; as a result, advanced machine learning techniques need to be employed to boost the classification accuracy. Moreover, as the cross-session and cross-subject variability of the EEG signal is very high, sessions are usually preceded by a calibration.

One of the most promising approaches is the use a classification method under the Riemannian framework, such as the Minimum Distance to Riemannian Mean (MDRM) algorithm, introduced in 2010 by A. Barachant to classify features in Motor Imagery based BCI [2]. The authors use the sample covariance matrix  $\Sigma_i$  of an epoched trial  $X_i$  as a feature, where  $X_i$  is a  $C \times N$  matrix,  $C$  is the number of electrodes used for the recording, and  $N$  depends on the length of the epoch and the sampling frequency. Using the Riemannian distance and the extrinsic definition of the mean by Fréchet, the algorithm estimates the centroid of each class in the training set by calculating the Riemannian mean of all the class features. For each new feature, its Riemannian distance to all centroids is calculated, and the smallest among these distances defines the winning class.

A number of adaptations of the MDRM algorithm have also been proposed for other BCI applications and problems [3],

[4], [5], [6]. In [6], Barachant et al. propose a calibration free, Riemannian distance based classification method for a P300 game called “Brain invaders”. In this work, an adaptive version of MDRM is implemented, where a new training set is learned during the testing of the algorithm. The reported Area Under Curve (AUC) score for both cross-session and cross-subject experiments is 80% in the worst case.

In the P300 Brain Invader game the target “alien” is selected by the program, on which the player has to focus his attention to destroy it, thus the classification result during the testing phase can be evaluated without feedback from the player. In most BCI applications however, such as the P300 speller, the classification results for the test set cannot be assessed without user feedback. Hence, to develop an adaptive algorithm similar to the one described in [6] that can potentially perform cross-session and cross-subject classification, a measure of confidence needs to be obtained for the classification results, so that we can decide whether a trial will be added to the learned training set or not.

Moreover, the MDRM classification method relies on the assumption that the point cloud consisting of the covariance matrices of a class forms a multidimensional sphere on the Riemannian manifold embedded with the chosen Riemannian metric. These assumptions cannot be easily verified, due to the dimensionality of the problem, the small amount of available data, the low signal to noise ratio, and the definition of the chosen Riemannian distance itself.

In the following sections, we present the contribution of our research. In section II we examine the Riemannian metric and the derived distance our feature space is subject to, and provide an intuitive explanation to what it represents. We present relevant geometric properties on high dimensional spaces and get a concept of the shape of the distributions our features are drawn from. Section III begins with an analysis of a specific dataset, and presents the *SM*-confidence method. The *SM*-confidence method makes use of the observations of our analysis to provide a marker of confidence that can be outputted along with the result of any classification algorithm that uses the Riemannian distance in its decision function. Sections IV and V close this paper with results and conclusions.

Our results show that the *SM*-confidence method provides a reliable confidence marker that can be used with an adaptive classification method such as the one proposed in [6], in the particular case where the features of the test set are learned

and used to update the training set without prior knowledge of the label, or user feedback.

## II. THEORETICAL FRAMEWORK

### A. The Riemannian Metric

Under the framework of classification methods based on Riemannian geometry [5], the sample covariance matrix has been emerging as a feature for the classification of mental tasks. The set of all sample covariance matrices is in fact a subset of the set of  $n \times n$  Symmetric Positive Definite (SPD) matrices  $P_n(\mathbb{R})$ , which constitutes a Riemannian manifold when it is embedded with the following Riemannian metric, first proposed by Forstner et al. [7].

$$\|\Sigma\| = \|\log(\Sigma)\|_F^2 = \sum_{i=1}^n \log^2 \lambda_i, \quad (1)$$

where  $\lambda_i$  are the eigenvalues of  $\Sigma$ . The structure of the resulting Riemannian manifold together with the above metric, which is often called the Affine Invariant metric, becomes highly regular, bearing much resemblance to a curved vector space. Under the Affine Invariant metric, the manifold transforms from a high dimensional cone into a regular and complete manifold of non positive curvature [8], [9].

