Channel Selection Procedure using Riemannian distance for BCI applications

Alexandre Barachant, Stephane Bonnet

Channel Selection Procedure using Riemannian distance for BCI applications

Alexandre Barachant, Stéphane Bonnet

***Abstract*— This article describes a new algorithm to select a subset of electrodes in BCI experiments. It is illustrated on a two-class motor imagery paradigm. The proposed approach is based on the Riemannian distance between spatial covariance matrices which allows to indirectly assess the discriminability between classes. Sensor selection is automatically done using a backward elimination principle. The method is tested on the dataset IVa from BCI competition III. The identified subsets are both consistent with neurophysiological principles and effective, achieving optimal performances with a reduced number of channels.**

I. INTRODUCTION

A Brain-Computer Interface (BCI) is a system for translating the brain neural activity into commands for external devices [1]. It aims at restoring communication and control in severely motor-disabled subjects that cannot use conventional communication channels like muscles or speech to interact with their environment. The targeted population concerns paralysed people suffering from severe motor disabilities: locked-in syndrome (LIS), spinal chord injury (SCI) in the range C4-C7. In such cases, cognitive functions are still preserved. As stated, BCI is based on the monitoring of the users brain activity and the translation of the users intention into commands. To do so, different measurement systems have been proposed in the past ranging from invasive recording techniques (micro-electrodes implanted into the cortex to record single-unit or multi-unit activity, ECoG) to non-invasive ones. Electro-EncephaloGraphic (EEG) is widely used in the current BCI realizations. It is a low-cost, practical modality that possesses a high temporal resolution. However, this technique suffers from a poor spatial resolution and it is very sensitive to noise.

The user is usually wearing a cap with electrodes placed directly onto his scalp. Positioning the electrodes in a very reproducible way, so to achieve low impedance is part of the EEG experimenter know-how. The clinical approach of placing a large number of wet electrodes (with gel) is usually cumbersome, time-consuming, and impractical for BCI applications. Furthermore, cost of such system and computational requirements should also be kept low in order to envisage out-of-the-lab BCI applications. This article is mainly focused on the **channel selection procedure** in EEG-based BCI experiments. For a given subject, such work could help to perform BCI paradigms with a reduced number of recording channels and still good performances.

In the literature, one can distinguish two types of electrode selection approaches. First, *patient-specific* methods allow a subset selection customized for each subject to increase the individual BCI performances. Second, *application-specific* methods seek the electrode subset that is shared between all subjects, for a particular application, to achieve the best global BCI performances [2].

In both cases, a first session is realized with a large number of electrodes in order to build a dataset on which the channel selection procedure can be assessed. Later on, the remaining sessions will be carried out with a reduced number of electrodes. Different criteria have been proposed in the literature for electrode subset selection : In [5], [14], the signal-to-noise ratio or the signal-to-signal plus noise ratio is used to select the best electrode subset. In [4], [6], the information content shared between electrode feature space and the training class is assessed either using mutual information or multiple correlation. These methods prevent the use of a spatial filter pre-processing. In [3], one ranks the absolute value of spatial filter coefficients and keeps a predefined number of electrodes. This simple method depends on the robustness of the computation of the spatial filter. But the most often used criterion is related to the classification accuracy for a given subset [2].

In few cases, channels are directly sorted according to the chosen criterion. Bust most often, the subset is iteratively built using stepwise research techniques, much like in regression procedures where one search for the most appropriate model. **Backward selection** amounts to exclude one electrode at a time from the subset. At each step, one removes the electrode for which the complementary subset gives the best score. Forward or backward-forward research techniques are two other possible options. These procedures may be more or less greedy if the criterion has a high computational cost or if the number of tested configurations is important. Moreover, it will depend on the possibility of reusing previous computations. Finally, it is necessary to define a stopping condition based on a predefined number of electrodes or by inspecting the criterion evolution curve [6].

This paper is organised as follows. Section II describes the proposed channel selection procedure using Riemannian distance, while numerical results are provided in Section III. Section IV concludes the paper with comments and perspectives.

Manuscript received December 17, 2010.

