A Minimax Approach to Optimize Spatial Filters for EEG Single-Trial Classification

Motoaki Kawanabe[1], Carmen Vidaurre[2] Benjamin Blankertz[2,1], and Klaus-Robert Müller[2]

[1] IDA group, FIRST, Fraunhofer Institute, Kekulestr. 7, 12489 Berlin
motoaki DOT kawanabe AT first DOT fraunhofer DOT de,
[2] Berlin Institute of Technology, Computer Science Faculty, Machine Learning department, Franklinstr. 28/29, 10587 Berlin

Abstract. EEG single-trial analysis requires methods that are robust with respect to noise, artifacts and nonstationarity among other problems. This work contributes by developing a minimax approach to robustify the common spatial patterns (CSP) algorithm. By optimizing the worst-case objective function within a prefixed set of the covariance matrices , we can transform the respective complex mathematical program into a simple generalized eigenvalue problem and thus obtain robust spatial filters very efficiently. We test our minimax CSP method with real world brain-computer interface (BCI) data sets in which we expect substantial fluctuations caused by day-to-day or paradigm-to-paradigm variability or different forms of stimuli. The results clearly show that the proposed method significantly improves the classical CSP approach in multiple BCI scenarios.

1 Introduction

Feature extraction is an important prerequisite for analyzing high dimensional real world data. For single-trial EEG classification tasks, spatial filters have become very popular feature extractors. Data driven approaches that optimize spatial filters for each subject individually have been proven useful [1], in particular in Brain-Computer Interfaces, which translate the users intent (coded by a small set of mental tasks) into control actions such as computer applications or neuroprostheses [2–4].

In the past years machine learning methods have led to significant advances in the analysis and modeling of neural signals. While early EEG-BCI efforts required neurofeedback training on the part of the user that lasted on the order of days, in ML-based systems it suffices to collect examples of EEG signals in a so-called *calibration measurement* during which the user is cued to perform repeatedly 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*). This step of adaption seems to be instrumental for effective BCI performance despite the large inter-subject variability of the respective brain signals [5]. After this preparation step, which is very short compared to the complementary subject training in the operant conditioning approach [6], the feedback application can start. Here, the users can actually transfer information through their brain activity and control applications. In this phase, the system is composed of the classifier that discriminates between different mental states and the control logic that translates the classifier output into control signals, e.g., a cursor position or a selection from an alphabet.

There are several aspects in which BCI research can profit from improvement, see the 'Challenges' section of [7]. One of them is to make the system more *robust* against non task-related fluctuations and/or non-stationarity of the measured EEG signals. These fluctuations may be caused by changes in the subject's brain processes, e.g. change of task involvement, fatigue etc., or by artifacts such as swallowing, blinking or yawning.

Recently Kim et al. [8] applied a minimax approach to Fisher Discrimant Analysis (FDA) and proposed a novel robust classification method. From their minimax theorem, the minimax FDA is guaranteed to have larger Rayleigh coefficients for any fluctuations within a prefixed tolerance set.

The present paper contributes by investigating a minimax approach in the spirit of [8] to common spatial patterns (CSP) [9], which is one of the working horses for spatial filtering in BCI applications. In contrast to the FDA case, we can obtain the worst case covariance matrices analytically, which leads a modified generalized eigenvalue problem.

2 Sensorimotor Rhythms and Common Spatial Patterns

Apart from transient components, EEG comprises rhythmic activity located over various areas. Most of these rhythms are so-called idle rhythms, which are generated by large populations of neurons in the respective cortex that fire in rhythmical synchrony when they are not engaged in a specific task. Over motor and sensorimotor areas in most subjects oscillations with a fundamental frequency between 9 and 13 Hz can be observed, the so called μ-rhythm. Due to its comb-shape, the μ-rhythm is composed of several harmonics, i.e., components of double and sometimes also triple the fundamental frequency with a fixed phase synchronization. These sensorimotor rhythms (SMRs) are attenuated when engagement with the respective limb takes place. As this effect is due to loss of synchrony in the neural populations, it is termed event-related desynchronization (ERD), see [10]. The increase of oscillatory EEG (i.e., the reestablishment of neuronal synchrony) is called event-related synchronization (ERS). The ERD in the motor and/or sensory cortex can be observed even when a subject is only thinking of a movement or imagining a sensation in the specific limb. This phenomenon makes the ERD/ERS feature attractive for BCIs.