Note that the same metric has been obtained in the bibliography under another framework as well. In their work, Skovgaard et al. obtained the Affine Invariant metric by using the Fisher Information and considering the set of all Multivariate Gaussian distributions of dimension  $n$  as a differentiable manifold [10].

The geodesic Riemannian distance  $d_R(\Sigma_1, \Sigma_2)$  can be derived from the Affine Invariant metric.

$$d_R(\Sigma_1, \Sigma_2) = \|\log(\Sigma_1^{-1}\Sigma_2)\|_F = \sqrt{\sum_{i=1}^n \log^2 \lambda_i}, \quad (2)$$

where  $\Sigma_1, \Sigma_2 \in P_n(\mathbb{R})$  and  $\lambda_i$  are the joint eigenvalues of the two matrices.

We distinguish two important differences between  $d_R$  and the Frobenius distance. These two differences motivate the choice of covariance matrices as an appropriate feature for the classification of epoched EEG signals.

The first one is that the two metrics reflect different measures of separability. The Frobenius distance between two  $n \times n$  matrices can be perceived as the action of vectorizing the matrices and computing the Euclidean distance between the resulting  $d$ -dimensional vectors, where  $d = n^2$ . The Riemannian metric reflects a different measure of separability. As shown in [10], it projects zero mean multivariate Gaussian distributions, by using their covariance matrices, onto a Riemannian manifold. On this manifold, the distance between two matrices corresponds to the informational difference between the Gaussians they represent.

Assuming that each column (time sample) of two epoched trials  $X_i^1, X_j^2$  that correspond to two different mental processes

is a random variable drawn from two different Gaussian distributions,  $\mathcal{N}(0, \Sigma^1), \mathcal{N}(0, \Sigma^2)$ , the Riemannian distance between their sample covariance matrices will be related to the informational difference between the two distributions.

The second difference concerns one of the properties that are specific only to  $d_R$ . In particular, it is invariant with respect to affine transformations on covariance matrices. This means that for any invertible matrix  $W$ ,  $d_R(\Sigma_1, \Sigma_2) = d_R(W\Sigma_1W^T, W\Sigma_2W^T)$ .

In [11], Congedo et Al. demonstrate the significance of that property in particular for BCI classification problems. Note that in [6], the authors hypothesize that the affine invariant property is what allows for the obtained classification results, under the assumption that cross-session and cross-subject variability can be described in terms of linear transformations.

### B. High Dimensional Spaces

While the Affine Invariant metric provides the Riemannian manifold of  $n \times n$  symmetric positive definite matrices with some very interesting properties, this manifold is a high dimensional space; its embedded dimension is equal to  $d = n(n+1)/2$ . In non-invasive EEG-based BCI where the classification feature is the sample covariance matrix,  $n$  corresponding to the number of electrodes used for the recording. For instance, the dimension of the manifold will be  $d = 2080$  for  $n = 64$  electrodes.

The geometry of high dimensional spaces presents us with a very counter intuitive set of phenomena; many examples can be found in [12]. Fortunately, this does not forbid the existence of some valuable properties that will allow us to gain significant insight on the shape of feature space. Under the assumption that the feature space of sample covariance matrices, embedded with the Riemannian metric, can be approximated as a set of random variables drawn from multidimensional Gaussian distributions, we use established properties of multidimensional Gaussians to develop our separability marker.

Regarding the separability of features that come from different classes, [12] provide a helpful insight on the distances between points that belong to different unit Gaussian distributions. They demonstrate that a mixture of two unit Gaussian distributions with centers  $\mu^1$  and  $\mu^2$  that are separated by a distance  $\delta$ , is separable when the distance between the two centers is  $\delta \in \Omega(d^{1/4})$ .

All of the above properties are proven in [12] for Euclidean high dimensional spaces. Nonetheless, our data lives in a Riemannian manifold, and we have to know whether the curvature of this particular space allows us to make the same assumptions. Since our previously presented theorems regarding distances between Gaussian distributions use trigonometric properties, we will use the works of [13] to get an insight on the effect of the curvature of the manifold.

