Optimizing Spectral Filters for Single Trial EEG Classification

Ryota Tomioka[1,2], Guido Dornhege[2], Guido Nolte[2], Kazuyuki Aihara[1], and Klaus-Robert Müller[2]

[1] Dept. Mathematical Informatics, IST, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo, 113-8656, Japan
[2] Fraunhofer FIRST.IDA, Kekuléstr. 7, 12 489 Berlin, Germany

Abstract. We propose a novel spectral filter optimization algorithm for the single trial ElectroEncephaloGraphy (EEG) classification problem. The algorithm is designed to improve the classification accuracy of Common Spatial Pattern (CSP) based classifiers. The algorithm is based on a simple statistical criterion, and allows the user to incorporate any prior information one has about the spectrum of the signal. We show that with a different preprocessing, how a prior knowledge can drastically improve the classification or only be misleading. We also show a generalization of the CSP algorithm so that the CSP spatial projection can be recalculated after the optimization of the spectral filter. This leads to an iterative procedure of spectral and spatial filter update that further improves the classification accuracy, not only by imposing a spectral filter but also by choosing a better spatial projection.

1 Introduction

A Brain-Computer Interface (BCI) system provides a direct control pathway from human intentions to computer. Recently, a considerable amount of effort has been done in the development of a BCI system [1–5]. We will be focusing on non-invasive, electroencephalogram (EEG) based BCI systems. Such a device can give disabled people direct control over a neuroprosthesis or over a computer application as tools for communicating solely by their intentions that are reflected in their brain signals (e.g. [2]).

Recently, machine learning approaches to BCI have proven to be effective by making the subject training required in the classical framework unnecessary and compensating for the high inter-subject variability.

The task in this approach is to extract subject-specific discriminability patterns from high-dimensional spatio-temporal signals. With respect to the topographic patterns of brain rhythm modulations, the Common Spatial Patterns (CSP) (see [6, 7]) algorithm has proven to be very useful in extracting discriminative spatial projections. On the other hand, the frequency band on which the classifier operates is either selected manually or unspecifically set to a broad band filter [7, 3]. Naturally, an automatic method also for the selection of the frequency band is highly desirable [8, 9]. Here, we present a method for the spectral filter optimization problem, which is based on a simple statistical criterion. The proposed method is capable of handling arbitrary prior filters based on neurophysiological insights. The proposed method outperforms broad-band filtered CSP in most datasets. Moreover, a detailed validation shows how much of the gain is obtained by the theoretically obtained filter and how much is obtained by imposing a suitable prior filter. Based on the spectral filter obtained by the proposed method, one can also recalculate the CSP projection; this leads to iterative updating of spatio-spectral filter. We show that further improvements in the classification accuracy can be achieved by iteratively updating.

2 The algorithm

Let us denote by $X \in \mathbb{R}^{d\times T}$ the EEG signal of a single trial of imaginary motor movement[3], where d is the number of electrodes and T is the number of sampled time-points in a trial. We consider a binary classification problem where each class, e.g. right or left hand imaginary movement, is called positive $(+)$ or negative $(-)$ class. The task is to predict the class label for a single trial X.

Throughout this paper, we use a feature vector, namely *log-power features*, defined as follows:

$$\phi_j(X; \boldsymbol{w}_j, B_j) = \log \boldsymbol{w}_j^{\dagger} X B_j B_j^{\dagger} X^{\dagger} \boldsymbol{w}_j \qquad (j = 1, \ldots, J), \tag{1}$$

where the upper-script $\dagger$ denotes a conjugate transpose or a transpose for a real matrix, $\boldsymbol{w}_j \in \mathbb{R}^d$ is a spatial projection that projects the signal into a single dimension and $B_j \in \mathbb{R}^{T\times T}$ denotes the linear time-invariant temporal filter, which is an identity matrix in the case of conventional CSP algorithm. The training of a classifier is composed of two steps. In the first step, the coefficients $\boldsymbol{w}_j$ and B_j are optimized. In the second step, the Linear Discriminant Analysis (LDA) classifier is trained on the feature vector.