For 'decoding' of different motor intentions from brain activity, the essential task is to distinguish different spatial localization of SMR modulations. Due to the topographical arrangement in the motor and somatosensori cortex, these locations are related to corresponding parts of the body. For example, left hand and right hand have corresponding areas in the contralateral, i.e., right and left motor cortex, respectively. Thus, spatial filters are an essential step for a meaningful feature extraction for the classification of motor intentions, and far from being a black-box methods, learned spatial filters can be visualized appropriately and checked with neurophysiologcal knowledge.

The common spatial pattern (CSP) algorithm is successful in calculating spatial filters for detecting modulations of the SMR or other ERD/ERS effects. Given two distributions in a high-dimensional space, the (supervised) CSP algorithm finds directions (i.e., spatial filters) that maximize variance for one class and simultaneously minimize variance for the other class. Since band-power can be calculated as the variance of band-pass filtered signals, this criterion corresponds to ERD/ERS effects.

Technically CSP analysis works as follows. Let Σ_+ and Σ_- be estimates of the covariance matrices of the band-pass filtered EEG signals under the two conditions. These two matrices are simultaneously diagonalized such that the eigenvalues of Σ_+ and Σ_- sum to 1. Practically this can be done by calculating the generalized eigenvectors W:

$$\Sigma_+ W = (\Sigma_+ + \Sigma_-) W D. \tag{1}$$

Here, the diagonal matrix D contains the (generalized) eigenvalues of Σ_+ (defined such that they are between 0 and 1) and the column vectors of W are the filters for the CSP projections. By this procedure a full decomposition of the sensor space is determined. Best contrast is provided by those filters with high eigenvalues (large variance for condition 1 and small variance for condition 2) and by filters with low eigenvalues (vice versa). Therefore, the common practice in a classification setting is to use several eigenvectors from both ends of the eigenvalue spectrum as features for classification. The solution for the eigenvector with the largest eigenvalue can also be obtained by maximizing the Rayleigh quotient:

$$\underset{w \in \mathbb{R}^C}{\text{maximize}} \qquad \frac{w^\top \Sigma_+ w}{w^\top (\Sigma_+ + \Sigma_-) w}. \tag{2}$$

This correspondence is often useful for algorithmic considerations. CSP filters can be visualized as scalp maps and chosen according to physiological plausibility. For more details, see the CSP tutorial [1].

3 Robust spatial filters based on minimax framework

The class covariance matrices Σ_+ and Σ_- used in CSP can vary substantially because of non task-related fluctuations and/or non-stationarity of the EEG signals. In BCI applications, it is important to make the features robust against various kinds of fluctuations, e.g. caused by change of task involvement or by changes in the subject's brain processes. Under such situations, the minimax approach which Kim et al. [8] successfully applied to Fisher discriminant analysis (FDA) could be one of the promising directions to construct robust CSP filters. The key idea is that, instead of just two single matrices, we consider convex sets $\mathcal{S}_+$ and $\mathcal{S}_-$ for the class covariance matrices Σ_+ and Σ_-, respectively. These sets specify the tolerant regions of fluctuations around the class covariance matrices. For simplicity, we assume that the sets $\mathcal{S}_+$ and $\mathcal{S}_-$ are independent of each other.

Based on the minimax framework, robust CSP can be constructed by maximizing the worst case Rayleigh quotient within all possible covariance matrices in the tolerant regions. That is, the optimization problems for our maxmin CSP can be expressed as

$$\begin{aligned} &\text{maximize} \min_{\Sigma_+ \in \mathcal{S}_+,\, \Sigma_- \in \mathcal{S}_-} \frac{w^\top \Sigma_+ w}{w^\top \left(\Sigma_+ + \Sigma_-\right) w} \\ &\text{subject to } w \neq 0 \end{aligned} \tag{3}$$

$$\begin{aligned} &\text{maximize} \min_{\Sigma_+ \in \mathcal{S}_+,\, \Sigma_- \in \mathcal{S}_-} \frac{w^\top \Sigma_- w}{w^\top \left(\Sigma_+ + \Sigma_-\right) w} \\ &\text{subject to } w \neq 0 \end{aligned} \tag{4}$$

For a moment, we will consider only the first optimization problem (3), because the other (4) can be handled in the same way.

3.1 Derivation of the maxmin filters

We will consider special cases where the subspaces $\mathcal{S}_+$ and $\mathcal{S}_-$ can be defined by balls in the space of $C \times C$ positive definite matrices $\mathrm{Pd}(C)$ centered at $\overline{\Sigma}_+$ and $\overline{\Sigma}_+$

$$\begin{aligned} \mathcal{S}_+ &= \left\{ \Sigma_+ \,\middle|\, \Sigma_+ \succeq 0,\ \|\Sigma_+ - \overline{\Sigma}_+\|_{P_+} \leq \delta_+ \right\}, \\ \mathcal{S}_- &= \left\{ \Sigma_- \,\middle|\, \Sigma_- \succeq 0,\ \|\Sigma_- - \overline{\Sigma}_-\|_{P_-} \leq \delta_- \right\}, \end{aligned} \tag{5}$$

where $\|X\|_P^2 = \mathrm{Tr}\left(P^{-1} X P^{-1} X\right)$ is the norm of a symmetric matrix X and the positive definite matrix P specifies the metric of $\mathrm{Pd}(C)$, or in other word, shape of the balls.

