

Machine Learning Methods of the Berlin Brain-Computer Interface

Carmen Vidaurre* Claudia Sannelli* Wojciech Samek**
Sven Dähne* Klaus-Robert Müller*,***

* *Berlin University of Technology, Machine Learning Dept. Marchstr.
23, 10587, Berlin, Germany.*

** *Fraunhofer HHI, Machine Learning Group, Dept. of Video Coding &
Analytics Einsteinufer 37, 10587 Berlin, Germany.*

*** *Dept. of Brain and Cognitive Engineering, Korea University,
Anam-dong, Seongbuk-ku, Seoul, Republic of Korea.*

Abstract: This paper is a compilation of the most recent machine learning methods used in the Berlin Brain-Computer Interface. In the field of Brain-Computer Interfacing, machine learning has been mainly used to extract meaningful features from noisy signals of large dimensionality and to classify them to transform them into computer commands. Recently, our group developed different methods to deal with noisy, non-stationary and high dimensional signals. These approaches can be seen as variants of the algorithm Common Spatial Patterns (CSP). All of them outperform CSP in the different conditions for which they were developed.

Keywords: Brain-Computer Interfacing, Motor Imagery, non-stationary analysis, electroencephalogram, multimodal analysis, adaptive systems.

1. INTRODUCTION

One relevant type of Human-Machine Interface is the Brain Computer Interface (BCI). It translates the intent of a subject measured from brain signals into control actions to command machines or computer applications. To measure the brain activity, the majority of BCI systems rely on electroencephalography (EEG), because it is relatively cheap, easy to acquire, minimally intrusive and does not involve any risks for the user. However, EEG has a very poor signal to noise ratio and is highly non-stationary. These are major challenges for BCI systems that limit their application in real-life settings. It has been shown that a non-negligible portion of the healthy population, estimated 30%, cannot achieve BCI control to an acceptable level Guger et al. (2003). This percentage may become as high as 100% for patients depending on their pathology. Early EEG-BCI efforts were based on neuro-feedback training on the part of the user that lasted on the order of days or even months, in Machine Learning-based systems it suffices to collect examples of EEG signals in a so-called calibration measurement during which the user is cued to perform a small set of mental tasks. This data is then used to adapt the system to the specific brain signals of each user (machine training) Blankertz et al. (2006). Typical algorithms used in this stage are Common Spatial Patterns (CSP) and linear classifiers Blankertz et al. (2008). CSP enhances the class differences of band-pass brain signals by optimized weighting of the sensors, assuming that the brain sources can be inferred with a linear inverse model. This step of adaptation, together with the individual selection of the motor imagery (MI) tasks to perform, is instrumental for effective BCI performance

*

despite the large inter-subject variability of the respective brain signals. However, CSP exhibits some drawbacks. It needs a considerable amount of training data to obtain robust spatial filters and avoid overfitting. CSP is a supervised algorithm whose number of parameters increases quadratically with the number of EEG channels involved (as the dimension of the covariance matrices increases), bearing a high risk of overfitting when using little training samples or in the presence of outliers. Also, it does not take into account the non-stationarities present in EEG data and it is not possible to combine it with different types of signals.

In this paper we present four machine-learning based approaches that have recently been developed to cope with these problems. In particular, Common Spatial Pattern Patches (CSPP) deal with the small sample size problem, stationary CSP (sCSP) allows finding more stationary filters, the divergence-based CSP framework permits the definition of robust filters and their regularization and Source Power Co-modulation (SPoC) finds components in the multivariate oscillatory signal (generally EEG in the case of BCI systems) that exhibit a co-modulation between spectral power and a given univariate signal.

2. COMMON SPATIAL PATTERNS

Common Spatial Patterns (CSP) (Ramoser et al., 2000) is a discriminative algorithm that determines the spatial filters \mathbf{W} from band-pass filtered EEG data such that the difference between the variances of the filtered data for the two classes is maximized.