We use Common Spatial Pattern (CSP) algorithm [6, 7], a well known technique for the spatial filter optimization. Given a set of trials and the labels $\{X_i, y_i\}_{i=1}^n$ $(X_i \in \mathbb{R}^{d\times T},\ y_i \in \{+1, -1\})$, the CSP is formulated so that the projection maximize the power of the projected signal for one class and minimize that for the other class. This principle can be written as follows:

$$\max_{\boldsymbol{w}\in\mathbb{R}^d} \quad \frac{\boldsymbol{w}^{\dagger} \left\langle XX^{\dagger} \right\rangle_{+} \boldsymbol{w}}{\boldsymbol{w}^{\dagger} \left\langle XX^{\dagger} \right\rangle_{-} \boldsymbol{w}}. \tag{2}$$

where the angled brackets denote expectation within a class. Furthermore, it is known that the solution is easily obtained by solving the following generalized eigenvalue problem:

$$\Sigma^{+} \boldsymbol{w} = \lambda \Sigma^{-} \boldsymbol{w}, \tag{3}$$

[3] For simplicity, we assume that the trial mean is already subtracted and the signal is scaled by the inverse square root of the number of time-points. This can be achieved by a linear transformation $X = \frac{1}{\sqrt{T}} X_{\mathrm{original}} \left(I_T - \frac{1}{T}\mathbf{1}\mathbf{1}^{\dagger} \right)$.

where we call $\Sigma^c := \left\langle XX^\dagger \right\rangle_c \in \mathbb{R}^{d \times d}$ ($c \in \{+, -\}$) the sensor covariance matrix. The eigenvector corresponding to the largest eigenvalue of Eq. (3) is the optimum of the problem (2). In addition, the minimization of the problem (2) gives another projection that may be equivalently powerful in the classification. Moreover, it is often observed that the second or the third eigenvectors have fairly good discrimination. Therefore, we take the eigenvectors corresponding to the largest and the smallest n_{of} eigenvalues for each side. Thus, $J = 2n_{\text{of}}$ in Eq. (1).

Given a spatial projection, the next question is how to optimize the temporal filter B in Eq. (1). We formulate this problem in the frequency domain, because any time-invariant operation B is diagonalized in the frequency domain. We state the problem as follows:

$$\begin{aligned} &\max_{\boldsymbol{\alpha}} \frac{\langle s(\boldsymbol{w}, \boldsymbol{\alpha}) \rangle_+ - \langle s(\boldsymbol{w}, \boldsymbol{\alpha}) \rangle_-}{\sqrt{\mathrm{Var}\left[s(\boldsymbol{w}, \boldsymbol{\alpha})\right]_+ + \mathrm{Var}\left[s(\boldsymbol{w}, \boldsymbol{\alpha})\right]_-}}, \\ &\text{s.t. } \alpha_k \geq 0 \quad (\forall k = 1, \ldots, T), \end{aligned} \tag{4}$$

where we write the power spectrum of the signal projected with $\boldsymbol{w}$ as $\{s_k(\boldsymbol{w})\}_{k=1}^T$, the spectrum of the filter as $\boldsymbol{\alpha} := \{\alpha_k\}_{k=1}^T$ and $s(\boldsymbol{\alpha}, \boldsymbol{w}) := \sum_{k=1}^T \alpha_k s_k(\boldsymbol{w})$.

The optimal filter coefficient is explicitly written as follows:

$$\alpha_k^{(+)\text{opt}} \propto \begin{cases} \dfrac{\langle s_k(\boldsymbol{w}) \rangle_+ - \langle s_k(\boldsymbol{w}) \rangle_-}{\mathrm{Var}\left[s_k(\boldsymbol{w})\right]_+ + \mathrm{Var}\left[s_k(\boldsymbol{w})\right]_-} & \langle s_k(\boldsymbol{w}) \rangle_+ - \langle s_k(\boldsymbol{w}) \rangle_- \geq 0, \\ 0 & \text{otherwise}, \end{cases} \tag{5}$$

because the spatio-temporally filtered signal $s(\boldsymbol{w}, \boldsymbol{\alpha})$ is linear with respect to the spectral filter coefficients $\{\alpha_k\}_{k=1}^T$ and we additionally assume that the signal is a stationary Gaussian process, where the frequency components are independent to each other for a given class label; thus $\mathrm{Var}\left[s(\boldsymbol{w}, \boldsymbol{\alpha})\right]_c = \sum_{k=1}^T \alpha_k^2 \mathrm{Var}\left[s_k(\boldsymbol{w})\right]_c$. Note that the labels ($+$ and $-$) are exchanged for $\{\alpha_k^{(-)\text{opt}}\}_{k=1}^T$, the filter for the "$-$" class. The norm of the filter coefficients cannot be determined from the problem (4). Therefore, in practice we normalize the coefficients so that they sum to one.

