ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

BCI Signal Classification using a Riemannian-based kernel

Alexandre Barachant[1], Stéphane Bonnet[1], Marco Congedo[2] and Christian Jutten[2]

1- CEA, LETI, DTBS/STD/LE2S,
17 rue des Martyrs, F-38054 Grenoble, France.

2- Team ViBS (Vision and Brain Signal Processing),
GIPSA-lab, CNRS, Grenoble Universities.
Domaine Universitaire, F-38402 Saint Martin d'Hères, France.

Abstract. The use of spatial covariance matrix as feature is investigated for motor imagery EEG-based classification. A new kernel is derived by establishing a connection with the Riemannian geometry of symmetric positive definite matrices. Different kernels are tested, in combination with support vector machines, on a past BCI competition dataset. We demonstrate that this new approach outperforms significantly state of the art results without the need for spatial filtering.

1 Introduction

Brain-Computer Interfaces (BCIs) based on motor imagery have been well studied in the literature. For this type of BCI, the electrophysiological source of BCI control is based on spontaneous signals induced by voluntary changes in the dynamics of brain oscillations such as event related (de)synchronization (ERD/ERS) [1]. Motor Imagery (MI) is often used where movements are imagined by the subjects, which results in the activation of dedicated cortical areas (sensorimotor cortex) in given frequency bands. Thus this type of BCI is well suited for asynchronous BCI applications since the user can potentially perform actions without any stimulus from the exterior world.

The standard approach in MI-based EEG signal classification is to perform band-pass filtering, spatial filtering and linear classification, generally using Fisher's Linear Discriminant Analysis (LDA). The spatial filtering can be seen as a data-driven dimension reduction method that aims at promoting variance differences between two conditions. Sample covariance matrices are generally handled in the Euclidean space without considerations about the curvature of the space of Symmetric Positive Definite (SPD) matrices.

This paper provides a simple way to take into account the Riemannian geometry for EEG signal classification. This approach has been successfully applied on radar signal processing and image processing [2]. A new kernel is derived by establishing a connection with the Riemannian geometry of SPD matrices. This kernel is tested in combination with support vector machines, although we could have applied the kernel trick to other classifiers like logistic regression. Encouraging results are presented that demonstrate the potential benefit of the approach. Another advantage of the presented method is that it can directly be applied without the need for spatial filtering [3]. In section 2, the kernel is

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

introduced based on the Riemannian geometry. In section 3, results on BCI competition dataset are provided.

2 A new kernel for symmetric positive definite matrices

2.1 Introduction

EEG signals are often analysed on short-time segments where the signal is assumed to be wide-sense stationary. Let $\mathbf{X} \in R^{E \times T}$ be a trial of imaginary movement; E being the number of electrodes and T the epoch duration in number of samples. We further assume that the different EEG signals have been band-pass filtered by the acquisition system, usually in the μ and β frequency bands in our application.
For each trial $\mathbf{X}_p$ of known class $y_p \in \{-1, 1\}$, one can estimate the covariance matrix of the EEG random signal by the $E \times E$ sample covariance matrix (SCM) : $\mathbf{C}_p = 1/(T-1)\ \mathbf{X}_p \mathbf{X}_p^T$. However note that SCM is sensitive to outliers and either robust covariance estimation techniques or regularization can be applied to improve the estimation.

It is common practice in MI-based BCI to use spatial filtering for dimension reduction and variance enhancement between EEG trials coming from different classes [4]. The Common Spatial Patterns method (CSP) is for instance successfully applied as a mean to extract relevant features for the classification of EEG trials recorded during two motor imagery tasks. This technique aims at simultaneously diagonalizing the two intra-class covariance matrices obtained in the two conditions. This observation motivated us to investigate the direct use of spatial covariance matrix as input feature for EEG-based BCI signal classification.