This is done by a simultaneous diagonalization of the estimated covariance matrices $\mathbf{\Sigma}_1 = \mathbf{X}_1\mathbf{X}_1^T$ and $\mathbf{\Sigma}_2 = \mathbf{X}_2\mathbf{X}_2^T$ of the data for the two classes:

$$\mathbf{W}^\top \boldsymbol{\Sigma}_1 \mathbf{W} = \boldsymbol{\Lambda}_1 \quad (1)$$

$$\mathbf{W}^\top \boldsymbol{\Sigma}_2 \mathbf{W} = \boldsymbol{\Lambda}_2, \quad (2)$$

$$\text{s.t. } \boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2 = \mathbf{I} \quad (3)$$

where $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ are diagonal matrices and each λ on the diagonal corresponds to an eigenvector \mathbf{w}^\top . In this way, the eigenvectors are the same for both decompositions and the same eigenvector, i.e. a spatial filter, corresponds to a large eigenvalue for one class and to a small eigenvalue for the other class. Since eigenvectors with large eigenvalues correspond to a large variance of the data, spatial filters with extreme eigenvalues maximize the difference in the variances for the two classes.

The sum of the formulas in Eq. 3 forms the generalized eigenvalue problem:

$$\boldsymbol{\Sigma}_2 \mathbf{W} = (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2) \mathbf{W} \boldsymbol{\Lambda} \quad (4)$$

Choosing D filters corresponding to extreme eigenvalues (either close to 1 or close to 0) the filtered data $\hat{\mathbf{s}}(t) = \mathbf{W}_D^\top \mathbf{X}$ will have smaller dimensionality $D < N$ and the two classes will be maximally separated by their variance. A CSP feature is the log-variance of the band-pass and CSP filtered data.

When the recorded EEG signal contains artifacts or is nonstationary, CSP may fail causing a significant drop of BCI performance. The left panel of Figure 1 shows the effect of artifacts on the CSP algorithm. Here CSP fails to compute spatial filters which capture MI related activity. The algorithm rather focuses on an artifactual trial resulting in a degenerated CSP pattern and poor classification performance. The right panel of Figure 1 depicts the nonstationarity problem in BCI. Here at the beginning (left) the classes are perfectly separated by a hyperplane (dashed line), whereas the same hyperplane does not separate the classes at later times (right) anymore. Such change in the feature distribution may lead to a significant drop in classification accuracy, Vidaurre et al. (2007).

3. COMMON SPATIAL PATTERNS PATCHES

CSPP is the application of CSP analysis to small sets of channels (patches), and the combination of the resulting features Sannelli et al. (2011). The hypothesis is that applying CSP analysis on just a few channels, the risk of over-fitting is reduced in comparison to usual CSP, which

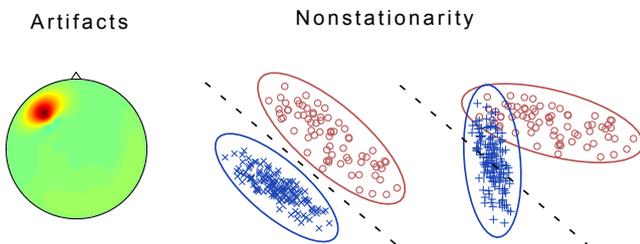


Fig. 1. Left: An artifact leads to a degenerated CSP pattern. The location of the artifact shows much larger power than the rest of the scalp. Right: A nonstationary feature distribution leads to poor classification performance because the hyperplane which separated the classes at the beginning (left) does not separate the classes at later times (right).

is calculated on all channels available. This is because the number of parameters to fit for each patch is less than for CSP. CSPP can be interpreted as a Laplacian filter where the weights of all channels are data driven optimized. We will call *center* of the patch, the channel that is weighted by 1 in Laplacian filtering.

Each patch can include a different number of surrounding channels and the position of the centers of the patches can be chosen, depending on the number of channels available, on the task and, when training data are already acquired, also depending on the subject-specific SMR activity. Here, eight patch forms in combination with a number of centers going from three to 18 are evaluated. The analyzed patch forms using the channel C3 as center are shown on the right of Fig. 2 and the centers of the configurations are shown on the left of Fig. 2. Similarly to CSP, CSPP are applied on band-pass and time filtered data. By computing CSP on one patch with n channels, n filters are obtained. With N_c patch centers, for each patch p with $p = 1, \dots, N_c$, one filter per class is selected by the extreme eigenvalues, i.e. two filters per patch are obtained. By concatenating all filters, a sparse filter matrix \mathbf{W} results with dimension $N_c \times N$ where N is the number of all available electrodes:

$$\mathbf{W} = [\mathbf{w}_{11}, \mathbf{w}_{12}, \dots, \mathbf{w}_{p1}, \mathbf{w}_{p2}, \dots, \mathbf{w}_{N_c1}, \mathbf{w}_{N_c2}] \quad (5)$$

The matrix \mathbf{W} is sparse since each column \mathbf{w}_{pi} contains just n non zero elements. From the resulting ensemble of filters \mathbf{W} , the most informative ones are chosen using the *ratio-of-medians* score.