Lemma 1. *For the sets $\mathcal{S}_+$ and $\mathcal{S}_-$ defined in Eq.* (5)*, the worst case Rayleigh quotient becomes*

$$\frac{w^\top \left(\overline{\Sigma}_+ - \delta_+ P_+\right) w}{w^\top \left(\overline{\Sigma}_+ + \overline{\Sigma}_- - \delta_+ P_+ + \delta_- P_-\right) w}, \qquad \forall w, \tag{6}$$

if $\overline{\Sigma}_+ - \delta_+ P_+ \succeq 0$.

3.2 Choice of the parameters of the tolerance sets

The remaining problem is how to determine the sets $\mathcal{S}_+$ and $\mathcal{S}_-$ of the class covariance matrices. We used the average class covariance matrices for the centers $\overline{\Sigma}_+$ and $\overline{\Sigma}_-$. There are several choices of the matrix P_+ and P_- which specify the shape of the tolerant balls. The standard norms of such forms are 'Frobenius' ($P_+ = P_- = I$) or $\overline{\Sigma}_+$ and $\overline{\Sigma}_-$ (the centers). However, in the latter case the filters coincide with those of CSP. Although the identity matrix ignores plausible directions of fluctuation in EEG signals, the maxmin CSP with this setting still improved the performance in the day-to-day transfer experiment. We conjecture that this is analogous to the fact that Bayesian regularization helps even with non-informative priors. If we have extra (prior) information about possible fluctuations as is the case with the real world BCI data in [11], the covariance of the distortions can be used for the matrices P_+ and P_-. This approach was called invariant CSP (iCSP). Future work should be done to analyse the scatter of the short-term covariance matrices in detail and to find a reasonable choice of these matrices. The size of the balls δ_+ and δ_- are assumed to be equal and are selected by cross-validation on the training set.

Fig. 1 is an illustrative explanation of our method. Although we developed the theory only for the first eigenvectors, in the experiments we will use a few eigenvectors each. Further work should be done to extend the minimax theorem and the lemma for multiple eigenvectors.

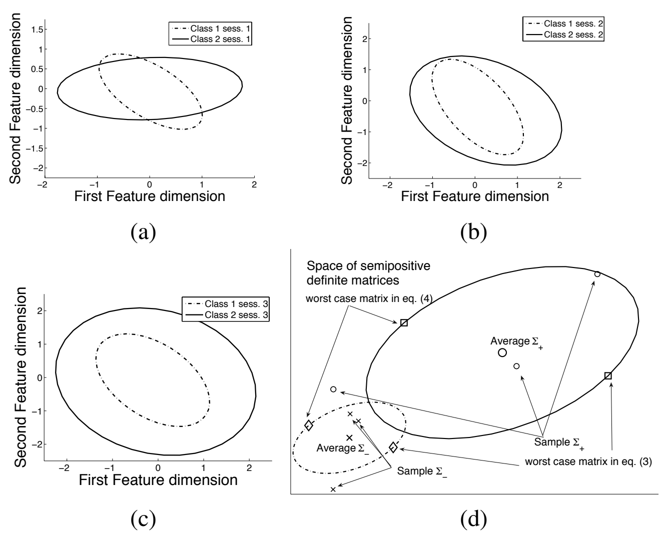

Fig. 1. Figs. (a), (b) and (c) represent Σ_+ and Σ_- at different time points. Mean of the features is 0, as it is bandpass filtered data. Fig. (d) represents the previous matrices as points in the space of positive definite matrices. The ellipsoids in Fig. (d) are the tolerant sets S_+ and S_- centered at the average matrices $\overline{\Sigma}_+$ and $\overline{\Sigma}_-$, respectively. From both ellipsoid, a pair of the worst case covariances is obtained for each optimization problem (3) or (4).