Let  $T$  be a a geodesic triangle, that is, a triangle on the manifold whose edges are minimizing geodesics. Because geodesics are uniquely defined on the Riemannian manifold of symmetric positive definite matrices under the Riemannian distance  $d_R$ ,  $T$  can be uniquely mapped onto the tangent space

$TpM$  of the manifold, by fixing  $p$  to one of its three vertices. The Topogonov theorem states that the edges and angles of  $\tilde{T} \in TpM$  have upper and lower bounds with respect to the bounds of the sectional curvature of the manifold [13].

It has been shown that the lower and upper bounds  $K_C^-$ ,  $K_C^+$  of the sectional curvature  $K_C$  of the SPD manifold are  $K_C^- = -1/2 \leq K_C \leq K_C^+ = 0$  [14], [9]. This means that, fortunately, the Riemannian distance between two features on the Riemannian manifold can be approximated with little error by the Euclidean distance between their projection on the tangent space, provided that we choose an appropriate reference point. Typically, that reference point is the Riemannian mean of all the features.

Therefore, we will use the above described properties of multidimensional Gaussian distributions to establish a separability marker between two sets of features, in order to assess the confidence of a classification result under a Riemannian classifier.

### III. MATERIALS AND METHODS

#### A. Dataset Description

The dataset consists of EEG signals recorded during P300 speller sessions that were conducted by adult patients suffering from Amyotrophic Lateral Sclerosis. For the recording, an Refa-8 amplifier (ANT) with 12 electrodes (Fz,C3,Cz,C4,P7,P3,Pz,P4,P8,O1,Oz,O2) was used, and the sampling rate equal to 256 Hz. The experiment took place in the premises of the Nice University hospital, and had been approved by the local ethics committee CPP Sud Méditerranée [15].

Each subject participated in three free-spelling sessions, each one preceded by a calibration session. Only the calibration sessions are used in this paper. Initially, 20 patients participated in the experiment. The patients where at least one session contained a great number of artifacts were discarded from the dataset and the remaining sessions correspond to 14 patients.

Concerning the pre-processing of the signals, we first apply a 5th order Butterworth filter between 5 and 20Hz to the EEG signals. Each signal is then separated into epochs  $X_i$  of size  $C \times N$ , where  $C = 12$  is the number of electrodes and  $N = 128$  to account for a 0.5sec epoch starting at the time of the flash.

The goal of a classification method for a P300 speller is to differentiate between Target and Nontarget trials. Target trials are those who contain the elicited P300 component. As mentioned previously, the feature that is used under the Riemannian framework is the sample covariance matrix  $\Sigma_i = \frac{1}{N} X_i^T X_i$ . This is the feature that we use in the rest of the document, for both the analysis of the shape of the classes, and the evaluation of the  $SM$ -confidence method.

#### B. Geometrical Analysis of the Feature Space

Regarding the theoretical framework we have presented on section II, we want to assert whether the two classes in this application,  $c \in \{T, N\}$  (Target/Nontarget) can be

approximated as high dimensional Gaussian distributions on the Riemannian manifold.

Lower-dimensional Gaussian distributions have their mass concentrated near their expected values. In high dimensions, Hopcroft et al. [12] prove that there is very little mass located near the expected value of a multivariate Gaussian distribution. The expected squared Euclidean distance of a random variable from the expected value of the Gaussian distribution it is drawn from, is in fact on the order of  $d\sigma^2$ , where  $d$  is the dimension and  $\sigma$  is the standard deviation of the distribution. They also demonstrate that most of the mass of a unit Gaussian multivariate distribution will be concentrated within an annulus of constant width.

Moreover, Hopcroft et al. [12] demonstrate that any two randomly drawn points will almost surely be orthogonal with respect to the expected value of the distribution. According to that, we expect that all features belonging to a class will be almost equidistant, and that the average distance  $\delta_I$  between two features will be related to the average distance to the distribution center through the equation  $\sqrt{2}\delta_C = \delta_I$ .

