Common Spatial Pattern revisited by Riemannian Geometry

Alexandre Barachant, Stephane Bonnet, Marco Congedo, Christian Jutten

HAL Id: hal-00602686
https://hal.science/hal-00602686

Submitted on 23 Jun 2011

Common Spatial Pattern revisited by Riemannian geometry

Alexandre Barachant [#], Stphane Bonnet [#], Marco Congedo [*], Christian Jutten [*]

[#] *CEA, LETI, DTBS/STD/LE2S,*
17 rue des Martyrs, F-38054 Grenoble, FRANCE

[*] *Team ViBS (Vision and Brain Signal Processing),*
GIPSA-lab, CNRS/UJF-INPG
Domaine Univiversitaire, 38402 Saint Martin d'Hères, FRANCE

Abstract—**This paper presents a link between the well known Common Spatial Pattern (CSP) algorithm and Riemannian geometry in the context of Brain Computer Interface (BCI). It will be shown that CSP spatial filtering and Log variance features extraction can be resumed as a computation of a Riemann distance in the space of covariances matrices. This fact yields to highlight several approximations with respect to the space topology. According to these conclusions, we propose an improvement of classical CSP method.**

I. INTRODUCTION

A Brain-Computer Interface (BCI) aims at providing a new communication channel for people who are suffering of important paralysis. Most of non-invasive BCIs use Electroencephalogram (EEG) in order to measure the cerebral activity. Through a paradigm, a mental task is associated to a command (for example imagine a left hand movement to turn left on a wheelchair). By analysis of cerebral activity, the BCI identifies and detects the user intention (the mental task) and performs the corresponding command. Each mental task has different properties in temporal, frequential and spatial domains. Since EEG provides very noisy measurements with low spatial resolution it is necessary to enhance the specific characteristics of each mental task by appropriate methods.

A lot of useful information lies in the spatial domain since the brain areas involved in the mental tasks are well separated. Spatial filtering methods try to collect this information and discard the irrelevant ones. Common Spatial Pattern (CSP) is a well known method in BCI domain [1]. The present work establishes a link between the CSP algorithm and the Riemannian geometry in the space of Symmetric Positive Definite (SPD) matrices.

II. SPATIAL FILTERING IN EEG-BCI

The conventional EEG-BCI signal chain begins with several pre-processing steps, such as time windowing, frequency filtering, etc. Furthermore, one important step consists generally in **spatial filtering** of EEG data [1] , i.e., in performing a linear combination of EEG signals, for enhancing signals coming from a particular area of the brain. Mathematically, this operation is parametrized by a matrix $\mathbf{W} \in \Re^{N \times J}$, which projects the EEG signal $\mathbf{x}(t) \in \Re^{N \times 1}$ from the original sensor space into the signal $\mathbf{z}(t) \in \Re^{J \times 1}$ which lives in the surrogate sensor space :

$$\boxed{\mathbf{z}(t) = [z_1(t) \ldots z_J(t)]^T = \mathbf{W}^T \mathbf{x}(t)} \quad (1)$$

Each column vector of $\mathbf{W}$, $\mathbf{w}_j$, denotes a spatial filter and $z_j(t) = \mathbf{w}_j^T \mathbf{x}(t)$ the corresponding spatial filtered signal at instant t. More generally, for a given EEG trial $\mathbf{X} \in \Re^{N \times K}$ of K time samples, we can define $\mathbf{Z} = \mathbf{W}^T \mathbf{X}$ with $\mathbf{Z} = [\mathbf{z}(1) \ldots \mathbf{z}(K)] \in \Re^{J \times K}$ and $\mathbf{z}_j = \mathbf{w}_j^T \mathbf{X}$.

Different design methods exist for the $\mathbf{w}_j$'s, depending on the problem at hand and in agreement with the chosen criterion. For EEG source localisation, beamforming methods are often used [2] but need an appropriate forward model. On the other hand, in BCI, blind methods are usually preferred [3].