Furthermore, we can incorporate our prior knowledge on the spectrum of the signal during the task. This can be achieved by generalizing from Eq. (5) to:

$$\alpha_k^{(c)} = \left(\alpha_k^{(c)\text{opt}}\right)^q \cdot (\beta_k)^p \qquad (c \in \{+, -\}), \tag{6}$$

where $\{\beta_k\}_{k=1}^T$ denotes the prior information, which we define specific to a problem (see Sec. 3). The optimal values for p and q should depend on the data, preprocessing, and the prior information $\{\beta_k\}_{k=1}^T$. Therefore one can choose them by cross validation.

Now, using the CSP projection $\boldsymbol{w}$ and the optimized spectral filter $\boldsymbol{\alpha}$, the log-power feature (Eq. (1)) is written as follows:

$$\phi_j(X; \boldsymbol{w}_j, \boldsymbol{\alpha}_j) = \log \sum_{k=1}^T \alpha_k^{(j)} \boldsymbol{w}_j^\dagger \hat{\boldsymbol{x}}_k \hat{\boldsymbol{x}}_k^\dagger \boldsymbol{w}_j \qquad (j = 1, \ldots, J), \tag{7}$$

where $\hat{\boldsymbol{x}}_k \in \mathbb{C}^d$ denotes the k-th component of the Fourier transform of X.

3 Results

3.1 Experimental setup

Data acquisition We use 162 datasets of motor-imagery BCI experiment from 29 healthy subjects. Each dataset contains EEG signal recorded during 70-600 (varying from a dataset to another at median 280) trials of one of the pairwise combinations of three motor imagination tasks, namely right hand (R), left hand (L) or foot (F) (see [9, 5] for the detail).

Preprocessing of the signals We band-pass filter the signal from 7-30Hz and cut out the interval of 500-3500ms after the appearance of the visual cue on the screen, which instructs the subject which imagination to perform, from the continuous EEG signal for each execution of imaginary movement as a trial. Only in Sec. 3.3, we also use the signal without the band-pass filter step, in order to investigate the effect of assuming this band (7-30Hz) on the design of a filter; except the band-pass filtering, the signal was equally processed as described above.

Classification We use the log-power feature (Eq. (1)) with $n_{\mathrm{of}} = 3$ features for each class and LDA as a classifier.

Prior information We test two prior filters $\{\beta_k\}_{k=1}^{T}$, namely:

- with the wide-band 7-30Hz assumption:

$$\beta_k = I_k^{[7,\,30]} \cdot \left(\langle s_k(\boldsymbol{w})\rangle_+ + \langle s_k(\boldsymbol{w})\rangle_-\right)/\,2, \tag{8}$$

- without the assumption:

$$\beta_k = \left(\langle s_k(\boldsymbol{w})\rangle_+ + \langle s_k(\boldsymbol{w})\rangle_-\right)/\,2, \tag{9}$$

where $I_k^{[7,\,30]}$ is an indicator function that takes value one only in the band 7-30Hz, and otherwise zero. Since we have already band-pass filtered the signal in order to calculate CSP, it is reasonable to restrict the resulting filter to take values only within this band. The second term, which is the average activity of two classes, express our understanding that in the motor imagery task, good discrimination is most likely be found at frequency bands that correspond to strong rhythmic activities, i.e., μ- and β-rhythms; the modulation of these rhythms is known as Event Related Desynchronization (ERD) and well studied. However, this might not be the case if we don't suppose the interesting signal to lie within the 7-30Hz interval as in the second prior filter (Eq. (9)). The comparison is shown in Sec. 3.3.