Fig. 3 is an example of how just using 13 channels with a *small* patch in C3, Cz and C4, the filters can change in comparison to Laplacian filters and can become better class related. The top row presents the classical Laplacian derivations with 4 neighboring channels in C3, Cz and C4, while the other three rows depict the CSPP involving the same channels respectively for the class combinations *Left/Right*, *Left/Foot* and *Foot/Right*. The CSPP are calculated on the calibration data of three good performing users.

3.1 Experimental Results

Data set The evaluation was performed with data of eleven volunteers. Six participants had very good performance, and five had not. All participants performed eight feedback runs, each of them consisting of 100 trials (50 trials of each class). Two different types of MI, chosen out of three possibilities (MI of left hand, right hand or foot) were selected in advance. For seven participants, previous data with MI performance was available. It was therefore known which two MI tasks should be used. For the other four volunteers no prior information could be used and they were asked to select two out of the three possible MI tasks. In the first run, the features were six CSPP obtained by CSP analysis on three *small* patches centered on C3, Cz and C4 using the band-pass filtered data. A broad band 8-32 Hz and time interval 750-3750 ms were used. CSPP features are the log-variance of the six CSPP filtered signals. For runs 2 and 3, CSPP analysis on 18 *small* patches was performed, resulting in 36 CSPP features (2 per patch). To reduce the dimension, a maximum of six (and minimum of one per class) CSPP features were automatically selected by *ratio-of-medians* score. During the runs, the CSPP filters were re-selected

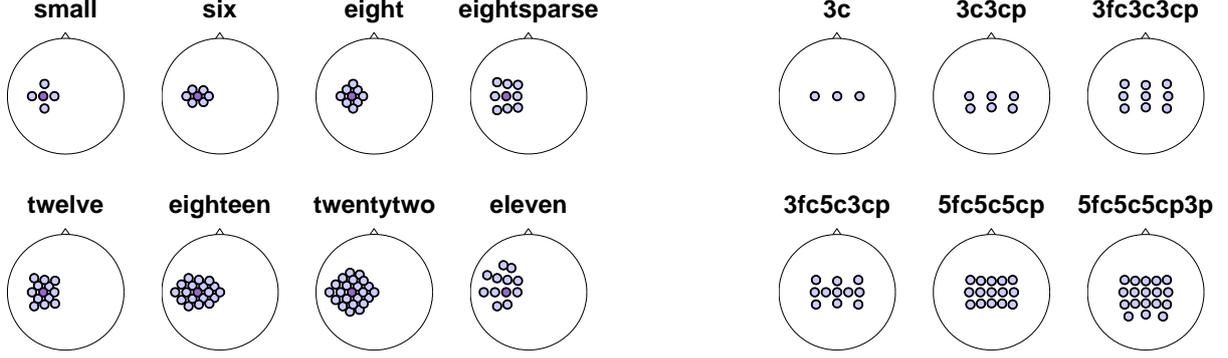


Fig. 2. Right: patch configurations centered on C3. Left: channel sets used as center of the patches/Laplacian filters.

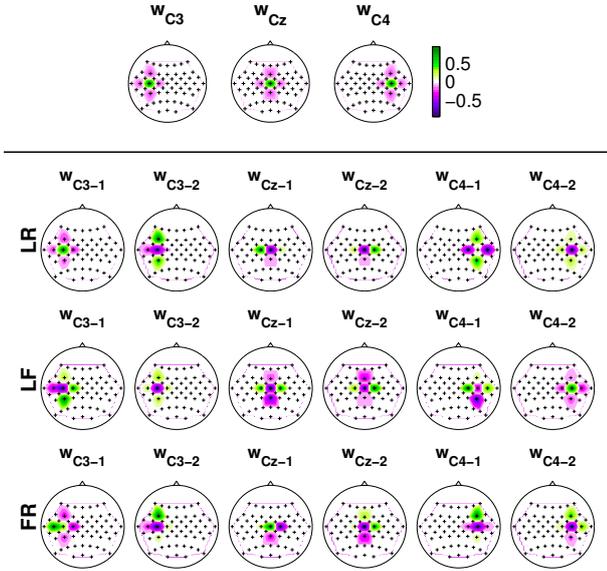


Fig. 3. Top: *Small* Laplacian filters for C3, Cz and C4. Bottom: *Small* CSPP filters centered in C3, Cz and C4, for the class combination *Left/Right* (first row), *Left/Foot* (second row) and *Foot/Right* (third row), calculated on the data of three good performing users.

