STATIONARY COMMON SPATIAL PATTERNS: TOWARDS ROBUST CLASSIFICATION OF NON-STATIONARY EEG SIGNALS

Wojciech Wojcikiewicz[†‡], *Carmen Vidaurre*[†], *Motoaki Kawanabe*[‡†]

[†]Technical University of Berlin, Franklinstr. 28 / 29, 10587 Berlin, Germany
[‡]Fraunhofer Institute FIRST, Kekuléstr. 7, 12489 Berlin, Germany

ABSTRACT

Brain-Computer Interfaces (BCIs) allow a user to control a computer application by brain activity as acquired, e.g., by EEG. A standard step in a BCI system is to project the EEG signals to a low-dimensional subspace using Common Spatial Patterns (CSP). However, non-stationarities in the data can negatively affect the performance of CSP, i.e. variation of the signal properties within and across experimental sessions coming from electrode artefacts, alpha or muscular activity, or fatigue may result in suboptimal projection directions. We alleviate this problem by regularizing CSP towards stationary subspaces and show that this especially increases classification accuracy of people who are not able to control a BCI i.e. have more than 30% of error. These users very often show non-stationarities in their EEG signals.

Index Terms— Brain-Computer Interface, Common Spatial Patterns, Non-Stationarity

1. INTRODUCTION

Brain-Computer Interface (BCI) systems aim to provide users control over a computer application by their brain activity. Many EEG-BCIs [1, 2] use the motor imagery paradigm for translating the user's intentions into commands. Commonly, subjects using these systems are asked to perform the imagination of movements with their hands, feet or mouth. Motor imagery alters the rhythmic activity that can be measured in the EEG over the sensorimotor cortex. The locations over the sensorimotor cortex are related to corresponding parts of the body. For example, left and right hand are localized in the contralateral hemisphere, i.e., right and left motor cortex, respectively. By using Common Spatial Patterns (CSP) [1] we can extract the relevant features from the EEG signal and thus distinguishing motor imagery tasks of different body parts.

We thank Klaus-Robert Müller for valuable discussions. This work was supported by the German Research Foundation (GRK 1589/1) and he European Union under the project TOBI (FP7-ICT-224631) and the Federal Ministry of Economics and Technology of Germany under the project THESEUS (01MQ07018). This publication only reflects the authors' views. Funding agencies are not liable for any use that may be made of the information contained herein.

One challenge of EEG-based BCIs is to solve the problem of lack of efficiency of BCI systems, which is that BCI control does not work for a non-negligible portion of users (estimated 15 to 30%), (c.f. Vidaurre et al. 2009 [3]). One reason for bad performance in BCI are non-stationarities in the EEG signal i.e. variation of the signal properties (e.g. covariances) within and across experimental sessions coming from electrode artefacts, alpha or muscular activity, or fatigue. Most machine learning algorithms implicitly assume stationarity in the data thus non-stationarities can negatively affect performance.

Recently, several approaches were performed to reduce the impact of non-stationarities in the data by co-adaptivity [3], channel selection [4], by including extra measurement (model-based) [1, 5] or by covariate shift adaptation [6]. In this paper we extend Common Spatial Patterns (CSP) to explicitly measure non-stationarities and regularize it towards stationary subspaces.

This paper is organized as follows. In Section 2 we present the stationary Common Spatial Patterns method (sCSP). After describing our experimental setup in Section 3, we compare the performance between CSP and sCSP and analyse the reasons for the performance gain on a specific subject. Section 5 concludes with a short summary and future research ideas.

2. STATIONARY COMMON SPATIAL PATTERNS

Common Spatial Patterns (CSP) has been widely used in BCI systems [2, 1] as they maximize the variance of signals of one class and at the same time minimize the variance of signals of another class, thus they are well suited to discriminate mental states that are characterized by ERD/ERS effects. The CSP spatial filter $\mathbf{w}$ can be obtained by maximizing / minimizing the Rayleigh coefficient

$$\max_{\mathbf{w}} \frac{\mathbf{w}^\top \mathbf{\Sigma}_+ \mathbf{w}}{\mathbf{w}^\top \{\mathbf{\Sigma}_+ + \mathbf{\Sigma}_-\} \mathbf{w}}, \tag{1}$$

where $\mathbf{\Sigma}_+$ and $\mathbf{\Sigma}_-$ are the average covariance matrices from class 1 and 2 respectively.

