Optimizing Spatial Filters by Minimizing Within-Class Dissimilarities in EEG based BCI

Mahnaz Arvaneh et al.

IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS (2013)

Presenter : Younghak Shin

GIST, Dept. of Information and Communication, INFONET Lab.

Gwangju Institute of Science and Technology

Introduction

- A brain-computer interface (BCI) provides a direct communication pathway between the brain and an external device that is independent from any muscular signals.
- Through motor imagery or movement intentions, brain activities can be voluntarily decoded to control signals.
- In the majority of current BCI systems, the brain signals are measured by electroencephalogram (EEG), due to its low cost and high time resolution compared to other modalities.
- However, a major challenge in EEG-based BCI research is the inherent nonstationarity in the recorded signals.
- Variations of the signal properties from intra and inter sessions can be caused by changes of task involvement and attention, fatigue, changes in placement or impedance of the electrodes.
- Variations in the EEG signal can lead to deteriorated BCI performances.

Introduction

- Dealing with nonstationary changes in EEG-based BCIs has remained a challenging issue.
- This paper aiming to extract BCI features that are robust and invariant against the nonstationarities.
- For this purpose, we optimize the CSP spatial filters by minimizing the dissimilarities and variations in the train data.
- The CSP is a well known feature extraction method for motor imagery based BCI.
- The CSP algorithm is a feature extraction method that computes spatial filters maximizing the discrimination of the two classes.

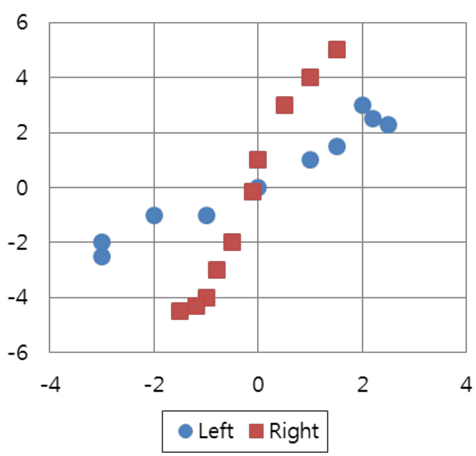
Motivation

- Despite the widespread use and the efficiency of CSP, its performance may be distorted by intrinsic variations in the signal properties.
- CSP only considers the separation of the means of the two classes, while the within-class scatter information is completely ignored.
- Since the EEG signals are nonstationary, there may be high trial-to-trial variations within a class that result in large scatters around the means in the feature space.
- This paper proposes a novel spatial filtering algorithm by defining a new criterion that simultaneously maximizes the discrimination between the class means, and minimizes the within-class dissimilarities.
- A Kullback–Leibler (KL) based term is defined to measure the within-class dissimilarities, the proposed algorithm is called KLCSP.

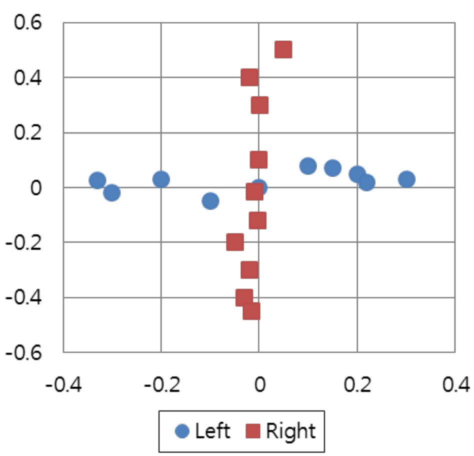
Method (Common Spatial Patterns)

- CSP linearly transforms EEG data to a spatially filtered space that the variance of one class is maximized while the variance of the other class is minimized.

[before CSP filtering]



[after CSP filtering]



- Since band-passed EEG measurements have approximately zero means, the normalized covariance matrix can be estimated as: $\Sigma = \frac{\mathbf{X}\mathbf{X}^{\mathrm{T}}}{\mathrm{trace}(\mathbf{X}\mathbf{X}^{\mathrm{T}})}$
- where $\mathbf{X} \in \mathbf{R}^{C \times S}$ denotes a single-trial EEG with C and S being the number of the channels and the measurement samples.
- The CSP algorithm projects **X** to spatially filtered **Z** as: $\mathbf{Z} = \mathbf{W}\mathbf{X}$
- Where the rows of the projection matrix **W** are the spatial filters.