We consider each class separately and calculate the average distance  $\delta_C = \text{avg}(d_R(\Sigma^c, \Sigma_i^c))$  between the features  $\Sigma_i^c$  that belong to class  $c$  and the centroid  $\Sigma^c$  of that class, which we compute using the method described in [8]. We also compute the average distance between same-class features,  $\delta_I = \text{avg}(d_R(\Sigma_i^c, \Sigma_j^c))$ .

TABLE I  
AVERAGE DISTANCE TO CENTROID AND AVERAGE DISTANCE BETWEEN FEATURES FOR TARGET (T) AND NONTARGET (N) CLASS

$\delta_C^T$	$\delta_I^T$	$\delta_C^N$	$\delta_I^N$
$4.59 \pm 0.78$	$5.88 \pm 0.75$	$4.53 \pm 0.73$	$5.79 \pm 0.72$

Table I displays the average over all subjects for  $\delta_C$  and  $\delta_I$  as well as the standard deviation of that average, calculated on the first session of each subject, for both classes. In this case study, our sample covariance matrices are  $12 \times 12$  matrices, so the dimension is  $d = 78$ , and  $\sqrt{d} \approx 8.83$ .

The result of that analysis show that, for both classes, the distance  $\delta_C$  between class centroids and class features appears to be on the order of  $\sqrt{d}$ . Additionally, the features are almost equidistant to each other, which we can deduct by observing the standard deviations on Table I.

Finally, if we compare the averages on Table I for the distances between features to the distances between each feature to its centroid, we can verify that  $\sqrt{2}\delta_C = \delta_I$  appears to be holding; we observe only a small deviation that is on average equal to  $\delta_I - \sqrt{2}\delta_C \approx 0.61$ ; it can be attributed to the effect of the curvature on the manifold.

Overall, we see that the Euclidean high dimensional properties of Gaussian distributions can be applied to our space. We proceed by making a cross-class comparison and try to calculate a marker of the amount of overlap that occurs between the two classes. This will give us a significant tool to evaluate the separability of Target and Nontarget classes.

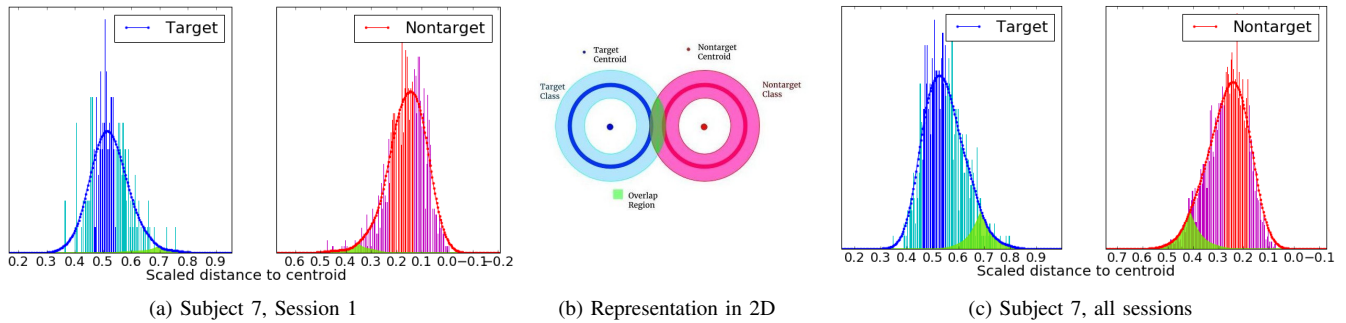


Fig. 1. A visualization of the distribution of distances between the class features and the estimated centroid. The distances have been scaled down by  $\sqrt{d}$ . The histograms represent the scaled distances distribution, approximated by a Gaussian probability distribution function. On 1a, the features are drawn from a single session, whereas 1c shows the distribution of features from three different sessions. 1b provides a 2D visualization of two multidimensional Gaussian distributions; the annulus is where the mass of the Gaussian is concentrated. The area in green represents a region where the two distributions have the highest chance of overlapping.