We extend CSP by regularizing it towards stationary subspaces. Our basic idea is the following:

- we consider local chunks of BCI data sequences in order to measure non-stationarity over time,
- we add a regularization term in the CSP optimization criterion 1 as invariantCSP [5, 7] in order to penalize non-stationary features.

Let $\mathbf{\Sigma}_+^{(k)}$ and $\mathbf{\Sigma}_-^{(k)}$ be the covariance matrices of the k-th chunk. A chunk is a collection of consecutive trials i.e. if the chunk size is one, we consider individual trials, otherwise we consider the mean covariance matrix of the trials in chunk k. Non-stationarity can be measured by comparing the covariance matrices of each chunks with the global average $\mathbf{\Sigma}_+$ and $\mathbf{\Sigma}_-$,

$$\mathbf{\Delta}_+^{(k)} \;:=\; \mathcal{P}\left(\mathbf{\Sigma}_+^{(k)} - \mathbf{\Sigma}_+\right), \tag{2}$$

$$\mathbf{\Delta}_-^{(k)} \;:=\; \mathcal{P}\left(\mathbf{\Sigma}_-^{(k)} - \mathbf{\Sigma}_-\right), \tag{3}$$

where $\mathcal{P}$ is an operator to make symmetric matrices be positive definite. More precisely, if a symmetric matrix $\mathbf{M}$ has eigen decomposition $\mathbf{M} = \mathbf{R}\,\mathrm{diag}(d_i)\,\mathbf{R}^\top$, the operator returns $\mathcal{P}(\mathbf{M}) = \mathbf{R}\,\mathrm{diag}(|d_i|)\,\mathbf{R}^\top$, i.e. the signs of all the negative eigenvalues are flipped. By doing so, the penalty term becomes (always) positive even in the case that power of a feature in the k-th chunk is smaller than its global average. We propose to use the averages of the difference matrices for regularization towards stationary subspaces

$$\overline{\mathbf{\Delta}}_+ \;:=\; \frac{1}{K}\sum_{k=1}^{K}\mathbf{\Delta}_+^{(k)}, \tag{4}$$

$$\overline{\mathbf{\Delta}}_- \;:=\; \frac{1}{K}\sum_{k=1}^{K}\mathbf{\Delta}_-^{(k)}. \tag{5}$$

In other words we project data in directions $\mathbf{w}$ computed as

$$\max_{\mathbf{w}} \frac{\mathbf{w}^\top\mathbf{\Sigma}_+\mathbf{w}}{\mathbf{w}^\top\{\mathbf{\Sigma}_+ + \mathbf{\Sigma}_- + \lambda(\overline{\mathbf{\Delta}}_+ + \overline{\mathbf{\Delta}}_-)\}\mathbf{w}}, \tag{6}$$

$$\max_{\mathbf{w}} \frac{\mathbf{w}^\top\mathbf{\Sigma}_-\mathbf{w}}{\mathbf{w}^\top\{\mathbf{\Sigma}_+ + \mathbf{\Sigma}_- + \lambda(\overline{\mathbf{\Delta}}_+ + \overline{\mathbf{\Delta}}_-)\}\mathbf{w}}. \tag{7}$$

The regularization terms in the denominators penalize non-stationary features, where the constant λ balances discriminativity (of the data at hand) and stationarity of features.

3. EXPERIMENTAL SETUP

In our experiments we use data recorded from 80 subjects performing motion imagery tasks with the left and right hand or with the feet. For each subject we select the best binary combination i.e. left hand vs. feet, right hand vs. feet or left hand vs. right hand in a calibration session consisting of 150 trials per combination. No feedback is provided in the calibration session. After that we perform a test session consisting of 300 trials using the best combination. In the test session we provide 1D feedback e.g. a cursor moves left and right on the screen based on the classification decision. All subjects in our study are BCI novices. We use recordings of 68 preselected electrodes, log-variance features, a Linear Discriminant Analysis (LDA) classifier and error rate or area over receiver operating characteristic (ROC) to measure performance. We select three directions of CSP / sCSP per class (i.e. use six-dimensional features), but also consider the case of using only one direction per class as this simplifies the visualization and analysis of results.