2.2 SVM formulation

Support Vector Machine (SVM) is a popular linear classifier in BCI applications [5]. Given a set of labelled feature vectors $\{(\mathbf{x}_p, y_p)\}_{p=1}^P$, this classification technique seeks to linearly separate data by finding an hyperplane (with normal vector $\mathbf{w}$) that maximizes the margin, i.e. the distance between the hyperplane and the nearest points from each class, called support vectors. We refer to [6] and references herein for a detailed discussion on SVM.
The decision function will be based on the sign of :

$$h(\mathbf{x}) = b + \sum_{p=1}^{P} \alpha_p y_p \langle \mathbf{x}_p, \mathbf{x} \rangle = b + \langle \mathbf{w}, \mathbf{x} \rangle \quad (1)$$

where $\langle ., . \rangle$ denotes the Euclidean scalar product between vectors. The term b is a bias and the $\{\alpha_p\}_{p=1}^P$ are the Lagrangian multipliers associated to the dual optimization problem [6]. Both quantities are estimated by quadratic programming. Most of the α_p's are null except for the support vectors. If data are not linearly separable in their native space, a mapping can be applied on the feature

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

vector $\mathbf{x}$ to another (high-dimensional) transformed space. The transformation ϕ is generally non-linear and the decision function can be rewritten as :

$$h(\mathbf{x}) = b + \sum_{p=1}^{P} \alpha_p y_p \langle \phi(\mathbf{x}_p), \phi(\mathbf{x}) \rangle_{\mathcal{H}} \tag{2}$$

where $\mathcal{H}$ is a reproducing kernel Hilbert space where the dot product is defined. The associated kernel $k(.,.)$ is defined by : $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}}$.
In most SVM applications, the ϕ-function is not explicitly expressed and solely the kernel is used. A common kernel is the Gaussian radial basis function : $k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left[-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right]$. SVM is known to possess good generalization properties and to perform well in high-dimensional feature space. Finally the margin maximization cost function is often penalized with the introduction of slack variables whose effects are controlled by the hyper-parameter λ. This allows to enable soft margins where some feature vectors can be reasonably misclassified.

2.3 SVM applied on covariance matrix

In order to use sample covariance matrix $\mathbf{C}$ as input feature to SVM, a simple choice consists in vectorizing it : $\mathbf{x} = \text{vec}(\mathbf{C})$ and apply linear SVM classification on the resulting set [7]. A more powerful approach consists in taking account the Riemannian geometry of the space of covariance matrices, i.e. the space of symmetric positive definite (SPD) matrices. Due to space constraints, we focus on the main points needed for the proposed algorithm.

The space of SPD $E \times E$ square matrices $P(E)$ forms a differentiable manifold $\mathcal{M}$ of dimension $E^\star = E\,(E{+}1)/2$. At each point $\mathbf{C}$ (i.e. each covariance matrix), a scalar product can be defined in the associated **tangent space** $\mathcal{T}_{\mathbf{C}}\mathcal{M}$. The manifold is locally homomorphic to the Euclidean tangent space and distance computations in the manifold can be well approximated by distance computations in the tangent space.
Let $\mathbf{S}_1$ and $\mathbf{S}_2$ be two tangent vectors (i.e. two symmetric matrices in our case), the scalar product in the tangent space at $\mathbf{C}$ can be defined by the relation :

$$\langle \mathbf{S}_1, \mathbf{S}_2 \rangle_{\mathbf{C}} = \text{tr}(\mathbf{S}_1 \mathbf{C}^{-1} \mathbf{S}_2 \mathbf{C}^{-1}). \tag{3}$$

Furthermore, we can define the logarithmic map to project locally all covariance matrices $\{\mathbf{C}_p\}_{p=1}^{P}$, onto the tangent plane by :

$$\mathbf{S}_p = \text{Log}_{\mathbf{C}}(\mathbf{C}_p) = \mathbf{C}^{1/2} \text{logm}\left(\mathbf{C}^{-1/2} \mathbf{C}_p \mathbf{C}^{-1/2}\right) \mathbf{C}^{1/2} \tag{4}$$

where logm denotes the logarithm of a matrix. From this connection with Riemannian geometry, we define the mapping function as : $\phi(\mathbf{C}) = \text{Log}_{\mathbf{C}_{\text{ref}}}(\mathbf{C})$.
This choice yields after some manipulations to the kernel :

$$\begin{aligned} k(\text{vec}(\mathbf{C}_i), \text{vec}(\mathbf{C}_j)) &= \langle \phi(\mathbf{C}_i), \phi(\mathbf{C}_j) \rangle_{\mathbf{C}_{\text{ref}}} \qquad (5) \\ &= \text{tr}\left[\text{logm}\left(\mathbf{C}_{\text{ref}}^{-1/2} \mathbf{C}_i \mathbf{C}_{\text{ref}}^{-1/2}\right) \text{logm}\left(\mathbf{C}_{\text{ref}}^{-1/2} \mathbf{C}_j \mathbf{C}_{\text{ref}}^{-1/2}\right)\right]. \end{aligned}$$