In the context of a two-class paradigm in EEG-BCI, the Common Spatial Pattern (CSP) algorithm is the technique of choice for spatial filtering [1]. Briefly, this method amounts to maximizing the variance of the (spatially) filtered signal under one condition while minimizing it for the opposite condition, as shown in Fig. 1.

CSP is a supervised technique since it requires known trials for both conditions in order to estimate $\mathbf{W}$. First, CSP builds an average covariance matrix for each class. Let $\mathbf{P}_1$ and $\mathbf{P}_2$ being such class-related mean covariance matrices. Spatial filter $\mathbf{W}$ is obtained by CSP using a joint diagonalization of both matrices and subject to an equality constraint. i.e. :

$$\begin{cases} \mathbf{W}^T \mathbf{P}_1 \mathbf{W} = \mathbf{D}_1 \\ \mathbf{W}^T \mathbf{P}_2 \mathbf{W} = \mathbf{D}_2 \\ \mathbf{D}_1 + \mathbf{D}_2 = \mathbf{I}_N \end{cases} \Leftrightarrow \begin{array}{l} \mathbf{P}_1 \mathbf{w}_j = \lambda_j (\mathbf{P}_1 + \mathbf{P}_2) \mathbf{w}_j \\ \text{with} \quad \lambda_j = (\mathbf{D}_1)_{jj} \end{array}$$

In the above equation $\mathbf{D}_1$ and $\mathbf{D}_2$ are two diagonal matrices and λ the generalized eigenvalues which also coincide with diagonal elements of $\mathbf{D}_1$. Because λ_j are strictly positives and $\mathbf{D}_1 + \mathbf{D}_2 = \mathbf{I}_n$, by definition $\lambda_j \in]0 : 1[$. The CSP algorithm is detailed below :
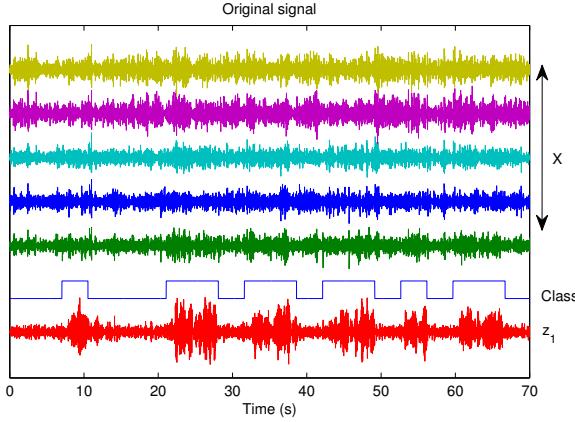
Figure 1. Illustration of CSP effect over filtered signal (at the bottom) and original signal (at the top). CSP maximizing the variance of the filtered signal under one condition while minimizing it for the opposite condition.

Algorithm 1 CSP Algorithm

Input: a set of I preprocessed EEG trials $\mathbf{X}_i$ with known class $Y_i \in \{1, 2\}$.

Output: $\mathbf{W}$ projection matrix

1: **for** $i = 1$ to I **do**
2:     $\mathbf{P}_{\mathbf{X}_i} = \text{Cov}(\mathbf{X}_i)$ {Compute Sample Covariance matrix}
3: **end for**
4: $\mathbf{P}_1 = \mathfrak{A}(\mathbf{P}_{\mathbf{X}_i}, Y_i = 1)$ {Arithmetic mean of covariance matrices for class 1}
5: $\mathbf{P}_2 = \mathfrak{A}(\mathbf{P}_{\mathbf{X}_i}, Y_i = 2)$ {Arithmetic mean of covariance matrices for class 2}
6: $(\mathbf{P}_1 + \mathbf{P}_2)^{-1}\mathbf{P}_1 = \mathbf{W}\mathbf{D}_1\mathbf{W}^T$ {Eigenvalue decomposition}
7: **return** $\mathbf{W}$

CSP algorithm gives us N spatial filters so $\mathbf{W} \in \Re^{N \times N}$, however it is usual to select a subset of the $J < N$ most relevant spatial filters and let the classification process operate on features extracted from the J filtered signals [4].