using all previous trials. Note, that only 36 EEG channels are required for this design. For run 4, CSP and CSPP features were concatenated. For each patch form in Fig. 2, a generalization error was calculated by 4-fold CV where for each fold: 1) two to six CSPP and CSP features were calculated on the training set, concatenated, and used to train the Linear Discriminant Analysis (LDA) and 2) the test set was spatially filtered by the selected CSPP+CSP filters and the resulting features classified by the trained LDA. The patch form with the best generalization error was selected. Finally, CSPP (with the chosen patch form), CSP and LDA were trained on the whole data set. During the runs, the calculated CSP, CSPP and LDA are used to calculate the features and for online classification, but the pooled mean of the features was adapted after each trial and used to update the bias of the classifier, Vidaurre et al. (2011).

BCI performance In Fig. 4 the accuracy obtained by offline-online simulation with the CSPP design is compared to the online performance. The average across users

of the same category (Cat. I are good performers, Cat. II are bad online performers with good calibration accuracy and Cat. III are users with bad calibration accuracy) is presented. Each point is the mean over 20 trials, while each bar is the mean over one run. Two different shades for LAP respectively CSPP design, light for LAP and dark for CSP. Purple is run 1, turquoise are runs 2 and 3 and green is run 4. The subject-independent CSPP improved the accuracy for all BCI users in run 1. While in the original study users with bad performance could only reach a 70% of accuracy in run 4 or 5, the CSPP design allows this to happen within the first three runs for all users except for one.

4. STATIONARY AND DIVERGENCE-BASED COMMON SPATIAL FRAMEWORK

4.1 Stationary CSP

The CSP algorithm maximizes the variance ratio of two MI classes, but does not optimize for stationarity of the feature distribution. The main idea of stationary CSP (sCSP) Samek et al. (2012) is to extract filters that maximize the variance ratio, but at the same time keep the variance estimation along the projected direction as stable as possible across trials (i.e., keep the feature distribution stationary). Formally, the following quantity is minimized for each class c

$$D_c(\mathbf{w}) = \mathbf{w}^\top \underbrace{\left(\frac{1}{n} \sum_{i=1}^n \mathcal{F}(\Sigma_c^i - \Sigma_c) \right)}_{\Delta_c} \mathbf{w} \approx \frac{1}{n} \sum_{i=1}^n |\mathbf{w}^\top \Sigma_c^i \mathbf{w} - \mathbf{w}^\top \Sigma_c \mathbf{w}| \quad (6)$$

where Σ_c^i is the covariance matrix of the i -th trial of class c and Σ_c is the average covariance matrix of class c and \mathcal{F} is an operator to make symmetric matrices be positive definite by flipping negative eigenvalues. Intuitively sCSP aims to extract features that maximize the “between-class distance” (i.e., the variance ratio) and minimize the “within-class variance” (i.e., measured as absolute deviation). Since sCSP applies the operator \mathcal{F} in Eq. (6), the spatial filters can be computed very efficiently by maximizing the Rayleigh quotient