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

Here $\mathbf{C}_{\text{ref}}$ is a free parameter that defines the point in SPD space where the tangent plane is computed. A common choice is to use the geometric mean of the P labelled covariance matrices $\{\mathbf{C}_p\}_{p=1}^P$ [8].
Alternatively, the choice $\mathbf{C}_{\text{ref}} = \mathbf{I}_E$ yields to the simplified log-Euclidean kernel :

$$\begin{aligned} k(\text{vec}(\mathbf{C}_i), \text{vec}(\mathbf{C}_j)) &= \text{tr}\left[\text{logm}\left(\mathbf{C}_i\right)\text{logm}\left(\mathbf{C}_j\right)\right] \\ &= \langle \text{logm}\left(\mathbf{C}_i\right), \text{logm}\left(\mathbf{C}_j\right)\rangle_F \end{aligned} \tag{6}$$

where subscript F stands for Frobenius. This kernel is investigated in [9].

3 Experiments

In order to evaluate the performance of the proposed method, we have compared it to the classical signal processing chain in asynchronous MI-based BCI. The standard approach consists in band-pass filtering, spatial filtering (using CSP approach), log-variance feature extraction and Fisher's LDA classification [4]. This method is compared to linear SVM applied on vectorized covariance matrices and to kernel-based SVMs according to equation 5. Two tangent planes have been considered : one at identity $\mathbf{I}_E$ and the other one at the geometric mean of all covariance matrices $\{\mathbf{C}_p\}_{p=1}^P$.

3.1 Dataset

Dataset IIa of BCI competition IV are used for analysis. 22 electrodes are used (Fz, FC3, FC1, FCz, FC2, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1, CPz, CP2, CP4, P1, Pz, P2, POz). The reference electrode is located on the left mastoid. A $8-35$ Hz band-pass filter has been applied on the original signals for all subjects. This dataset is composed of 9 subjects who performed 576 trials of right-hand (RH), left-hand (LH), tongue (TO) and both feet (BF) motor imagery (i.e. 144 trials per class). Since CSP is designed for binary classification, we have evaluated the average performance per subject, for all 6 possible pairs of mental tasks : {LH/RH, LH/BF, LH/TO, RH/BF, RH/TO,

	SVM kernel Geometric Mean	SVM kernel Identity	SVM plain	CSP
LH/RH	**83.4**	82.8	76.9	81.4
LH/BF	**89.9**	89.5	82.2	86.9
LH/TO	**89.3**	88.8	84	87.3
RH/BF	**88.8**	88.2	81.1	86.3
RH/TO	**88.7**	88	83.4	85.7
BF/TO	**82.1**	81.4	74.9	80.8
mean	**87**	86.4	80.4	84.7

Table 1: Average classification accuracy across the 9 subjects for 6 pairs of mental tasks.

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

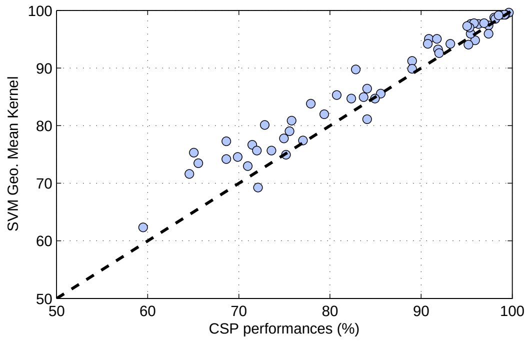

Fig. 1: Kernel SVM versus CSP classification accuracy comparison for each individual pair of mental tasks.

BF/TO}. We have used 30-fold cross-validation to evaluate the generalization performance of the algorithms and the significance of the classification accuracy.