Features are naturally related to variance of the filtered signals with the usual choice :

$$\mathbf{F_X} = \begin{pmatrix} \log\left(\text{Var}(\mathbf{w}_1^T\mathbf{X})\right) \\ \vdots \\ \log\left(\text{Var}(\mathbf{w}_J^T\mathbf{X})\right) \end{pmatrix} = \begin{pmatrix} \log\left(\text{Var}(\mathbf{z}_1)\right) \\ \vdots \\ \log\left(\text{Var}(\mathbf{z}_J)\right) \end{pmatrix} \quad (2)$$

We will see in next sections that such feature choice is quite natural according to space topology.

III. CSP INTERPRETED VIA DIFFERENTIAL GEOMETRY

A. Differential geometry framework

1) Distance: A Riemannian manifold is a differentiable manifold in which the tangent space at each point is a finite-dimensional Euclidean space. We denote by $S(N) = \{\mathbf{S} \in M(N), \mathbf{S}^T = \mathbf{S}\}$ the space of all $N \times N$ symmetric matrices in the space of square matrices and denote by $P(N) = \{\mathbf{P} \in$

$S(N), \mathbf{P} > 0\}$ the set of all $N \times N$ symmetric positive-definite (SPD) matrices.

$N \times N$ spatial covariances matrices, which are used in CSP algorithm, are by construction SPD and therefore lie in $P(N)$. According to the topology of $P(N)$, rigorous manipulation of covariance matrices must be done through Riemannian geometry.

The Riemannian distance between two SPD matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ in $P(n)$ is given by [5] :

$$\delta_R(\mathbf{P}_1, \mathbf{P}_2) = \|\text{Log}\left(\mathbf{P}_1^{-1}\mathbf{P}_2\right)\|_F = \left[\sum_{n=1}^{N} \log^2 \beta_n\right]^{1/2} \quad (3)$$

where $\beta_n, n = 1\ldots N$ are the real eigenvalues of $\mathbf{P}_1^{-1}\mathbf{P}_2$. In Eq. (3), $\text{Log}(\mathbf{P})$ is the logarithm of the matrix $\mathbf{P}$, defined in [6] from the diagonalization of $\mathbf{P}$.

Riemannian distance has the following properties :

- $\delta_R(\mathbf{P}_1, \mathbf{P}_2) = \delta_R(\mathbf{P}_2, \mathbf{P}_1)$
- $\delta_R(\mathbf{P}_1, \mathbf{P}_2) = \delta_R(\mathbf{P}_1^{-1}, \mathbf{P}_2^{-1})$
- let $\mathbf{W} \in Gl(N)$ be an invertible matrix, $\delta_R(\mathbf{P}_1, \mathbf{P}_2) = \delta_R(\mathbf{W}^T\mathbf{P}_1\mathbf{W}, \mathbf{W}^T\mathbf{P}_2\mathbf{W})$

The third property is very important, since it implies that the space of SPD matrices is invariant by projection. It allows us to manipulate such space with tools like PCA without incidence over the distance.

2) Mean: It is well known that in Euclidean space, the arithmetic mean of I given SPD matrices $\mathbf{P}_1, \ldots, \mathbf{P}_I$ can be defined as follows :

$$\mathfrak{A}(\mathbf{P}_1, \ldots, \mathbf{P}_I) = \underset{\mathbf{P} \in P(N)}{\text{argmin}} \sum_{i=1}^{I} \delta_E^2(\mathbf{P}, \mathbf{P}_i) = \frac{1}{I}\sum_{i=1}^{I}\mathbf{P}_i \quad (4)$$

with the Euclidean distance between two matrices $\mathbf{M_1}$ and $\mathbf{M_2}$ in $M(N)$ given by the Froebenius norm [5] :

$$\delta_E(\mathbf{M}_1, \mathbf{M}_2) = \|\mathbf{M}_1 - \mathbf{M}_2\|_F \quad (5)$$

where $\|X\|_F = tr\left(XX^T\right)^{1/2}$.