For sCSP we need to set the λ parameter and the chunk size. We select the best λ from the set of candidates $\{0, 0.1, 0.025, 0.05, 0.075, 0.1, 0.25, 0.5, 0.75, 1, 2.5, 5, 10\}$ by 5-fold cross-validation (CV) on the calibration data. Furthermore we consider chunk sizes of 1, 3, 5 and 8 and select the best combination of λ and chunk size using cross-validation for each subject individually. Apart from that we also run experiments with a priori fixed chunk size and select λ by cross-validation.

4. RESULTS

We compare the performance of CSP and sCSP using two different error measures, error rate and area over ROC. The latter measure make sure that the performance difference between CSP and sCSP is not only due to a bias shift but comes from a better feature representation[1]. Figure 1 shows a scatter plot of CSP and sCSP performances and we can see that our method in most cases lies below the black dotted line i.e. it performs better than CSP. This is especially obvious for subjects which perform poorly with CSP. We also see that there is not much difference between using error rate (top panels) and area over ROC (bottom panels) as error measure. This indicates that sCSP provides better separable features or in other words it is less affected by noise and non-stationarities than CSP.

Since we want to know whether sCSP is significantly better than CSP, we apply an one-sided t-test to the experiment which uses three CSP directions. Table 1 shows the p-values for different fixed chunk sizes and the case where chunk size was selected by cross-validation. Furthermore we divide the subjects into three groups according to their performances: good performers, medium performers and subjects lacking BCI efficiency. We clearly see that our method is significantly better than CSP in the group of subjects lacking BCI efficiency and in total irrespectively of the chunk size.

In the following we would like to analyse the reasons for the performance gain of subject 20. We consider the experiment with one CSP direction per class because it is better for visualization. Subject 20 has an error rate of 39.3 when using CSP and 19.3 when sCSP is used, we selected λ as 0.5 and a chunk size of 8 by cross-validation. Subject 20 uses motion imagery of left vs. right hand in order to discriminate

[1]Changes in bias can be alleviated by shifting the classification boundary, whereas a better separability of feature is much harder to obtain.

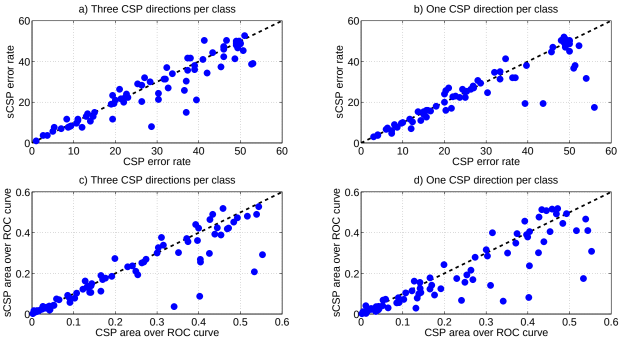

Fig. 1. Scatter plot of CSP and sCSP performances using error rate (top) or area over ROC curve measure (bottom) for three (left) and one (right) CSP direction per class.

Chunk size	Error rate			
	$0-15$	$15-30$	>30	all
CV selected	0.0599	0.6642	**0.0069**	**0.0068**
1	0.0567	0.5194	**0.0085**	**0.0066**
3	0.1749	0.3547	**0.0121**	**0.0087**
5	0.1568	0.6690	**0.0076**	**0.0091**
Chunk size	Area over ROC			
	$0-0.15$	$0.15-0.3$	>0.3	all
CV selected	0.3084	0.2337	**0.0122**	**0.0091**
1	0.3940	0.2629	**0.0151**	**0.0122**
3	0.3766	0.3001	**0.0271**	**0.0178**
5	0.2336	0.4183	**0.0133**	**0.0113**

Table 1. Overview of p-values for different chunk sizes and different error regions when using one-sided t-test with the hypothesis that sCSP performs better than CSP. Bold values are significant when $\alpha = 0.05$.

between classes. So the question is why do we have such a large improvement when using sCSP ?