Furthermore, since the optimal filter (Eq. (5)) and the prior filter (Eqs. (8) or (9)) scale with powers -1 and 1 of the spectrum, respectively, we reparameterize the hyperparameters as $p = p' + q'$ and $q = q'$. Thus, if $p' = c$ the filter scales with the power c regardless of which q' is chosen. Therefore, the contributions of the scale and the discriminability are separated in the new parameterization. Now, for the prior filter (8), using p', the scaling exponent of the filter and q', the intensity of the label information, we can write Eq. (6) as follows:

$$\alpha_k \propto I_k^{[7,\,30]} \cdot \begin{cases} \left(\frac{\left(s_k^{(+)} - s_k^{(-)}\right)\left(s_k^{(+)} + s_k^{(-)}\right)}{v_k^{(+)} + v_k^{(-)}} \right)^{q'} \cdot \left(s_k^{(+)} + s_k^{(-)}\right)^{p'} & s_k^{(+)} - s_k^{(-)} \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

where the following short-hands are used: $s_k^{(c)} := \langle s_k(\boldsymbol{w}) \rangle_c$ and $v_k^{(c)} := \mathrm{Var}\left[s_k(\boldsymbol{w})\right]_c$. The filter with the prior filter (9) is simply Eq. (10) without the indicator $I_k^{[7,\,30]}$.

3.2 Comparison with CSP

First, we compare the proposed method with the prior filter (Eq. (8)) with conventional CSP [6, 7] algorithm. The spatial projection for the proposed method is the CSP itself. Therefore, the only difference is that we incorporate a non-homogeneous weighting of the spectrum (see Eq. (7)). The hyperparameters for the proposed method were fixed at $p' = 0$ and $q' = 1$ ($p = 1$ and $q = 1$ in the original parameterization), which corresponds to the direct product of Eqs. (5) and (8).

Figure 1 shows the 10×10 cross-validation errors of CSP and the proposed method for each dataset as a single point. Data-points lower than the diagonal correspond to datasets where the proposed method outperforms CSP.

As a visualization, the spectral filter corresponding to conventional CSP, the theoretically obtained filter (Eq. (5)), the prior filter (Eq. (8)) and the resulting spectral filter are shown in Fig. 2 for a CSP projection in a single dataset. The conventional CSP is purely an operation in the spatial domain. Therefore, as a spectral filter it has a flat spectrum as shown in the top-left corner. The proposed method (bottom-left corner) is a combination of the theoretically obtained filter (Eq. (5)) shown in the bottom center and the prior filter (Eq. (8)) shown in the bottom-right corner. The theoretically obtained filter (Eq. (5)) scales with the power -1 of the spectrum. This means that it compares frequency components with different ranges in a fair manner; the signal is first scaled down by a factor $1 / \sqrt{v_k^{(+)} + v_k^{(-)}}$ (whitening) and then summed with a weighting $(s_k^{(+)} - s_k^{(-)})_+ / \sqrt{v_k^{(+)} + v_k^{(-)}}$. This effect is clearly seen in the bottom center. The theoretically obtained filter has two peaks, one approximately at 12Hz and the other at 24Hz, although in the original scale the difference between two classes around 24Hz is hardly seen (top center). The scale -1 is also favorable from another point of view, namely invariance; one can apply an arbitrary (non-zero) spectral filter to the signal before calculating Eq. (5) yet the effect is canceled out by Eq. (5). On the other hand, since the signal is already band-pass filtered from 7-30Hz, a prior filter is always peaked at frequency components corresponding to strong rhythmic activities (e.g.., μ- or β-rhythm) regardless of whether they have discriminative information or not. The resulting filter (bottom left), which is a direct product of the two filters in this case (because $p = 1$ and $q = 1$), has two peaks but the peak at 12Hz is larger than the peak at 24Hz. The optimal combination of the theoretical optimum and the prior filter is discussed in the next session.

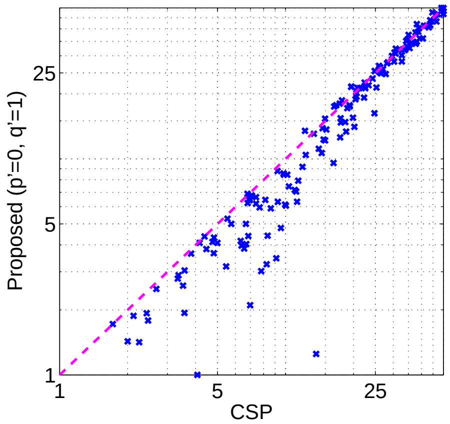


Fig. 1. The 10×10 cross-validation errors of conventional CSP and the proposed method on 162 datasets. Points lower than the diagonal correspond to datasets where the proposed method outperforms CSP. The conventional CSP weights the spectrum homogeneously (Eq. (1) with $B_j = I_T$ or Eq. (7) with $\alpha_k = 1\,(\forall k)$) while the proposed method weights the spectrum according to Eq. (10)). The hyperparameters were fixed at $p' = 0$ and $q' = 1$ (the direct product of Eqs. (5) and (8)).