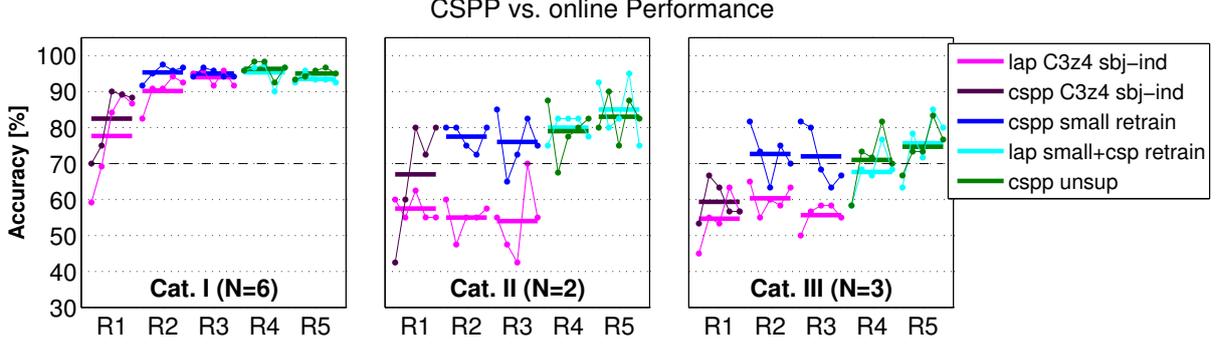


Fig. 4. Performance of CSPP during the first 5 runs of the co-adaptive calibration study compared with the original performance. Each point is the mean of the accuracy across 20 trials (thin lines) and across the users of the corresponding category. Each bold line is the mean across 100 trials of the corresponding run. LAP original experiment: runs 1-3 with subject-independent classifier and supervised adaptation, runs 4-5 with CSP+*small* LAP re-selection (supervised adaptation). CSPP design: run 1 with subject-independent classifier and supervised adaptation, runs 2-3 with *small* CSPPs re-selection (supervised adaptation), runs 4-5 with CSPP+CSP (automatic patch form) and unsupervised adaptation.

$$R_c(\mathbf{w}) = \frac{\mathbf{w}^\top \Sigma_c \mathbf{w}}{\mathbf{w}^\top (\Sigma_1 + \Sigma_2 + \lambda(\Delta_1 + \Delta_2)) \mathbf{w}} \quad (7)$$

where λ is a regularization parameter.

4.2 Divergence-based Spatial Filtering Framework

Recently, the authors of Samek et al. (2013) showed that spatial filter computation can be cast into a divergence framework (divCSP). Mathematically, spatial filters are computed by maximizing a symmetric divergence between the probability distribution of both classes:

$$\mathbf{V}^* = \underset{\mathbf{V}}{\operatorname{argmax}} \tilde{D}(\mathcal{N}(\mathbf{0}, \mathbf{V}^\top \Sigma_1 \mathbf{V}) \parallel \mathcal{N}(\mathbf{0}, \mathbf{V}^\top \Sigma_2 \mathbf{V}))$$

where $\mathcal{N}(\mathbf{0}, \Sigma)$ denotes the Gaussian distribution, with mean $\mathbf{0}$ and covariance Σ . Note that for the case of Kullback-Leibler divergence, the method extracts CSP filters.

In the divergence framework, robustness to artifacts can be achieved by decomposing the divergence between the average class distributions into the sum of trialwise divergences and limiting the influence of single (potentially outlier) terms (see Samek et al. (2013)). This changes the objective function into

$$\mathbf{V}^* = \underset{\mathbf{V}}{\operatorname{argmax}} \sum_i \tilde{D}(\mathcal{N}(\mathbf{0}, \mathbf{V}^\top \Sigma_1^i \mathbf{V}) \parallel \mathcal{N}(\mathbf{0}, \mathbf{V}^\top \Sigma_2^i \mathbf{V}))$$

Beta divergence¹, Bhattacharyya distance and Gamma divergence have been proposed as robust measures in this framework (see Samek et al. (2013); Brandl et al. (2015)).

By adding a regularization term to the objective one can easily use data from additional subjects or enforce specific properties such as stationarity on the solution. The objective function then has the form

$$\mathcal{L}(\mathbf{V}) = \underbrace{(1 - \nu) \tilde{D}(\mathbf{V}^\top \Sigma_1 \mathbf{V} \parallel \mathbf{V}^\top \Sigma_2 \mathbf{V})}_{\text{CSP Term}} - \underbrace{\nu \Delta}_{\text{Regularization Term}} \quad (8)$$

where Δ is the regularization term that can be arbitrarily defined, depending on the type of invariance we want to achieve, and ν is a regularization parameter trading-off the influence of the CSP objective function and the regularization term. Similarly as for the sCSP algorithm (but without using an approximation²) one can enforce stationarity on the feature distribution by using the following regularization term

$$\Delta = \frac{1}{2n} \sum_{c=1}^2 \sum_{i=1}^n D(\mathbf{V}^\top \Sigma_c^i \mathbf{V} \parallel \mathbf{V}^\top \Sigma_c \mathbf{V}), \quad (9)$$

4.3 Experimental Results

Data Set The Vital BCI data set Blankertz et al. (2010) contains EEG recordings from 80 healthy subjects performing MI tasks with the left and right hand or with the feet. It consists of one calibration and one feedback session, both recorded on the same day. In the calibration session visual cues (arrows pointing left, right, down) indicated which MI task should be performed and three runs with 25 trials of each motor condition were recorded. Then, the best binary combination of MI tasks were selected and the subjects performed feedback with three runs of 100 trials each (some users performed only one or two runs). Visual feedback, i.e., a cursor moving on the screen, was provided to the user while performing MI. Note that this feedback was lacking in the calibration phase. The signals were recorded from 118 Ag/AgCl electrodes, band-pass filtered between 0.05 and 200 Hz and downsampled to 100 Hz. All subjects in this study were BCI novices.