In order to answer the question we at first visualize the training and test features and the classification boundary of CSP and sCSP (see Figure 3). Inspecting Figure 3, one can observe that the features of class 2 suffer a considerable change between calibration and test phase, whereas this is not the case for the stationary features. We also see that the separability of the sCSP features is much better than the separability of the CSP features.

If we want to understand why the sCSP projection is much better than the CSP one, we should look at the activation patterns of CSP and sCSP in Figure 2. Since subject 20 is performing a left vs. right hand motion imagery, we should see an activation in the right and left hemisphere respectively. This is exactly what we obtain when using sCSP (bottom), the pattern in the left panel corresponds to right-hand motor imagery and the one in the right panel to left hand motor imagery. Unfortunately, we do not get these patterns in the CSP case (top), instead of the right hand motion imagery pattern we obtain an artefact in electrode CCPC3 (according to extended 10-20 system of electrode placement). This explains the change of class 2 features in Figure 3 which was mentioned above as class 2 corresponds to right hand motion imagery and a CSP filter which concentrates on electrode artefact instead of the underlying right-hand motor imagery pattern will produce very non-stationary and noisy features. From that we conclude that subject 20 has such a high error rate when using CSP as one CSP filter is affected by an electrode artefact, thus is not optimal.

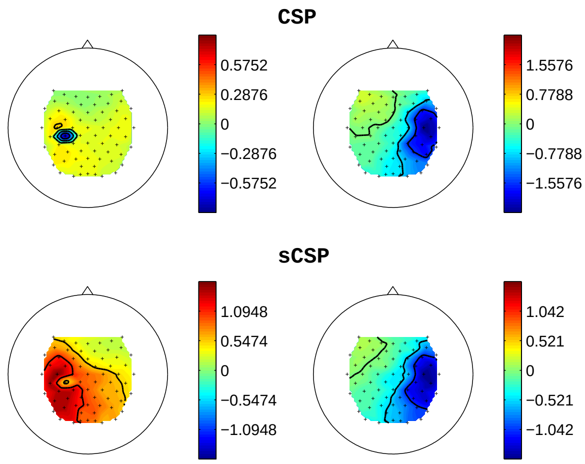

Fig. 2. Activation patterns of left hand (right panels) and right hand (left panels) motion imagery of subject 20. CSP pattern (top panels) contains electrode artefact at CCPC3, which is not present in sCSP pattern (bottom panels).

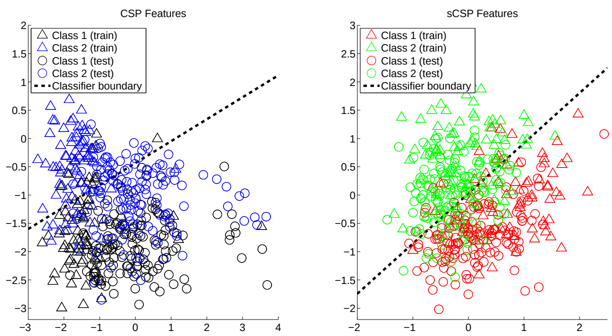

Fig. 3. Comparison of training (triangle) and test (circle) features of CSP (left) and sCSP (right) for subject 20.

We can prove our artefact hypothesis when analysing the band-passed filtered signal at CCPC3. Figure 4 shows the concatenated EEG signals at CCPC3 and CCPC4. Both sig-

λ	e_R	λ	e_R	λ	e_R
0 and 0.01	39.3	0.1	28	1	25.7
0.025	40.7	0.25	20	2.5	37.7
0.05	41.7	0.5	19.3	5	41.7
0.075	38.7	0.75	23.3	10	42.7

Table 2. Error rate e_R of subject 20 for different λ.

nals should look more or less similar, but CCPC3 (top panel) seems to be more non-stationary and it contains a clear outlier in trial 15 (around 0.5×10^4). This may be due to e.g. a loose electrode. So the reason why sCSP performs so well in subjects which perform poorly with CSP is simply that it is more robust to noise and non-stationarities in the data.