3.3 Comparison of the two prior filters

In the previous section, we have shown that the combination of the theoretical optimum (Eq. (5)) and the prior filter (Eq. (8)) outperforms CSP in most datasets. However, it is still unclear whether the hyperparameters $p' = 0$ and $q' = 1$ are optimal or not. Furthermore, the range of validity of the prior filter (Eq. (8)) is not clear.

Therefore, in this section, we investigate two prior filters (Eqs. (8) and (9)). The first prior filter (Eq. (8)) focuses on the strong activity within the interval 7-30Hz. The second filter (Eq. (9)) also focuses on the strong activity but without the constraint, i.e., the wide-band 7-30Hz assumption.

In order to compare these two prior filters appropriately, we take the following two steps approach. In the first step, we optimize the spatial filter. Each dataset is band-pass filtered from 7-30Hz and the CSP projection with $n_{\text{of}} = 3$ patterns for each class is calculated on the whole dataset and fixed. In the second step, in order to investigate the optimal design of a spectral filter, we conduct a cross validation on the signal without pre-filtering.

Note that this validation differs from that in the previous section in two folds: firstly, the optimization of the spatial filter was done on the whole dataset in the first step and fixed during the validation, secondly, the spatial filter was calculated on the pre-filtered signal but applied on the signal without pre-filtering.

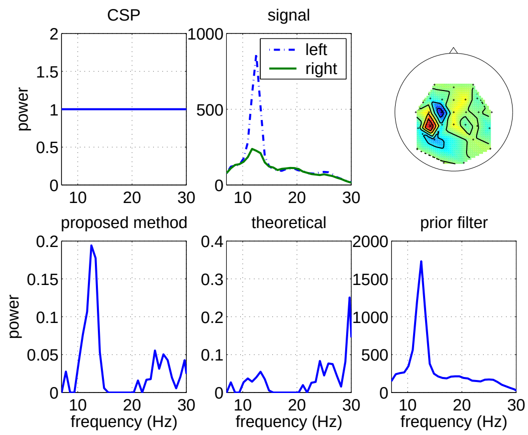
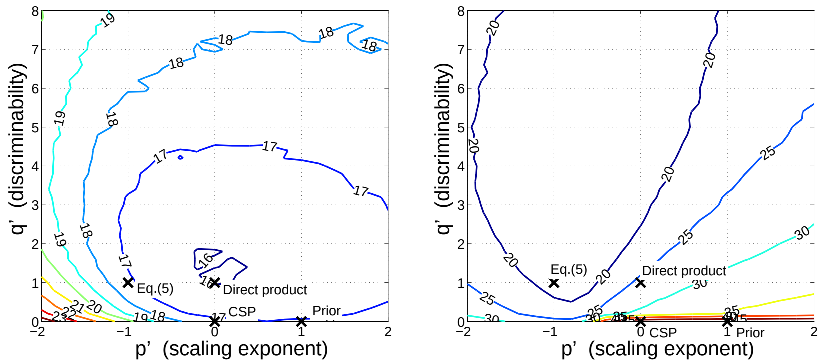


Fig. 2. (top center) The class-averaged spectrum of the original signal projected with a CSP projection shown in the top right corner. (top left) The conventional CSP in the spectral domain. (bottom left) The filter spectrum obtained by the proposed method. (bottom center) The theoretically obtained filter (Eq. (5)). (bottom right) The prior filter (Eq. (8)). (top right) The CSP projection topographically mapped on a head viewed from above. The head is facing the top of the paper.

Figures 3(a) and 3(b) show the contour plot of the average cross-validation error for all combinations of $p' \in [-2, 2]$ and $q' \in [0, 8]$ on a 0.2 interval grid for the two prior filters (Eqs. (8) and (9)), respectively. Figure 3(a) shows that the non-homogeneous weighting of the spectrum improves the classification accuracy ($p' = 0, q' = 1$ is better than $p' = 0, q' = 0$), which is consistent with Fig. 1, and incorporating the prior filter is also effective ($p' = 0, q' = 1$ is better than $p' = -1, q' = 1$). On the other hand, Fig. 3(b) shows a completely different picture. Since the wide-band assumption is not adopted in the prior filter (9), it weights not only μ- or β-band but also the strong brain activity lower than 7Hz, which does not correspond to motor imagery task or even which cannot be considered a rhythmic activity. Thus the prior information is not so much useful anymore. The basin of the classification error is now shifted to approximately $p' = -1$ where the spectrum is whitened. The theoretical optimum (Eq. (5)) is now in the region that gives minimum classification error. Note that however the overall error is lower in Fig. 3(a) compared to that in Fig. 3(b). Therefore, in practice the wide-band assumption appears to help though the aim of this section was to show that in general, without the wide-band assumption, it is necessary that one scales down the filter inversely to the power of the signal (Eq. (5)).