The mean in the Riemannian sense, i.e. associated with the metric of (3), of I given SPD matrices $\mathbf{P}_1, \ldots, \mathbf{P}_I$ is defined according to [5] :

$$\mathfrak{G}(\mathbf{P}_1, \ldots, \mathbf{P}_I) = \underset{\mathbf{P} \in P(N)}{\text{argmin}} \sum_{i=1}^{I} \delta_R^2(\mathbf{P}, \mathbf{P}_i) \quad (6)$$

There is no closed-form expression for such mean computation and optimisation algorithms should be employed. Refer to [6] for efficient algorithmic implementation.

B. CSP revisited

1) Class-related mean covariance matrices: Referring to CSP algorithm, class-related mean covariance matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ usually obtained using arithmetic mean. As explained in Section III-A1, the topology of the space of covariance matrices follows Riemannian geometry. In order to obtain a rigorous estimation of class-related mean covariance matrices,

a simple and effective improvement is to change the Euclidean mean $\mathfrak{A}\,(.)$ for the Riemannian mean $\mathfrak{G}\,(.)$.

However Eq. (6) shows its limitations in term of computation time when the number of sensor is large (> 50). In this case, an approximation of Riemannian metrics like log-Euclidian metrics can be used instead [7]

2) Spatial filter selection: All CSP filters are not relevant for subsequent classification. The discriminative power of a spatial filter $\mathbf{w}_j$ is linked with its associated eigenvalue λ_j in :

$$(\mathbf{P}_1 + \mathbf{P}_2)^{-1}\mathbf{P}_1\mathbf{w}_j = \lambda_j\mathbf{w}_j$$

It is common to select an arbitrary number of filters, taking for example the filters which correspond to the three highest and the three lowest values of λ_j [1]. The most usual criterion consists on keeping the spatial filters which maximize the following quantity :$|\lambda_j - 0.5|$ [1].

However some smarter criterion can be established. Indeed if we look at the Riemannian distance between the two class-related mean covariance matrices, the full ($J = N$) CSP projection matrix $\mathbf{W}$ being invertible, it holds :

$$\delta_R(\mathbf{P}_1, \mathbf{P}_2) = \delta_R(\mathbf{W}^T\mathbf{P}_1\mathbf{W}, \mathbf{W}^T\mathbf{P}_2\mathbf{W}) = \delta_R\,(\mathbf{D}_1, \mathbf{D}_2)$$

Since $\mathbf{D}_1 + \mathbf{D}_2 = \mathbf{I}_N$, it holds :

$$\boxed{\delta_R(\mathbf{P}_1, \mathbf{P}_2) = \sqrt{\sum_{j=1}^{N} \log^2\left(\frac{\lambda_j}{1 - \lambda_j}\right)}} \quad (7)$$

Thus Riemannian distance between the two class-related mean matrices is directly linked with the set of CSP eigenvalue λ_j. Let denote $\mathbf{d}_R$ a $N \times 1$ column vector with elements $(\mathbf{d}_R)_j = |\log\left(\frac{\lambda_j}{1-\lambda_j}\right)|$. It holds