4 Application to EEG data from Brain-Computer Interfacing

In this paper we evaluate the proposed algorithm on off-line data in which substantial fluctuations are expected. First we show that maxmin CSP can work under the same settings as iCSP, that is robustifying the filters in situations in which a known and measurable distortion affects the data. Later we also test the algorithm to robustify the filters against session-to-session (day-to-day) variability which may be caused by different mental conditions, materials (cap and electrodes) and different preparation of the measurement devices. In the last example, we test whether classifiers trained with recordings from multiple paradigms can be transfered to data in yet another paradigm. The nonstationarity is induced by having different background conditions (presentation of the cues) for the same primary task (motor imagery). In this paper we use ‘calibration measurements' in which no feedback was provided to avoid bias toward any method. The trials of these 'calibration data' have fixed length of 3.5 seconds in which an imagery motor task is performed (right or left hand or foot movement). The cue is given as auditory ('imag_audi') or visual stimuli. The visual stimuli can be a letter L, R, F ('imag_lett'), a fixed arrow pointing to left, right or down ('imag_arrow') or a randomly moving small rhomboid ('imag_move') with either its left, right or bottom corner filled to indicate the task. The movement of the object in 'imag_move' was independent from the indicated targets with the aim to induce target-uncorrelated eye movements. For the evaluation, we only considered binary classification. From the three motor imaginary tasks, the best pair was selected based on separability for each subject. LDA was used as the classifier and the performance was measured by error rate.

Maxmin CSP working as iCSP. First we want to show that maxmin CSP can work under the same conditions of iCSP obtaining similar results. Therefore we realized two experiments. The first one is a replica of that presented in [11]. As in [11], we investigated whether it is possible to robustify CSP against different demands in visual processing, which cause substantial differences in the background brain activity [11]. We trained CSP using 'imag_move'. Additional data ('sham_feedback') with task uncorrelated eye movements and visual actitivity was recorded to create a disturbance matrix (for more information of the paradigm please refer to [11]). Fig. 2(a) depicts the results of 4 subjects in which we see that iCSP and maxmin CSP exhibit very similar performance.

The second experiment was done with 5 subjects to test maxmin CSPs invariance properties in another setting. We used 'imag_move' data for training CSP and some data recorded while subjects were told to keep their eyes closed, was used to characterize the expected disturbance (matrices $P_+ = P_-$). Maxmin CSP was then tested in 'imag_lett', in which the visual activity is supposed to be less intense (more similar to an 'eyes-closed' condition). Results are depicted in Fig. 2(b). Again iCSP and maxmin CSP perform very similarily, however, better than the original CSP.

Session-to-session transfer. For this test we used data from four subjects for whom we had recorded several sessions in different days (even with more than one year apart). For subject *zq* there were 6 datasets available, 5 datasets for *cm* and *zp* and finally 4 for subject *zk*. All files except one were used for training the maxmin filters. Matrices P_+ and P_- were the identity matrix. The last file was used to test the performance of each subject. All datasets correspond to 'calibration measurements' of type 'imag_lett'. The performance of the subjects is described using error rate and the trials are classified using LDA. Fig. 2(c) shows this error rate when using CSP or maxmin CSP for preprocessing the data. Again the new maxmin CSP method outperforms CSP in all cases.

Paradigm-to-paradigm transfer. To test whether it is possible to use 'calibration measurements' recorded using different paradigms we performed an experiment in which 'imag_move' and 'imag_audi' data were used to train maxmin CSP, whereas the performance test was done in 'imag_arrow', in which the visual task is less demanding than 'imag_move'. The number of trials used from 'imag_audi' was found subject-specifically by cross validation on the training set. Note that no additional and specific recording is necessary to estimate maxmin CSP. In this setting, we had data from five subjects and their performance using CSP and maxmin CSP is depicted in Fig. 2(d). Again, matrices P_+ and P_- were the identity matrix. We see that the performance of all five subjects could be improved by maxmin CSP.

All parameters of the experiments were fixed on the training set using 10 fold cross-validation before testing in unseen data: δ for maxmin CSP, ξ for iCSP and the number of trials from 'imag_audi' in the paradigm-to-paradigm transfer experiment.