The following preprocessing is applied in the experiments performed on this data set. We manually select 62 elec-

² The sCSP approximation fails if the trialwise covariance matrices are not jointly diagonalizable.

¹ Beta divergence between distributions p and q is defined as $D_\beta(p \parallel q) = \frac{1}{\beta} \int (p^\beta - q^\beta) p dx - \frac{1}{\beta+1} \int (p^{\beta+1} - q^{\beta+1}) dx$

trodes densely covering the motor cortex and filter the data in the frequency range 8-30 Hz with a 5th order Butterworth filter. Furthermore, we use a fixed time segment from 750 to 3500 ms after the trial start for feature extraction.

BCI Performance In the following we evaluate the stationary CSP method (sCSP), the robust divCSP algorithm based on symmetric beta divergence (β -divCSP) and the divCSP algorithm with regularization towards stationarity (*reg-divCSP*) on the Vital BCI data set. The following parameters are used in the experiment: $\lambda, \nu = \{0, 2^{-10}, 2^{-9}, \dots, 2^0\}$ and $\beta = \{0, 0.0001, 0.001, 0.01, 0.05, 0.1, 0.15, 0.2, 0.25, 0.5, 0.75, 1, 1.5, 2, 5\}$. The parameters are selected by 5-fold cross-validation on the calibration data. The one-sided Wilcoxon sign-rank test is applied to test significance.

Figure 5 visualizes the results. One can see that all three methods significantly improve the BCI performance over the CSP baseline by reducing nonstationarity and the influence of artifacts.

5. SOURCE POWER CO-MODULATION

The core idea of the SPoC approach Dähne et al. (2014) is to (i) decompose the multivariate EEG data into a set of source components and (ii) to use the information contained in a target variable (z , a scalar function of time) to guide the decomposition. The result of this approach is a set of spatial filters, \mathbf{W} , which *directly* optimize the co-modulation between the target variable z and the power time course of the spatially filtered signal. In a neuroimaging context, this target variable will typically either represent a behavioral measure as the final output of the central nervous activity (e.g. reaction time, sensory detection, task rating, motor evoked potentials, etc.) or a parameter of external stimuli (e.g. when studying how amplitude modulation of neuronal oscillations correlate with stimulus properties). Also, SPoC has an advantage over blind source separation methods such as for example Independent Component Analysis (ICA) Comon (1994), because it has more information at its disposal (the target variable). The SPoC algorithm is intimately related to the Common Spatial Pattern (CSP) algorithm family. When the target variable z is binary, classical CSP is obtained as a special case of SPoC. One may thus view SPoC as a regression extension of CSP to continuous target variables.

The derivation of SPoC is as follows: we assume that the EEG data $\mathbf{x}(t)$ has been band-pass filtered in the frequency band of interest. Thus, the power of the projected signal $\mathbf{w}^\top \mathbf{x}(t)$ within a small time interval is well approximated by the variance of $\mathbf{w}^\top \mathbf{x}(t)$ within that interval. We refer to such time intervals as *epochs* and assume that the EEG data can be divided up into consecutive or overlapping epochs of suitable length³. Epochs will be indexed by the index e . We assume the target variable z to only

³ Working with epoched data instead of continuous data does not represent a loss of generality, because all of the following derivations can be reformulated for continuous data as well, provided that the target variable changes slowly enough. We choose to work with epoched data because it resembles the format of data obtained in trial-based experiments.

have a single value per epoch, which can be achieved by appropriate re-sampling. Furthermore we assume without loss of generality that z has zero mean and unit variance, which can be achieved by normalization.