$$\delta_R(\mathbf{P}_1, \mathbf{P}_2) = \|\mathbf{d}_R\|_2 \quad (8)$$

Consequently every spatial filter $\mathbf{w}_j$ supports a part of the total distance through its weight $(\mathbf{d}_R)_j$. It is interesting to observe the function $f_R(\lambda_j) = (\mathbf{d}_R)_j^2$, see Fig. 2.
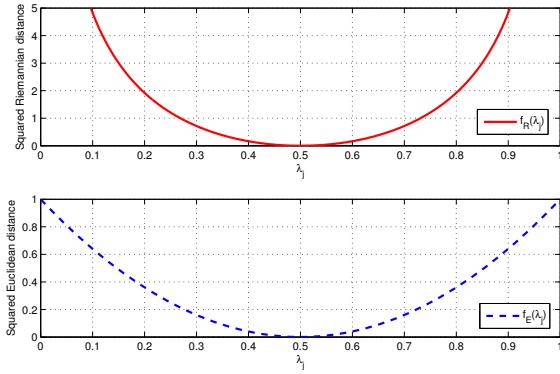

Figure 2. Evolution of squared distance supported by spatial filter $\mathbf{w}_j$ as a function of λ_j (solid line: Riemann , dashed line: Euclidean)

As it can be observed, the further is the eigenvalue from the symmetric point $1/2$, the more the distance supported

by the spatial filter between the two matrices is important. If we consider the distance between the class means as the most important factor to discriminate the two classes, the Riemannian distance gives us a tool to control which spatial filter should be chosen and its contribution to the whole distance. The function $f_R(\lambda_j)$ yields directly the spatial filters according to the intrinsic distance between $\mathbf{P}_1$ and $\mathbf{P}_2$. The sum across the different values of $f_R(\lambda_j)^2$ enable us to automatically choose the optimal number of spatial filters.

Using this approach, we can also define a similar criterion with Euclidian distance. In the projected space, $\mathbf{D}_1$ and $\mathbf{D}_2$ are the class-related mean matrices. We want to maximise the distance between the two classes in the projected space. If we take the Euclidean distance, we obtain the following statement :

$$\delta_E(\mathbf{D}_1, \mathbf{D}_2) = \|\mathbf{D}_1 - \mathbf{D}_2\|_F = 2\sqrt{\sum_{j=1}^{N}\left(\lambda_j - \frac{1}{2}\right)^2} \quad (9)$$

Let denote $\mathbf{d}_E$ a $N \times 1$ column vector with elements $(\mathbf{d}_E)_j = 2|\lambda_j - 0.5|$ It holds

$$\delta_E(\mathbf{D}_1, \mathbf{D}_2) = \|\mathbf{d}_E\|_2 \quad (10)$$

$\mathbf{d}_E$ represents a decomposition of Euclidean distance between $\mathbf{D}_1$ and $\mathbf{D}_2$ over the different spatial filters. It is interesting to study the evolution of the function $f_E(\lambda_j) = (\mathbf{d}_E)_j^2$. This function has a slower increase compared to its Riemannian counterpart.

Finally, the Riemannian approach provides us an estimation of the distance between the class-related matrices which is supported by each spatial filter. As we can see Section III-B3, the Riemannian distance is directly linked to the distance in feature space and therefore gives us an estimation of classification performance. That being not the case of Euclidean approach, even though it is almost equivalent, we would always prefer the use of Riemannian distance.

3) Feature extraction: CSP spatial filtering is followed by feature extraction, as mentioned in Section II. In this case, only the log-variance of each signal $\mathbf{z}_j$ is considered. In our setting, this amounts to approximate the covariance matrix $\mathbf{P_Z} = \mathrm{Cov}(\mathbf{Z})$ by its diagonal counterpart $\mathbf{D_Z} = \mathrm{diag}\,(\mathrm{diag}\,(\mathbf{P_Z}))$. In most cases, this should not be a bad approximation because for a given trial covariances matrices $\mathbf{P_Z}$ are statistically close either to $\mathbf{D}_1$ or $\mathbf{D}_2$ and so $\mathbf{P_Z}$ should be almost diagonal.