Although the performance of sCSP is not highly affected by the chunk size (as shown in Table 1), it is very sensitive to the selection of λ. Table 2 shows the test performances of different λ for subject 20. The performance curve has an U-shape with a minimum at $\lambda = 0.5$, which is also the λ value selected by cross-validation. The most popular value selected by cross-validation in all subjects is $\lambda = 0.25$. The selection frequencies of λ for all subjects can be seen in Figure 5.

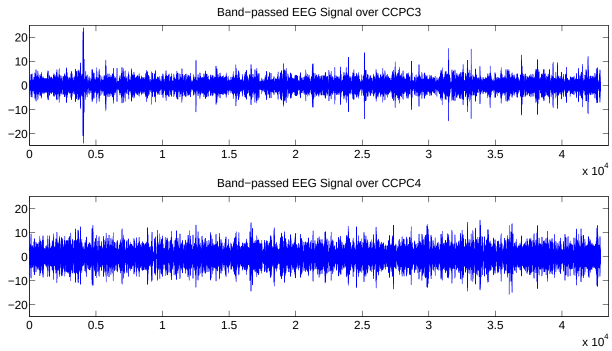

Fig. 4. Comparison of EEG signals over CCPC3 (top) and CCPC4 (bottom). The upper one contains an electrode artefact which deteriorate performance.

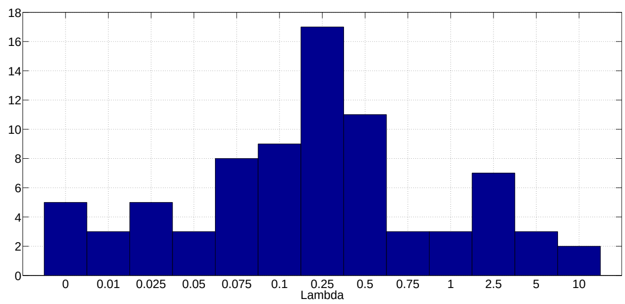

Fig. 5. Selection frequencies of λ using 5-fold CV.

5. CONCLUSION

We presented an extension of CSP which explicitly measures non-stationarities and regularizes the CSP directions towards stationary subspaces. We showed that sCSP significantly outperforms CSP in subjects with high error rates since their signals often contains a lot of noise and non-stationarities.

We also showed that sCSP is able to extract meaningful neurophysiological filters even when the data in use is noisy. We have also demonstrated that the stationary features suffer less shift than the ones computed using the usual CSP method, thus are easier to discriminate.

Furthermore, unlike other methods, such as invariantCSP, our approach is completely data-driven and does not require additional recordings or models of the expected change that occurs in the EEG.

In future research we would like to combine stationarity features like sCSP with classifier adaptation to further improve classification performance on non-stationary and noisy BCI data.

6. REFERENCES

[1] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K-R. Müller, "Optimizing spatial filters for robust eeg single-trial analysis," in *IEEE Signal Proc. Magazine*, 2008, pp. 581–607.

[2] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement," *IEEE Trans. Rehab. Eng.*, vol. 8, pp. 441–446, 1998.

[3] C. Vidaurre, C. Sannelli, K-R. Müller, and B. Blankertz, "Machine-learning based co-adaptive calibration," *Neural Comput.*, 2010, in press.

[4] M. Kawanabe, C. Vidaurre, B. Blankertz, and K-R. Müller, "A maxmin approach to optimize spatial filters for EEG single-trial classification," in *Proceedings of IWANN 09, Part I, LNCS*, 2009, pp. 674–682.

[5] B. Blankertz, M. Kawanabe R. Tomioka, F. U. Hohlefeld, V. Nikulin, and K-R. Müller, "Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing," in *Ad. in NIPS 20*, 2008, pp. 113–120.

[6] Y. Li, H. Kambara, Y. Koike, and M. Sugiyama, "Application of covariate shift adaptation techniques in brain-computer interfaces.," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 6, pp. 1318–24, 2010.

[7] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, A. Smola, and K-R. Müller, "Invariant feature extraction and classification in kernel spaces," in *Ad. in NIPS 12*, 2000, pp. 526–532.