3.4 Iterative update of spatio-spectral filter

Although we have so far used the CSP projection as a spatial projection and focused on the optimization of the spectral filter, one can also recalculate the spatial projection after the optimization of the spectral filter. In order to incorporate the spectral filter, we generalize the definition of the sensor covari-



(a) with the wide-band 7-30Hz assumption (see Eq. (8)).

(b) without the assumption (see Eq. (9)).

Fig. 3. The contour plot of the average cross-validation errors over 162 datasets in the two dimensional hyperparameter space. Unlike in Sec. 3.2 or in Sec. 3.4 the cross-validation was carried out on the signal without pre-filtering with a pre-computed spatial pattern. Points corresponding to the CSP, the theoretically derived filter (Eq. (5)), the prior filter, and the direct product of the two filters ($p' = 0, q' = 1$) are marked. The cross validation is 4×4.

ance matrix Σ^c. Since the covariance matrix of the temporally filtered signal can be written as $V(\boldsymbol{\alpha}) := \sum_{k=1}^{T} \alpha_k V_k$, where $V_k = \hat{\boldsymbol{x}}_k \hat{\boldsymbol{x}}_k^\dagger \, (k = 1, \ldots, T)$ are the cross spectrum matrices, we solve the generalized eigenvalue problem $\Sigma^+(\boldsymbol{\alpha})\boldsymbol{w} = \lambda \Sigma^-(\boldsymbol{\alpha})\boldsymbol{w}$ instead of Eq. (3) for the recalculation of the spatial projection, where $\Sigma^c(\boldsymbol{\alpha}) := \langle V(\boldsymbol{\alpha}) \rangle_c$. Starting from uniform spectral coefficients $\alpha_k = 1 \, (\forall k)$, we alternately update the spectral filter and the spatial projection until convergence, because both steps depend on each other.

Figure 4 shows the improvements in the cross-validation error by iteratively updating spatio-spectral filter for six subjects. The odd steps correspond to the spatial projection updates and the even steps are spectral updates. Since the first step is CSP with homogeneous spectral filter and the second step is the proposed method without recalculation of the spatial projection, one can see that the major improvements occur by imposing a spectral filter (the second step). However, further improvements after the third step (e.g,. in subject C) were observed for many datasets. For some subjects (e.g., in subject F) no improvement in the cross-validation was observed, most likely due to artifacts whose effects are not localized in the frequency spectrum (e.g. blinking, chewing or other muscle movements).

4 Conclusion

In this paper, we have proposed a novel spectral filter optimization technique for CSP [6, 7] based single-trial EEG classifiers. The method is formulated in

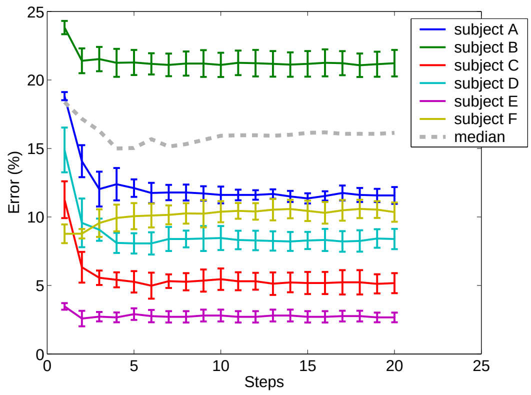


Fig. 4. The cross-validation errors of the iterative updating method for each step are shown for six subjects from very good classfication accuracy (subject E) to moderate accuracy (subject B). The median over 162 datasets is also shown (dashed line). The hyperparameters were fixed at $p' = 0$ and $q' = 1$ (the direct product of Eqs. (5) and (8)). The odd steps correspond to spatial projection updates and the even steps are spectral filter updates. Note that the first step is the conventional CSP itself and the second step is the proposed method without the recalculation of spatial projection.

the spectral domain, based on a simple statistical criterion. Thus the result is highly interpretable. The method is capable of handling arbitrary prior filter, which one can design based on the neuro-physiological understanding of the EEG signal during the task.