A. Barachant is with CEA Leti, MINATEC Campus, DTBS, F-38054 Grenoble (FRANCE) alexandre.barachant@cea.fr

S. Bonnet is with CEA Leti, MINATEC Campus, DTBS, F-38054 Grenoble (FRANCE) stephane.bonnet@cea.fr

II. A Riemmanian distance-based selection procedure

A. Notations

During the calibration phase, different mental task realizations are usually performed on a cue-based paradigm [7] and recorded using a large number of electrodes N_e. Without loss of generality, we consider here a two-class paradigm [7].

First, the EEG recording is band-pass filtered in the appropriate frequency band. Second it is divided into epochs that correspond to the different known brain patterns. Let denote $\mathbf{X}_i^{(c)}$ the i-th trial for the c-th condition. Each trial consists of a $N_e \times N_t$ matrix, where N_t is the number of samples in time. The sample spatial covariance matrix is computed using the relation $\mathbf{C}_i^{(c)} = \frac{1}{N_t}\ \mathbf{X}_i^{(c)}\mathbf{X}_i^{(c)T}$. Such second-order information has been shown to be well adapted to catch the relevant information between two mental tasks [8]. We have then a set of covariance matrices akin to each mental task.

B. Riemannian manifold of SPD matrices

Covariance matrices are symmetric positive definite (SPD) matrices that live in a connected Riemannian manifold [10]. Furthermore, this manifold has been well studied and analytical formulae exist to manipulate such matrices in their native space [9]. For this article, two concepts will be mainly used, the Riemannian distance between two covariance matrices and the mean of a set of covariance matrices.

The Riemannian distance between two SPD covariance matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ is given by [9] :

$$\delta_R(\mathbf{C}_1, \mathbf{C}_2) = \|\mathrm{Log}\left(\mathbf{C}_1^{-1}\mathbf{C}_2\right)\|_F = \sqrt{\sum_{n=1}^{N_e} \log^2 \lambda_n}, \quad (1)$$

where the λ_n's are the real and strictly positive eigenvalues of the matrix $\mathbf{C}_1^{-1}\mathbf{C}_2$, $\mathrm{Log}(.)$ is the log-matrix operator and $\|.\|_F$ is the Frobenius norm of a matrix. The Riemannian distance, in (1) is different from the usual Euclidean distance $\delta_E(\mathbf{C}_1, \mathbf{C}_2) = \|\mathbf{C}_1 - \mathbf{C}_2\|_F$ since it includes the geometry of the manifold.

For the c-th class condition, the mean of $N^{(c)}$ spatial covariance matrices $\mathbf{C}_i^{(c)}$ can be defined by :

$$\bar{\mathbf{C}}^{(c)} = \underset{\mathbf{C}}{\mathrm{argmin}} \sum_{i=1}^{N^{(c)}} \delta_R^2(\mathbf{C}, \mathbf{C}_i^{(c)}). \quad (2)$$

This geometric mean can be iteratively computed using efficient iterative algorithms, like in [11].

C. Patient-specific channel subset selection

The proposed method starts with the computation of the two class-conditional mean matrices using (2). Indeed, Common Spatial Pattern (CSP), one of the most popular algorithms in BCI, is also based on such class-conditional mean covariances matrices, albeit usually formulated in Euclidean space [8]. It is shown in [12] that the computation of the spatial filters relies implicitly on the computation of the Riemannian distance between these two mean matrices. Since CSP algorithm computes spatial filters in order to maximize variance of the signal in one condition while minimizing it for the other, CSP is an efficient pre-processing step to linearly transform the data to make both classes well separated.

Using the same point-of-view, we suggest to use the Riemannian distance between class-conditional mean matrices as a criterion for channel selection.

$$\mathrm{Crit} = \delta_R(\bar{\mathbf{C}}^{(1)}, \bar{\mathbf{C}}^{(2)}) \quad (3)$$

Maximimizing this distance should increase in some sense the discriminability between the two mental tasks. Furthermore, the proposed criterion is well adapted for subsequent CSP.

The proposed algorithm is based on backward selection starting from N_e electrodes to $N^\star$ remaining electrodes. The backward selection is achieved by taking into account the fact that removing an electrode of the subset will only impact one row and one column of the mean covariance matrices. Thus, it is necessary to compute the mean covariances matrices only one time at the beginning of the selection procedure, making this algorithm fast and computationally efficient. This procedure is explained in detail below.