It is our goal to approximate the target variable z with the bandpower/variance of a source component. We denote this estimate by ϕ , which depends on a spatial filter \mathbf{w} . Let $\text{Var}[\mathbf{w}^\top \mathbf{x}(t)](e)$ denote the variance of $\mathbf{w}^\top \mathbf{x}(t)$ in a given epoch e . This epoch-wise variance of the projected signal will serve as the approximation of z . Thus we have

$$z(e) \approx \phi(e) = \text{Var}[\mathbf{w}^\top \mathbf{x}(t)](e) = \mathbf{w}^\top \mathbf{C}(e) \mathbf{w}, \quad (10)$$

where $\mathbf{C}(e)$ denotes the covariance matrix of the e -th epoch. Let us further define the matrix

$$\mathbf{C}_z := \langle \mathbf{C}(e) z(e) \rangle, \quad (11)$$

where the $\langle \cdot \rangle$ denotes the average over epochs. Then the objective function that is optimized by SPoC is given by

$$\text{Cov}[\phi(e), z(e)] = \mathbf{w}^\top \mathbf{C}_z \mathbf{w}, \quad (12)$$

with respect to the following norm constraint:

$$\text{Var}[\mathbf{w}^\top \mathbf{x}(t)] = \mathbf{w}^\top \mathbf{C} \mathbf{w} \stackrel{!}{=} 1. \quad (13)$$

This constraint optimization problem can be solved using the method of Lagrange multipliers. Setting the first derivative of the corresponding Lagrangian to zero leads to the following generalized eigenvalue equation:

$$\mathbf{C}_z \mathbf{w} = \lambda \mathbf{C} \mathbf{w}, \quad (14)$$

where the eigenvalue λ corresponds to the covariance between ϕ and z .

5.1 Applications

SPoC has been successfully applied in the context of a visuomotor workload paradigm. In such a setting where binary labels may not be available, a suitable target variable can be extracted from task-related error rates or other physiological indicators of stress level such as skin conductance or heart rate. In the study by Schultze-Kraft et al. (2013), an extension of SPoC to multivariate target variables was employed to find a linear combination of the three mentioned workload indicators that best correlated with the band-power modulations of brain rhythms. The scalp maps of the SPoC components clearly reflected the known impact of workload on the human EEG. On the one hand, modulations of the theta frequency band-power, as extracted by SPoC, showed a decrease during low workload condition and an increase during high workload condition. As for the alpha frequency band the power modulations of the SPoC components showed the opposite effect.

In very recent study, Meinel et al. (2015) used SPoC to identify individual oscillatory components that predict reaction times in a sequential visual isometric pinch task (SVIPT). Such a task is applied to patients in the sub-acute and chronic phase of stroke in order to reveal hand motor deficits. The utilization of the SPoC approach allowed for the extraction of a component showing band power co-modulation with the reaction time of the hand

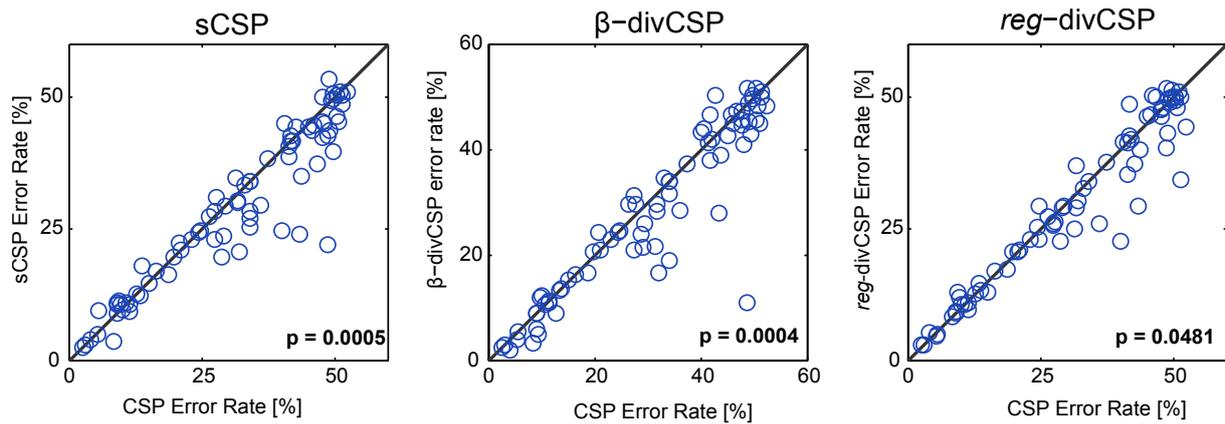


Fig. 5. Performance comparison between CSP and sCSP, β -divCSP and reg-divCSP.

motor task. Despite of the relative small number of epochs and the low SNR of EEG, components are surprisingly stable over subjects and test folds. This finding is encouraging the expansion of the workflow to a larger frequency band range and other performance metrics of the SVIPT.