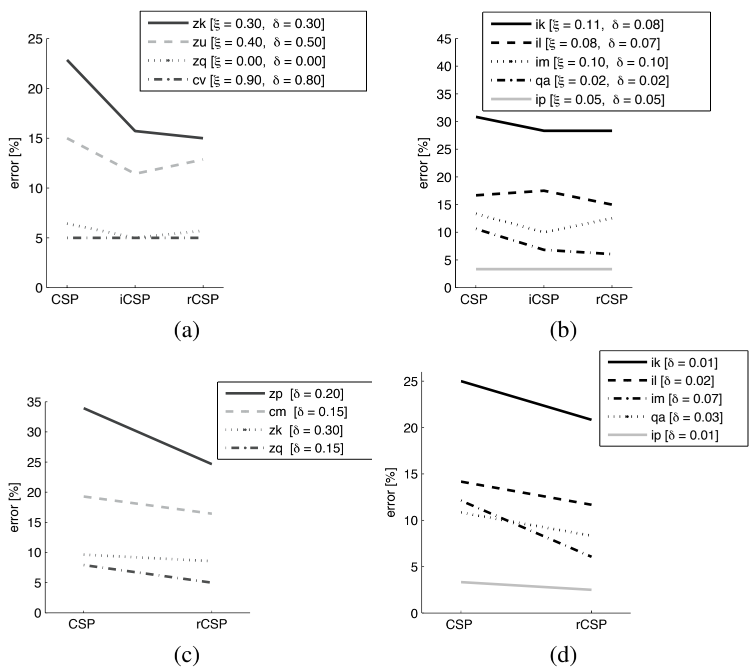

Fig. 2. Test errors of our maxmin CSP compared with CSP and iCSP. The legend provides subject codes (two letter codes) and the selected parameters in square brackets. (a) maxmin CSP working as iCSP (with sham_feedback), (b) maxmin CSP working as iCSP (with 'eyes closed' measurement), (c) session-to-session transfer, only possible with maxmin CSP, (d) paradigm transfer, only possible with maxmin CSP.

5 Conclusions

BCI data is contaminated by a variety of noise sources, artifacts, nonstationarities and outliers that make it indispensable to strive for more robust learning methods. In this paper we proposed a novel classification algorithm that is inspired by the work of Kim et al. [8].

In particular, we analyze the worst case performance among possible class covariance matrices and optimize the respective CSP-like filters based on such a criterion.

When we take hyperspheres in the matrix space for the sets of the covariances, the algorithm can be elegantly reduced to a generalized eigenvalue problem similar to the original CSP, but with modified covariance matrices. Extensive simulations show that our new CSP framework is indeed more robust as it allows to transfer BCI classifier knowledge from session to session. This is again a further step towards a BCI system that is more stable with respect to nonstationarities and non-task related fluctuations.

In future studies we will continue working towards more robust BCIs and also evaluate the present approach in feedback experiments. Since the minimax approach is an extreme case, we will furthermore pursue Bayesian approaches which may bridge the gap between this extreme method and classical CSP.

References

1. Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and Klaus-Robert Müller. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Proc Magazine*, 25(1):41–56, January 2008.
2. Jonathan R. Wolpaw, Niels Birbaumer, Dennis J. McFarland, Gert Pfurtscheller, and Theresa M. Vaughan. Brain-computer interfaces for communication and control. *Clin Neurophysiol*, 113:767–791, 2002.
3. J. del R. Millán. *Handbook of Brain Theory and Neural Networks*. MIT Press, Cambridge, 2002.
4. Guido Dornhege, José del R. Millán, Thilo Hinterberger, Dennis McFarland, and Klaus-Robert Müller, editors. *Toward Brain-Computer Interfacing*. MIT Press, Cambridge, MA, 2007.
5. Benjamin Blankertz, Guido Dornhege, Steven Lemm, Matthias Krauledat, Gabriel Curio, and Klaus-Robert Müller. The Berlin Brain-Computer Interface: Machine learning based detection of user specific brain states. *J Universal Computer Sci*, 12(6):581–607, 2006.
6. Thomas Elbert, Brigitte Rockstroh, Werner Lutzenberger, and Niels Birbaumer. Biofeedback of slow cortical potentials. I. *Electroencephalogr Clin Neurophysiol*, 48:293–301, 1980.
7. Anton Nijholt, Desnay Tan, Gert Pfurtscheller, Clemens Brunner, José Millán, Brendan Allison, Bernhard Grainmann, Florin Popescu, Benjamin Blankertz, and Klaus-Robert Müller. Brain-computer interfacing for intelligent systems. *IEEE Intelligent Systems*, 23(3):72–79, 2008.
8. Seung-Jean Kim, Alessandro Magnani, and Stephen Boyd. Robust fisher discriminant analysis. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 659–666. MIT Press, Cambridge, MA, 2006.
9. Z. J. Koles. The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. *Electroencephalogr Clin Neurophysiol*, 79(6):440–447, 1991.
10. Gert Pfurtscheller and F.H. Lopes da Silva. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin Neurophysiol*, 110(11):1842–1857, Nov 1999.
11. B. Blankertz, M. Kawanabe, R. Tomioka, F. Hohlefeld, V. Nikulin, and K-R Müller. Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing. In *Advances in Neural Information Processing Systems 20*. MIT Press, Cambridge, MA, 2008. in press.