3.2 Results

The hyperparameter λ is chosen equal to 10 in our experiments. As it can be observed from Table 1, kernel-based SVM approaches outperform both vectorized-SCM SVM method and CSP. A mean classification accuracy of 87% can be obtained whereas the CSP method is limited to 85%. Locating the tangent plane at the mean of all covariance matrices yields better results on average compared to the log-Euclidean choice of equation 6.

Comparing CSP and proposed kernel-based SVM on the individual sessions yields to Figure 1. It can be appreciated that except for few localized pairs of mental task, kernel-based SVM consistently outperforms CSP results. This remark is especially true for difficult binary classification cases (CSP performance below 80%) where the improvement brought by kernel-based SVM is significant.

We have also tested the significance of methods with respect to each other. Results are presented in Figure 2. This figure illustrates the p-values obtained for the 6 pairs of mental tasks. These values are obtained with a one-tailed dependent t-test (8 degrees of freedom) for paired samples, for test hypothesis $\{\mathcal{H}_0 : \mu_1 > \mu_2\}$. Both kernel-SVM proposed methods perform significantly better than the CSP ($p < 0.05$) for 4 out of 6 pairs of mental tasks. The performance of the two kernel-based SVMs differs significantly only for pair LH/BF. Nonetheless, the kernel using the tangent space at the geometric mean is always superior to the kernel using the tangent space at the identity. The slight increase of p-value in the RH/LH case could be futher reduced by optimizing the hyperparameter λ.

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence
and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

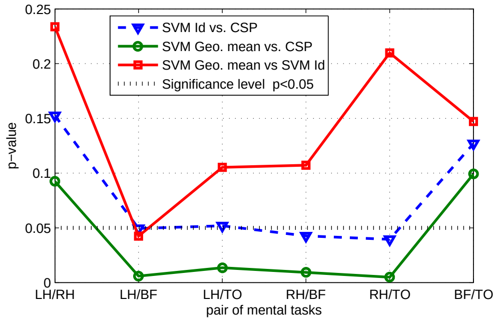

Fig. 2: P-values for the 6 pairs of mental tasks in a subject-independent manner. Hypothesis $\mathcal{H}_0 : \mu_1 > \mu_2$, one-tailed dependent t-test (8 df) for paired samples.

4 Conclusion

This paper has proposed a new kernel for directly handling covariance matrices in classification methods. The approach is tested on a BCI competition dataset and outperforms significantly the conventional CSP method. This kernel could be employed in different applications where covariance matrices are the main ingredients of the feature extraction process. We have demonstrated in this work that the spatial filtering of electrodes, could be avoided without loss of performance. A future work will investigate the online use of this algorithm and the location update of the tangent space between BCI sessions and the minimum number of trials required to properly estimate the covariance matrices.

References

[1] G. Pfurtscheller and F.H.L. da Silva, Event-related EEG/MEG synchronization and desynchronization: basic principles, *Clin. Neurophysiol.*, 11(110):1842-1857, 1999.

[2] O. Tuzel, F. Porikli and P. Meer, Pedestrian detection via classification on Riemannian manifolds, *IEEE Trans. PAMI*, 10: 1713–1727, 2008.

[3] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Riemannian geometry applied to BCI classification In LVA ICA-2010, St Malo, France, Sept. 27-30, pp. 629-636.

[4] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe and K.-R. Müller, Optimizing spatial filters for robust EEG single-trial analysis, *IEEE Signal Proc. Mag.*, 25(1):41-56, 2008.

[5] N. Jrad, M. Congedo, R. Phlypo et al., sw-SVM: sensor weighting support vector machines for EEG-based brain-computer interfaces, *J. Neur. Eng.*, 8:1-8, 2011.

[6] B. Schölkopf, and A.J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and beyond*, MIT Press, Cambridge, MA, 2002.

[7] J. Farquhar, A linear feature space for simultaneous learning of spatio-temporal filters in BCI, *Neural Networks*, 22:1278-1285, 2009.

[8] M. Moakher, A differential geometric approach to the geometric mean of symmetric positive-definite matrices, *SIAM J. Matrix Anal. Appl.*, 26:735-747, 2005.

[9] E.-Y. Wang, W. Guo, L.-R. Dai et al., Factor analysis based spatial correlation modeling for speaker verification In ISCSLP-2010, Tainan, China, Nov. 29-Dec. 3, pp. 166-170.