6. CONCLUSIONS

This paper presents a summary of current machine learning methods for spatial filtering developed at the Machine Learning Group of the Technical University of Berlin. All of them represent an improvement to the state-of-the-art regarding different problems that occur in electroencephalographic signals.

ACKNOWLEDGEMENTS

The authors acknowledge funding by the German Research Foundation (DFG) grant nos. MU 987/19-1, MU 987/14-1, MU 987/3-2, and support by the Bernstein Center for Computational Neuroscience Berlin through the graduate program GRK 1589/1

REFERENCES

- Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.R., Kunzmann, V., and Curio, F.L. (2006). The berlin brain-computer interface: Eeg-based communication without subject training. *IEEE Transactions Neural System and Rehabilitation Engineering*.
- Blankertz, B., Sannelli, C., Halder, S., Hammer, E.M., Kübler, A., Müller, K.R., Curio, G., and Dickhaus, T. (2010). Neurophysiological predictor of smr-based bci performance. *NeuroImage*, 51(4), 1303–1309. doi:DOI: 10.1016/j.neuroimage.2010.03.022.
- Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Müller, K. (2008). Optimizing spatial filters for robust EEG Single-Trial analysis. *Signal Processing Magazine, IEEE*, 25(1), 41–56.
- Brandl, S., Müller, K.R., and Samek, W. (2015). Robust common spatial patterns based on bhattacharyya distance and gamma divergence. In *Proceedings of the 2015 International Winter Workshop on Brain-Computer Interface (BCI)*. To appear.
- Comon, P. (1994). Independent component analysis: a new concept? *Signal Processing*, 36, 289–314.
- Dähne, S., Meinecke, F.C., Haufe, S., Höhne, J., Tangermann, M., Müller, K.R., and Nikulin, V.V. (2014). SPoC: a novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters. *NeuroImage*, 86(0), 111–122. doi: http://dx.doi.org/10.1016/j.neuroimage.2013.07.079.
- Guger, C., Edlinger, G., Harkam, W., Niedermayer, I., and Pfurtscheller, G. (2003). How many people are able to operate an EEG-based brain-computer interface (BCI)? *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 11(2), 145–147.
- Meinel, A., Castano-Candamil, J.S., Reis, J., Dähne, S., and Tangermann, M. (2015). EEG band power predicts single-trial reaction time in a hand motor task. In *Proceedings of the 7th International IEEE EMBS Conference on Neural Engineering*.
- Ramoser, H., Muller-Gerking, J., and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *Rehabilitation Engineering, IEEE Transactions on*, 8(4), 441–446.
- Samek, W., Blythe, D., Müller, K.R., and Kawanabe, M. (2013). Robust spatial filtering with beta divergence. In *Advances in Neural Information Processing Systems*, 1007–1015.
- Samek, W., Vidaurre, C., Müller, K.R., and Kawanabe, M. (2012). Stationary common spatial patterns for brain-computer interfacing. *Journal of Neural Engineering*, 9(2), 026013.
- Sannelli, C., Vidaurre, C., Müller, K.R., and Blankertz, B. (2011). CSP patches: an ensemble of optimized spatial filters. an evaluation study. *Journal of Neural Engineering*, 8(2), 025012.
- Schultze-Kraft, M., Dähne, S., Blankertz, B., and Curio, G. (2013). Temporal and spatial distribution of workload-induced power modulations of EEG rhythms. In *International BCI Meeting, Asilomar, 2013*. doi: 10.3217/978-3-85125-260-6-116.
- Vidaurre, C., Kawanabe, M., von Büna, P., Blankertz, B., and Müller, K.R. (2011). Toward unsupervised adaptation of LDA for brain-computer interfaces. *IEEE Transactions on Bio-Medical Engineering*, 58(3), 587–597.
- Vidaurre, C., Scherer, R., Cabeza, R., Schlögl, A., and Pfurtscheller, G. (2007). Study of discriminant analysis applied to motor imagery bipolar data. *Medical & biological engineering & computing*, 45(1), 61–68.