In terms of distance, the approximation yields to :

$$\begin{aligned}\delta_R\,(\mathbf{P}_1, \mathbf{P_X}) &= \delta_R\left(\mathbf{W}^T\mathbf{P}_1\mathbf{W}, \mathbf{W}^T\mathbf{P_X}\mathbf{W}\right) \\ &= \delta_R\,(\mathbf{D}_1, \mathbf{P_Z}) \simeq \delta_R\,(\mathbf{D}_1, \mathbf{D_Z})\end{aligned}$$

which is expressed explicitly as the vector norm of :

$$\begin{pmatrix} \log\left(\frac{(\mathbf{D_Z})_{11}}{(\mathbf{D}_1)_{11}}\right) \\ \vdots \\ \log\left(\frac{(\mathbf{D_Z})_{NN}}{(\mathbf{D}_1)_{NN}}\right) \end{pmatrix} = \begin{pmatrix} \log\left(\mathrm{Var}(\mathbf{z}_1)\right) - \log\left((\mathbf{D}_1)_{11}\right) \\ \vdots \\ \log\left(\mathrm{Var}(\mathbf{z}_N)\right) - \log\left((\mathbf{D}_1)_{NN}\right) \end{pmatrix}$$

Usually we use $J < N$ spatial filters, so the previous vector will be truncated. However, if the spatial filters selection is

correctly done, this truncation can be interpreted as denoising. If spatial filters are selected in order to conserve an arbitrary percentage of total distance between mean matrices, for example about 99%, we can consider that this truncation will discard irrelevant information.

It is interesting to observe that the log variance of the $\mathbf{z}_j$'s appears in the previous vector. Remember Eq. (2), log variance is the usual choice for features. The feature vector $\mathbf{F_X}$, which describes the EEG trial $\mathbf{X}$ in front of the classification algorithm, is directly linked to an approximation of Riemannian distance :

$$\delta_R\left(\mathbf{P}_1, \mathbf{P_X}\right) \simeq \|\mathbf{F_X} - \mathbf{F}_1\|_2 \qquad (11)$$

The feature space represents an approximate decomposition of Riemannian distance between the test covariance matrix $\mathbf{P_X}$ and the class-wise mean covariance $\mathbf{P}_1$. Conversely, **the Riemannian distance represents an approximation of the Euclidean distance in the Euclidean feature space**, in the context of CSP. Thus, the spatial filters selection based on the highest Riemannian distance insures implicitly a maximal separability of classes in feature space.
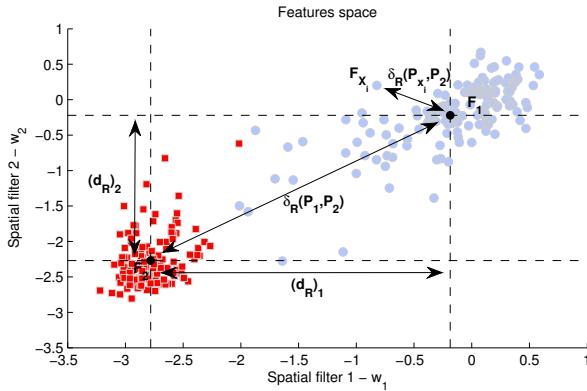

Figure 3.   Feature Space, class-related means and distances

An illustration of feature space is shown in Fig. 3. Each element of this space can be linked to an approximation of Riemannian distance, in the original space of covariance matrices.

IV. RESULTS

A. Datasets

In order to evaluate classification performances, we benchmark the proposed methods, with respect to the standard CSP method [1] with the following signal processing chain : frequency filtering, **CSP spatial filtering**, log-variance feature extraction and finally Fisher's LDA classification, as described in section II.

Datasets IVa of BCI competition III are used for analysis [8]. In our work, the goal is not to achieve the best performance but to compare different improvements of CSP method using the Riemannian framework. Thus no frequency filter is optimized nor electrode subset selection so that performances

could be less than those observed on the BCI competition or in the literature.