We perform an analysis that is based on the description of the shape of a high dimensional Gaussian distribution, given in sections II-B and III-B. A 2-dimensional schematic of this analysis is presented in Fig. 1b. Our goal is to obtain a measure of the amount of overlap between classes, in the annular regions where their distributions are concentrated.

We begin by scaling the distances between the features and their respective centers by  $\sqrt{d}$ , to obtain distributions that are no longer affected by the dimensionality of the space. We plot a histogram of these distances, and approximate their distribution with a Gaussian kernel, as shown in Fig. 1a and 1c. Note that, for the Nontarget class, the distances to the centroid are reversed, so that the histograms are coherent with the representation of Fig. 1b.

We display the results of this analysis for a single subject, which we have randomly chosen; subject 7. We perform the analysis twice, one to see the separability of the two classes within a single session (the first session), and once more for the union of all sessions. Observe that the distance distributions suggest a Gaussian probability density function; this is in accordance with our theoretical assumptions. We can also see that the features of the Nontarget class are closer to their centroid; the radius of the annulus is smaller. This can also be seen on Table I by comparing  $\delta_C^N$  to  $\delta_C^T$ .

This analysis allows us to visualize the width of the annulus in which the features are contained, and use that information to define a region where the possibility of class overlap is increased, by taking into account the distance between the two centroids. This overlap region is represented in 2D for a general case in Fig. 1b; histograms 1a and 1c can be seen as 1D projections of the general case. On Fig. 1a we can observe that, for a single session, the histograms that represent the two classes do not significantly overlap. On the other hand, the overlap is more important in Fig. 1c when the class features come from three different sessions. This is due to the cross-session variability, which is causing an increase in the width of the annulus.

Taking into account the previous observations on the overlap region, we calculate the separability marker  $SM$  for any pair

of classes in the following way. We estimate the distance distributions of the two classes from the distance distribution histograms using an appropriate Gaussian kernel. The two obtained curves that model the estimated distance distributions are superimposed, so that the distance between their expected values is equal to the distance between the two centroids,  $\delta = d_R(C^T, C^N)$ ; if there is an overlap region, the curves will intersect. Let  $U$  be the area under the union of these two curves and  $I$  the area under the intersection of the two curves. Then,  $SM$  is equal to  $SM = (U - I)/U$ .

Intuitively, this marker gives us a comparative measure of the overlap. A small value suggests a big overlap, so that the classes are harder to separate, whereas a large value suggests a small overlap.

On Fig. 2, we calculate the  $SM$  marker and compare it to the result obtained from training and testing a MDRM classifier; the reported statistic to assess the classification accuracy is Cohen's kappa value. We have also calculated the optimal distance needed for the classes to be perfectly separable. Recall that it is equal to  $d_{R, opt}(\Sigma^T, \Sigma^N) = \delta_{opt} = d^{1/4}$ . Figure 2 shows the ratio of the actual distance over the optimal distance,  $\delta/\delta_{opt}$ . Figure 2a displays the analysis and results for each subject within a single session, where a 5-fold cross-validation is performed to select a training and testing set. Figure 2b is a cross-session evaluation, where the training set of the classification contains the first two sessions, and the test set contains the third.

We can immediately see that the three markers are correlated. High  $SM$  scores accompany high classification scores, and distances that are near the optimal distance are also a strong marker of separability, as we see for example for subject 6, in both the single-session and the cross-session cases.

Now that we have analyzed the information that can be conveyed from calculating Riemannian distances between features and centroids, we proceed to suggest a probabilistic classifier whose methods will be based on principles of Riemannian geometry and high dimensional geometry.

TABLE II

THE CLASSIFICATION ACCURACY OF THE MDRM CLASSIFICATION ALGORITHM WITH RESPECT TO  $p_{SM}$  IN A CROSS-SESSION EXPERIMENT, FOR A SELECTED SET OF SUBJECTS, FOLLOWED BY THE AVERAGE NUMBER OF FEATURES THAT WERE SELECTED OVER THE WHOLE TEST SET.