Suppose a subset of N electrodes has already been selected. The criterion associated with the removal of the i-th electrode is computed by removing the i-th row and i-th-column from both class-conditional mean matrices. Denote by $\mathcal{C}\left(.,i\right)$ such matrix reduction. After having removed independently each electrode from the current subset, one obtains N performance scores. The subset with the highest score is kept for next iteration of the algorithm.

Algorithm 1 Channel subset selection

Input: $\bar{\mathbf{C}}^{(1)}$ and $\bar{\mathbf{C}}^{(2)}$
Input: $N^\star$
Output: $Subset$

```
Subset = [1 ... N_e]
for k = 1 to N_e − N* do
    for i = 1 to N_e − k + 1 do
        D̄^(1) = C(C̄^(1), i)
        D̄^(2) = C(C̄^(2), i)
        Crit(i) = δ_R(D̄^(1), D̄^(2))
    end for
    i* = argmax_i Crit(i)
    C̄^(1) = C(C̄^(1), i*)
    C̄^(2) = C(C̄^(2), i*)
    Subset(i*) = [ ] {Remove from the Subset the corresponding channel}
end for
return Subset
```

D. Application-specific channel subset selection

Using the patient-specific procedure described in section II-C we can obtain the application-specific one by simply selecting the $N^\star$ electrodes which appears among all subsets for all subjects. This procedure is usually done over a large population for good generalization performances of the chosen subset [2].

E. Multi-class extension

A possible extension to a K-class BCI paradigm is given by the following criterion to be maximized :

$$\text{Crit} = \sum_{k=1}^{K} \sum_{j>k}^{K} \delta_R(\bar{C}^{(k)}, \bar{C}^{(j)}).$$

In this case, one seeks to maximize the average Riemannian distance between all pairwise class-conditional matrices.

III. Numerical results

A. Data

The proposed method is benchmarked on the dataset IVa from the BCI competition III [1]. This dataset is well suited for the issue of channel selection since it is composed by EEG recording using 118 electrodes. The experiment is a classical cue-based motor imagery paradigm in which 5 users have performed a total of 280 trials of right hand and foot motor imagery. EEG signals are bandpass filtered in the large frequency band 8-30 Hz by a 5-th order Butterworth filter. The time interval is restricted to the segment located from 0.5s to 4s after the cue.

B. Illustrated results

Fig. 1 shows the evolution of the criterion against the number of selected sensors for the five subjects. Here the criterion is normalized by dividing it with the total Riemannian distance between the two classes covariances matrices using N_e electrodes.

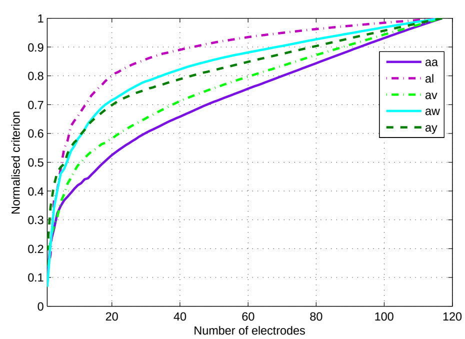

Fig. 1. Evolution of the criterion against the number of selected sensors for the five subjects.

For all subjects, we can notice the same behaviour with a rapid growth of the criterion, an inflexion and then a linear increase. This last trend informs us that the added electrodes are equivalent in terms of distance and thus are not relevant for the discrimination of the two mental tasks. As we can see, only a small number of electrodes ($N^\star$ <20) holds a large part of the distance between classes. For the subject al, the subset of 20 electrodes carries 80 $\%$ of the total distance.

[1] http://www.bbci.de/competition/iii/

Fig. 2 exhibits the *patient-specific* selection for the 5 subjects and the corresponding *application-specific* selection. The location of the 10-electrode subset is over the sensory-motor cortex which is in good accordance with the performed mental tasks. The *application-specific* selection is even clearer, with a majority of electrodes selected over the area dedicated to the right hand. Indeed, hand movements, imagined by the subjects, result in the activation of dedicated cortical areas (close to C3, C4 electrodes) in given frequency bands.