Only 9 electrodes are used : F_3, F_Z, F_4, C_3, C_Z, C_4, P_3, P_Z, P_4. This electrode subset represents a typical case of an every-day use BCI. A general $[10, 30]$ Hz band-pass filter has been applied on the original EEG signals for all subjects.

This dataset is composed by 5 subjects who performed 280 trials of right-hand or right-foot motor imagery. It was provided by Fraunhofer FIRST, Intelligent Data Analysis Group. leave-one-out cross-validation is used to evaluate properly the performances.

B. CSP results

When performing a CSP, the first step is to observe the so called spatial patterns. Spatial patterns are obtained by inverting the transpose of the full projection matrix :

$$\mathbf{A} = \left(\mathbf{W^T}\right)^{-1}$$

Spatial patterns $\mathbf{a}_j$ are the column vectors of matrix $\mathbf{A}$, and can be viewed as a correlation map between original electrodes signal $\mathbf{X}_i$ and the spatially filtered signal $\mathbf{z}_j$. Further informations about spatial patterns can be found in [1].

Fig. 4 shows us the first spatial filter and the corresponding spatial pattern for user al. We can see that the maximum weight of the spatial filter is for electrodes C_3, which covers the area dedicated to the right-hand movement. The spatial pattern is also consistent with the mental task of the motor imagery and the contralateral operation of the motor cortex.
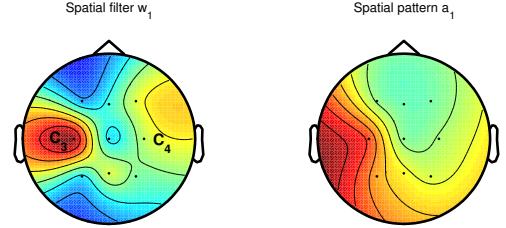

Figure 4.   First spatial filter $\mathbf{w}_1$ and the corresponding spatial pattern $\mathbf{a}_1$ for user al.

Fig. 5 shows the percentage of distance covered by the subset of filters. We see that 95% of total distance is covered by the 2-3 first spatial filters, and at least 99% of distance between the two matrices is covered with 5-6 filters.

Fig. 6 gives the repartition and weights, in terms of percentage of total squared distance, of the λ_j's for the user al. The best filter $\mathbf{w}_1$ associated with λ_1 covers about 60% of total squared distance. The use of Euclidean distance leads to underestimate the contribution of $\mathbf{w}_1$ and overestimate the others.

C. Classification results

In this work, we have proposed different ameliorations of CSP algorithm : they are implemented both separately and
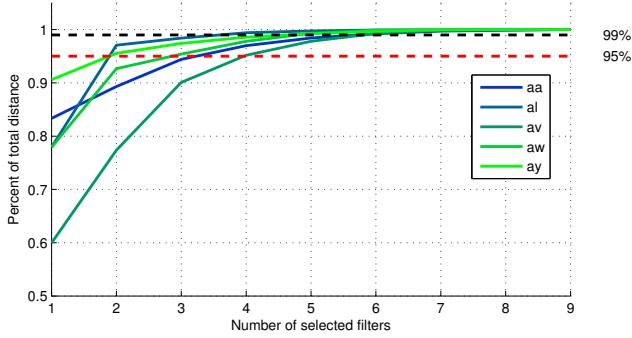
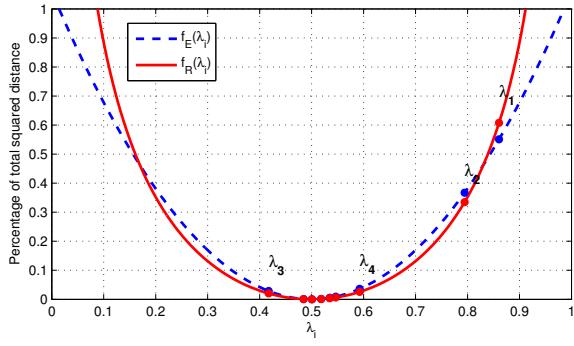Figure 5. Evolution of total distance as a function of the number of filters

Figure 6. Percentage of squared distance supported by the spatial filters and corresponding eigenvalues for user al.

all together. They are compared with an implementation of standard CSP algorithm.