Threshold value $thr$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Subject 1	0.831	0.831	0.834	0.841	0.867	0.871	0.926	0.916	0.916	0.921
Subject 6	0.833	0.888	0.924	0.947	0.959	0.970	0.975	0.986	0.991	0.990
Subject 9	0.803	0.839	0.874	0.892	0.906	0.940	0.954	0.959	0.947	0.945
Subject 11	0.206	0.192	0.196	0.210	0.218	0.254	0.285	0.431	0.524	0.558
Subject 12	0.812	0.821	0.825	0.831	0.834	0.839	0.849	0.850	0.856	0.870
Subject 13	0.317	0.325	0.333	0.361	0.371	0.402	0.459	0.451	0.416	0.366
Subject 14	0.746	0.721	0.714	0.710	0.713	0.695	0.772	0.767	0.628	0.592
Subject 18	0.330	0.321	0.320	0.325	0.345	0.366	0.392	0.440	0.349	0.406
$N^{thr}/N(\%)$	100%	87 %	73 %	60%	48%	37%	29%	22%	17%	12%

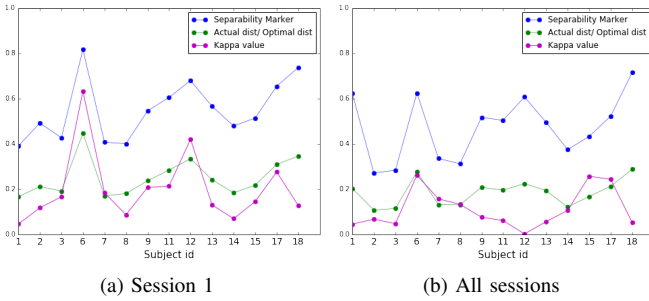


Fig. 2. Comparison between the separability marker  $SM$ , the ratio  $\delta/\delta_{opt}$ , and value of Cohen's kappa after classifying the data with the MDRM algorithm. (a) concerns only Session 1, and (b) concerns all sessions.

### C. The $SM$ -confidence Method

We propose  $SM$ -confidence, a two part method that appends a confidence value to the prediction result of the MDRM classification method. The first part of  $SM$ -confidence is based on the analysis presented in section II, and can be embedded to the training step of MDRM. The second part concerns the estimation of a confidence value that can be appended to the output of the prediction process. The  $SM$ -confidence method can also be adapted to any classification method that uses symmetric positive definite matrices as features, and the minimum Riemannian distance to Riemannian class mean, to decide on a binary classification problem.

The training step of MDRM consists of calculating the centroids  $\Sigma^T$  and  $\Sigma^N$  for each class, where  $T$  and  $N$  are the labels of the two classes, and  $I^c$  the set of features corresponding to class  $c$ . Our method adds the computation of three more values;  $\delta_{max}^T$ ,  $\delta_{max}^N$  and  $SM$ .

$\delta_{max}^c$  is the maximum value of the differences of distances to each centroid:

$$\delta_{max}^c = \max(d_R(\Sigma_i^c, \Sigma^{\bar{c}}) - d_R(\Sigma_i^c, \Sigma^c)),$$

where  $i \in \{1, 2, \dots, |I_c|\}$ ,  $\Sigma_i^c \in I_c$ ,  $\bar{c}$  is the other class. The separability marker  $SM$  is calculated according to the method that has been presented in section III-B.

The second part of our method follows the classification part of the MDRM classification method, and appends the confidence marker  $p_{SM}^c$  to the classification label of a new feature  $\Sigma_{new}$  that has been attributed label  $c$ .