C. Classification results

The classification accuracy (CA) is assessed through a 30-fold cross-validation procedure. For each training dataset, the described channel selection procedure is first applied. Then for each selected electrode subset, a set of 6 CSP filters is computed (corresponding to the 3 maximum and 3 minimum eigenvalues of $\left[\bar{\mathbf{C}}^{(1)}\right]^{-1}\bar{\mathbf{C}}^{(2)}$, as recommended in [8]). Finally, each training trial is spatially filtered to yield 6 aggregate signals from which the log-variance is computed. This feature extraction procedure results for each training trial in a set of 6-D feature vectors on which a standard LDA is applied. For each test dataset, the learned classification rule is applied onto the (learned) spatial filtered test trials. CA is computed by averaging the CAs obtained on the 30 test datasets.

As it can be observed in Fig. 3, the channel selection procedure succeeds in reaching nearly optimal performances with very reduced subset. For instance, only 3 electrodes are sufficient in subject al to achieve above 97% correct classification. In average, a 10-electrode subset seems sufficient to catch the relevant information and achieves the best mean CA, as seen on the dotted curve in Fig. 3. In addition, the proposed criterion is strongly related to the CA since the curves are nearly identically sorted in Fig. 1 and Fig. 3.

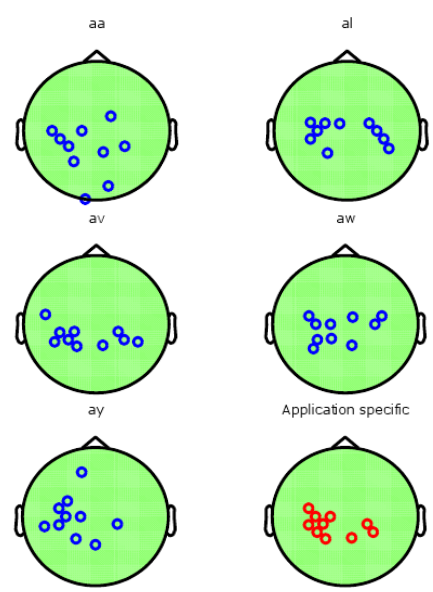

Fig. 2. Patient-specific selection of 10 electrodes and application-specific selection

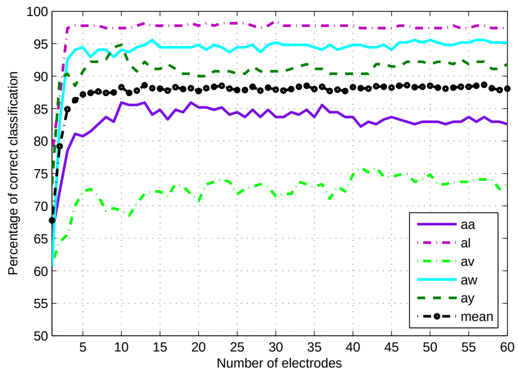

Fig. 3. Classification accuracy versus number of selected sensors.

TABLE I
CLASSIFICATION ACCURACIES (MEAN IN %) OBTAINED FOR EACH SUBJECT:

User | (tr/te)	118 ch.	10 ch. PS	10 ch. AS
aa | (168/112)	67	**74.1**	72.3
al | (224/56)	96.4	**98.2**	94.6
av | (84/196)	48.5	59.2	**71.4**
aw | (56/224)	74.6	**77.7**	64.7
ay | (28/252)	48.4	**80.6**	67.5
mean	67	**78**	74.1

D. Comparison

The whole procedure is repeated in the same conditions as in the BCI competition III. In this case, the training dataset is fixed to 168, 224, 84, 56, and 28 trials for subject aa, al, av, aw, and ay and the CA is assessed on the remaining test dataset (total 280 trials). Table I shows the results obtained when no channel selection is applied (118 ch.), when 10 channels are selected with the Patient Specific procedure (10 ch. PS) and when 10 channels are selected with the Application Specific procedure (10 ch. AS). The AS procedure is obtained on the 5 subjects' training datasets.

The PS channel procedure outperforms the raw procedure by almost 10% of CA. As expected, the AS channel selection provides lower CA than the PS one but still much better than the raw method. In addition, it is observed that, for subject ay, that the channel selection procedure is effective on small size training set. Finally, these results rank our methods between the 3rd and the 4th place of the BCI Competition III, without optimisation on time interval nor frequency band.