The cross validation on 162 BCI datasets show improved classification accuracy compared to the conventional CSP [6, 7]. In comparison to CSP, we have shown that the non-homogenous weighting of the spectrum improves the classification accuracy.

Moreover, we have investigated the best combination of the theoretically obtained filter (Eq. (5)) and the prior filter. We have tested two prior filters, namely the filter with the wide-band 7-30Hz assumption (Eq. (8)) and that without the assumption (Eq. (9)). We have found that with the wide-band assumption (Eq. (8)), the best combination is achieved approximately at $p' = 0, q' = 1$, which corresponds to the direct product of the theoretically obtained filter (Eq. (5)) and the prior filter (Eq. (8)); it is better than the conventional CSP ($p' = 0, q' = 0$), the theoretical optimum alone ($p' = -1, q' = 1$) or the prior filter ($p' = 1, q' = 0$). However, without the wide-band assumption, the prior filter, which assumes the discrimination to be found at frequency regions that is strongly active, fails because the activity below 7Hz will tend to dominate without contributing to discriminability. On the other hand, the theoretically optimal scale $p' = -1$, which whitens the signal, has proved to be favorable without the assumption. Thus, the prior filter is only valid with the wide-band assumption. In fact, we note that either CSP or the best combination $p' = 0, q' = 1$ already incorporates this prior knowledge that “strong activity implies good discrimination”, because both of them have the scale $p' = 0$.

Furthermore, we have tested an iterative updating algorithm of spatio-spectral filter. We have generalized the CSP algorithm to incorporate a non-homogeneous weighting of the cross spectrum matrices. The spatial filter and the spectral filter were updated alternately. We have found by cross validating each step of iteration that although for most datasets the major drop in the cross-validation error is observed when a spectral filter (the second step) was imposed on the original CSP pattern (the first step), further improvements in the cross-validation error were observed after the recalculation of CSP pattern (the third step) in many datasets.

The proposed method gives highly interpretable spatial filter naturally because we solve the generalized CSP problem. In addition, the spectral representation of the temporal filter is favorable not only from the interpretability but also from providing possibility to incorporate any prior information about the spectral structure of the signal as we have demonstrated in Sec. 3.

Acknowledgment: This research was partially supported by MEXT, Grant-in-Aid for JSPS fellows, 17-11866, 2006, by BMBF-grant FKZ 01IBB02A and by the PASCAL Network of Excellence (EU # 506778).

References

1. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain-computer interfaces for communication and control. Clin. Neurophysiol. **113** (2002) 767–791
2. Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., Perelmouter, J., Taub, E., Flor, H.: A spelling device for the paralysed. Nature **398** (1999) 297–298
3. Pfurtscheller, G., Neuper, C., Guger, C., Harkam, W., Ramoser, R., Schlögl, A., Obermaier, B., Pregenzer, M.: Current trends in Graz brain-computer interface (BCI). IEEE Trans. Rehab. Eng. **8**(2) (2000) 216–219
4. Blankertz, B., Dornhege, G., Schäfer, C., Krepki, R., Kohlmorgen, J., Müller, K.R., Kunzmann, V., Losch, F., Curio, G.: Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis. IEEE Trans. Neural Sys. Rehab. Eng. **11**(2) (2003) 127–131
5. Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.R., Kunzmann, V., Losch, F., Curio, G.: The Berlin Brain-Computer Interface: EEG-based communication without subject training. IEEE Trans. Neural Sys. Rehab. Eng. **14**(2) (2006) in press.
6. Koles, Z.J.: The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. Electroencephalogr. Clin. Neurophysiol. **79** (1991) 440–447
7. Ramoser, H., Müller-Gerking, J., Pfurtscheller, G.: Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans. Rehab. Eng. **8**(4) (2000) 441–446
8. Lemm, S., Blankertz, B., Curio, G., Müller, K.R.: Spatio-spectral filters for improved classification of single trial EEG. IEEE Trans. Biomed. Eng. **52**(9) (2005) 1541–1548
9. Dornhege, G., Blankertz, B., Krauledat, M., Losch, F., Curio, G., Müller, K.R.: Combined optimization of spatial and temporal filters for improving brain-computer interfacing. IEEE Trans. Biomed. Eng. (2006) accepted.