The calculation of  $p_{SM}^c$  is carried out in the following manner:

$$p_{SM}^c = \left( \frac{\delta_{new}^c}{\delta_{max}^c} \right) \cdot \left( 1 - \frac{\delta_{new}^{\bar{c}}}{\delta_{max}^{\bar{c}}} \right) \cdot SM$$

$$\text{where } \delta_{new}^{c(\bar{c})} = d_R(\Sigma_{new}, \Sigma^{\bar{c}(c)}) - d_R(\Sigma_{new}, \Sigma^{c(\bar{c})})$$

The first two components give us an measure of how probable it is that the new feature belongs to class  $c$ , by taking into account the Riemannian distances to both centroids. The fraction  $\delta_{new}^{c(\bar{c})}/\delta_{max}^{c(\bar{c})}$ , where  $\bar{c}$  is the other label, will return a higher value depending on how close  $\Sigma_{new}$  is to the centroid of class  $c(\bar{c})$ .

However, if the classes are barely separable, that is, the centroids are very close to each other, or the standard deviation of the sample is affecting the maximum distances, the product of the first two components may be higher for the wrong class, and lower for the correct one, due to the overlap of the two classes. To account for that factor, the product of the first two components is scaled by an amount that is equal to the the separability marker  $SM$ , which is calculated during the calibration of the classifier.

Since the  $p_{SM}$  marker is calculated for both classes, it can be used to classify new trials with a given confidence. It can also be used to select trials that cannot be classified. We see an example of the latter for a cross-classification experiment in the next session, and discuss the results.

## IV. RESULTS

We perform a series of cross-session classification experiments, using the dataset presented in section III-A. For each subject, we train and test the MDRM classifier with trials that belong to this subject alone. The training set consists of trials that are drawn from two out of three sessions, and the test set corresponds to the third session. The experiments are carried out with all three possible combinations of sessions, and for



each subject the reported result is the average accuracy. We use the  $SM$ -confidence method and compute the previously described values during the training step of the classifier. The output of the testing step is the label of each trial,  $l_i = T(N)$  along with the corresponding confidence marker  $p_{SM}^c[i]$ .

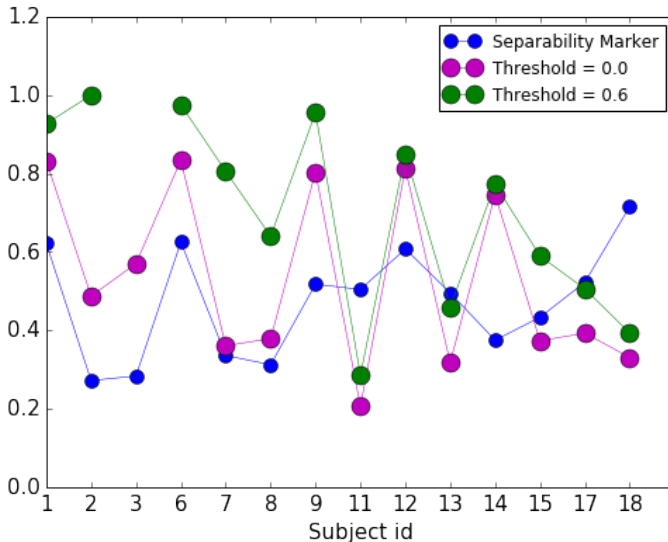


Fig. 3. Evaluation of the confidence marker  $p_{SM}$  for a cross-session experiment. When the threshold over  $p_{SM}$  is equal to zero, the result is equivalent to the classification accuracy of the MDRM algorithm. We compare it to the classification accuracy obtained from the trials where  $p_{SM} > 0.6$ . The value of  $SM$  is displayed as well.

To assess the level of certainty provided by our method, we display, on Fig. 3, the classification accuracy for each subject and compare it to a subset of the result where  $p_{SM} > 0.6$ . We can see that in most cases, the classification accuracy of the selected subset is higher. For instance, while subject 2 has low classification accuracy, approximately equal to  $acc = 0.5$ , the selected subset has an accuracy equal to  $acc = 1.0$ ; these features can be used to efficiently train an adaptive classifier.