IV. Conclusion and Perspectives

In this article, we have presented a new channel selection procedure based on the Riemannian distance between covariance matrices. The proposed criterion is efficient in the context of a CSP-LDA classification framework. Excellent results were demonstrated on a BCI competition dataset with a major reduction from 118 electrodes to less than 10 electrodes, with no major loss of BCI performance. This work opens interesting perspectives for cost-effective attractive EEG cap design. Patient specific and application specific channel selection strategies have also been proposed.

A perspective of this work is to investigate other distances between covariance matrices like the Log-affine distance [13]. Indeed, for large number of electrodes, the calibration phase can be quite long since it relies on numerous eigenvalue decompositions. Another possibility to decrease computational effort would be to recursively compute the Riemannian distance during the backward procedure from the matrix pencil of $(\bar{\mathbf{C}}^{(1)}, \bar{\mathbf{C}}^{(2)})$.

The employed criterion is optimal for CSP algorithm but does not take into account the dispersion of data. For a more general approach, a better criterion should be used in order to measure the discriminability of data by also taking into account the intra-class variance and not only the inter-class distance. One solution is to use a Fisher criterion or a Student t-test defined in the Riemannian space of covariance matrices. Such criterion can be helpful to solve the ambiguity between two subsets with a similar distance between the intra-class covariance matrices.

References

[1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller and T.M. Vaughan "Brain-computer interfaces for communication and control", *Clinical Neurophysiology*, 2002, pp. 767-791.

[2] C. Sannelli, T. Dickhaus, S. Halder, E. Hammer, K.-R. Müller and B. Blankertz "On Optimal Channel Configurations for SMR-based BrainComputer Interfaces", *Brain Topography*, 2010.

[3] Y. Wang, S. Gao and X. Gao, "Common Spatial Pattern Method for Channel Selection in Motor Imagery Based Brain-computer Interface", *in 27th Annual International Conference of the Engineering in Medicine and Biology Society*, 2005, pp. 5392-5395.

[4] T. Lan, D. Erdogmus, A. Adami, M. Pavel and S. Mathan, "Salient EEG Channel Selection in Brain Computer Interfaces by Mutual Information Maximization", *in 27th Annual International Conference of the Engineering in Medicine and Biology Society*, 2005, pp. 7064-7067.

[5] B. Lou, B. Hong, X. Gao and S. Gao, "Bipolar electrode selection for a motor imagery based braincomputer interface",, Sept. 2008.

[6] A. Barachant, T. Aksenova and S. Bonnet, "Filtrage spatial robuste partir d'un sous-ensemble optimal d'lectrodes en BCI EEG", *GRETSI*, 2009.

[7] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller and G. Curio, "The non-invasive Berlin Brain-Computer Interface: fast acquisition of effective performance in untrained subjects", *NeuroImage*, 2007, pp. 539-550.

[8] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K.-R. Müller, "Optimizing Spatial filters for Robust EEG Single-Trial Analysis", *IEEE Signal Processing Magazine*, 2008, pp. 41-56

[9] M. Moakher, "A Differential Geometric Approach to the Geometric Mean of Symmetric Positive-Definite Matrices", *SIAM Journal on Matrix Analysis and Applications*, 2005, pp. 735-747.

[10] F. Barbaresco, "Innovative tools for radar signal processing Based on Cartan's geometry of SPD matrices & Information Geometry", *IEEE Radar Conference*, 2008, p. 1-6.

[11] A. Barachant, S. Bonnet, M. Congedo and C. Jutten, "Riemannian geometry Applied to BCI classification", *Latent Variable Analysis and Signal Separation - (LVA-ICA)* , 2010, pp. 629-636, Volume 6365 LNCS 2010.

[12] A. Barachant, S. Bonnet, M. Congedo and C. Jutten, "Common Spatial Pattern revisited by Riemannian geometry", *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, 2010, pp. 472-476.

[13] V. Arsigny, P. Fillard, X. Pennec and N. Ayache, "Geometric Means in a Novel Vector Space Structure on Symmetric Positive-Definite Matrices", *SIAM Journal on Matrix Analysis and Applications*, 2007, pp. 328-347.

[14] H. Cecotti, B. Rivet, M. Congedo, C. Jutten, O. Bertrand, E. Maby and J. Mattout, "Suboptimal sensor subset evaluation in a p300 brain-computer interface", *European Signal Processing Conference (EUSIPCO)*, 2010.