The first improvement, named CSP_+, consists in the Riemannian mean computation as described in section III-B1.

The second one, named CSP_{++}, adds to CSP_+ an automatic spatial filter selection with a threshold of 99% of total distance as described in section III-B2.

For state-of-the-art CSP and CSP_+, we select the six best spatial filters according to classical criterion $|\lambda_i - 0.5|$. Table I shows the results in terms of classification error rates and number of selected filter for CSP_{++} :

Table I
CLASSIFICATION ERROR RATES IN LEAVE-ONE-OUT CROSS-VALIDATION AND NUMBER OF SELECTED FILTERS (SECOND COLUMN OF CSP_{++})

User	CSP	CSP_+	CSP_{++}
aa	24.2	**23.5**	**23.5**|6
al	**2.8**	3.2	3.2|4
av	35	**31.7**	**31.7**|6
aw	6	6.7	**5.3**|5
ay	**8.9**	9.6	9.2|5
Mean	15.4 ± 13.7	14.9 ± 12.2	$\mathbf{14.6 \pm 12.4}$

Classification error rates for CSP_+ and CSP_{++} methods shows an interesting decrease, due to a better estimation of mean matrices and a smarter selection of spatial filter.

For CSP_+ methods, the whole improvement is owned by the Riemannian mean estimation. For predisposed subjects i.e.

subject al, aw and ay, the improvements are not significant but for aa and av the ameliorations are important. Indeed, for predisposed subjects the intrinsic distance between the two class-related mean covariance matrices is large enough to compensate for approximation due to a poor estimation of class mean.

CSP_{++} shows a slight improvement compared to CSP_+. This is due to a better spatial filters selection. Since the number of kept filter does not change for aa and av, only the other subjects support the improvement. For these subjects, the distance supported by the spatial filters grows quickly and leads to a difficult choice of filters as it can be observed in Fig. 6. The automatic filter selection will be very helpful in this case.

V. CONCLUSION

We have shown that the Common Spatial Pattern algorithm can be viewed and interpreted through Riemannian geometry. This point of view leads to two possible enhancements of the classical CSP method.

First, the arithmetic mean, which estimates the intrinsic mean of covariance matrices is non-optimal in the sense of topology of symmetric-positive-definite matrices space. This can be solved using a tailored mean [6].

Second, an efficient and automatic solution for selection of spatial filters was presented. This can be achieved directly using the eigenvalues given by CSP algorithm and need no extra computation. However this method does not take into account the classes dispersion. Future works will provide a solution which includes data dispersion around the mean, like t-statistics, instead of distance between mean matrices.

Third, the feature space obtained by log variance extraction, and used for classification, actually describes an approximation of Riemannian distance between covariance matrices.

Finally, the combination of these improvements appears to produce better results than classical CSP due to a better handling of space of covariances matrices.

REFERENCES

[1] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller. Optimizing spatial filters for robust EEG Single-Trial analysis. *Signal Processing Magazine, IEEE*, 25(1):41–56, 2008.
[2] B. D. Van Veen, W. van Drongelen, M. Yuchtman, and A. Suzuki. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Bio-Medical Engineering*, 44(9):867–880, September 1997.
[3] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda. Recipes for the linear analysis of EEG. *NeuroImage*, 28:326—341, 2005.
[4] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 4, 2007.
[5] M. Moakher. A differential geometric approach to the geometric mean of symmetric Positive-Definite matrices. *SIAM J. Matrix Anal. Appl.*, 26(3):735–747, 2005.
[6] P. T. Fletcher and S. Joshi. Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*, pages 87–98. 2004.
[7] 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*, 29(1):328–347, 2007.
[8] BCI competition III, dataset IVa. www.bbci.de/competition/iii/index.html.