In Table II, one can see the improvement of the classification accuracy of the MDRM algorithm over a threshold value for the  $p_{SM}$  that varies from  $thr = 0.0$  to  $thr = 0.9$ . Because some subjects do not have a high average  $p_{SM}$  value, and for such subjects very few features are selected for high thresholds, we display only the subjects where at least 10% of the features were selected.

We can see that the best accuracy values are obtained when the threshold is equal to  $thr = 0.6$  and  $thr = 0.7$ . This is mostly due to the percentage of selected features, which becomes very low for  $thr \geq 0.8$ . In almost every case, there is a net improvement in comparison to the classification accuracy reported by the algorithm.

## V. CONCLUSIONS

In this research, we have studied the shape of distributions of sample covariance matrices on the Riemannian manifold of symmetric positive definite matrices. When the sample covariance matrices are classification features of a binary MDRM

classifier, obtained from BCI applications such as the P300 speller, we can approximate their shape by using theorems that apply to high-dimensional Gaussian distributions. We can also get an insight on how separable the two classes are, and use that knowledge to get a confidence marker on the classification results. The  $SM$ -confidence method provides such a marker, and the experimental results we have obtained suggest that in future works it can be used to improve the prediction accuracy of an adaptive MDRM classifier, where each new feature will be added to the learned training set only when the  $p_{SM}$  marker is above a certain threshold.

## VI. ACKNOWLEDGEMENTS

The Nice University Hospital is acknowledged for providing the dataset used in this paper.

## REFERENCES

- [1] E. Thomas, M. Dyson, and M. Clerc, "An analysis of performance evaluation for motor-imagery based bci," *Journal of neural engineering*, vol. 10, no. 3, p. 031001, 2013.
- [2] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Riemannian geometry applied to bci classification," in *Latent Variable Analysis and Signal Separation*. Springer, 2010, pp. 629–636.
- [3] C. Lindig-León, N. Gayraud, L. Bougrain, and M. Clerc, "Comparison of hierarchical and non-hierarchical classification for motor imagery based bci systems," in *The Sixth International Brain-Computer Interfaces Meeting*, 2016.
- [4] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain-computer interface classification by riemannian geometry," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 4, pp. 920–928, 2012.
- [5] M. Congedo, A. Barachant, and A. Andreev, "A new generation of brain-computer interface based on riemannian geometry," *arXiv preprint arXiv:1310.8115*, 2013.
- [6] A. Barachant and M. Congedo, "A plug&play p300 bci using information geometry," *arXiv preprint arXiv:1409.0107*, 2014.
- [7] W. Förstner and B. Moonen, "A metric for covariance matrices," in *Geodesy-The Challenge of the 3rd Millennium*. Springer, 2003, pp. 299–309.
- [8] X. Pennec, P. Fillard, and N. Ayache, "A riemannian framework for tensor computing," *International Journal of Computer Vision*, vol. 66, no. 1, pp. 41–66, 2006.
- [9] X. Pennec, "Statistical computing on manifolds: from riemannian geometry to computational anatomy," in *Emerging Trends in Visual Computing*. Springer, 2009, pp. 347–386.
- [10] L. T. Skovgaard, "A riemannian geometry of the multivariate normal model," *Scandinavian Journal of Statistics*, pp. 211–223, 1984.
- [11] M. Congedo, B. Afsari, A. Barachant, and M. Moakher, "Approximate joint diagonalization and geometric mean of symmetric positive definite matrices," *PLoS one*, vol. 10, no. 4, p. e0121423, 2015.
- [12] J. Hopcroft and R. Kannan, "Computer science theory for the information age," 2012.
- [13] M. Berger, *A panoramic view of Riemannian geometry*. Springer Science & Business Media, 2012.
- [14] M. R. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*. Springer Science & Business Media, 1999, vol. 319.
- [15] M. Clerc, J. Mattout, E. Maby, D. Devlaminck, T. Papadopoulo, V. Guy, and C. Desnuelle, "Verbal communication through brain computer interfaces," in *Interspeech-14th Annual Conference of the International Speech Communication Association-2013*, 2013.