Method (Common Spatial Patterns)

- **W** is generally computed by solving the eigenvalue decomposition problem:

$$\Sigma_1 \mathrm{W}^{\mathrm{T}} = \Sigma_2 \mathrm{W}^{\mathrm{T}} \Lambda$$

- Σ_1 and Σ_2 are, respectively, the average covariance matrices of each class.
- Λ is the diagonal matrix that contains the eigenvalues of $\Sigma_2^{-1}\Sigma_1$
- Since the eigenvalues in Λ indicate the ratio of the variances under the two conditions,
- The first and the last *m* rows of **W**, corresponding to the *m* largest and the *m* smallest eigenvalues, are generally used as the most discriminative filters.
- The CSP algorithm, in computing the projection matrix **W**, can be formulated as an optimization problem:

$$\min_{\mathrm{w}_i} \sum_{i=1}^{i=m} \mathrm{w}_i \Sigma_2 \mathrm{w}_i^{\mathrm{T}} + \sum_{i=m+1}^{i=2m} \mathrm{w}_i \Sigma_1 \mathrm{w}_i^{\mathrm{T}}$$

$$\text{Subject to:} \quad \mathrm{w}_i(\Sigma_1 + \Sigma_2)\mathrm{w}_i^{\mathrm{T}} = 1 \quad i = \{1, 2, \ldots, 2m\}$$

$$\mathrm{w}_i(\Sigma_1 + \Sigma_2)\mathrm{w}_j^{\mathrm{T}} = 0 \quad i, j = \{1, 2, \ldots, 2m\} \; i \neq j$$

- where the unknown weights $\mathbf{w}_i \in \mathbf{R}^{1 \times C}$, $i = \{1, \ldots, 2m\}$, respectively, indicate the first and the last *m* rows of the CSP projection matrix.

Method (Minimizing Within-Class Dissimilarities in CSP Filters)

- The CSP filters consider the discrimination of the average powers (means) of the two classes.
- However, the large discrimination between the class means does not guarantee to have compact features with small scatters around the means.
- Since the EEG signals are nonstationary, there may be high trial-to-trial variations within a class resulting in deteriorated BCI performances.
- This issue motivates to modify the CSP algorithm such that simultaneously the discrimination between the class means is maximized, and the within-class dissimilarities are minimized.
- For this purpose, first, the variations and dissimilarities between the trials of each class require to be measured.
- A natural choice for a dissimilarity metric is one that compares the probability distribution functions.
- A common possible choice, used in this paper, is the KL divergence or relative entropy.

Method (KLCSP)

- Given two probability distributions, $P_1(i)$ and $P_2(i)$ (taken as reference), the KL(Kullback Leibler) divergence is defined as

$$D(P_1(i)|P_2(i)) = \sum_i P_1(i) \ln\left(\frac{P_1(i)}{P_2(i)}\right).$$

- the KL divergence evaluates the dissimilarity between two distributions via the logarithm of their ratio weighted by the occurrence probability.
- In this paper, it is assumed that the nonstationarities exist only in the first two moments of the single-trial EEG (i.e.,mean and covariance)
- To measure the within-class dissimilarities of the EEG data, we split the training trials of each class into a number of consecutive epochs,
- And then we measure the dissimilarities between the distributions of each epoch and the average trials from the same class using the first two moments.
- The average distribution of a group of band-pass filtered EEG trials can be defined by a zero mean and a covariance matrix computed from averaging the covariance matrices over the multiple EEG trials.

Method (KLCSP)

- The most practical model for modeling the distribution of the EEG trials that is consistent with zero mean and a covariance matrix is Gaussian.
- The KL divergence between multivariate Gaussian distributions, *N*0*(μ*0*,*0*)* and *N*1*(μ*1*,*1*)*, has a closed-form expression:

$$D(N_0|N_1) = 0.5\Big[(\mu_1-\mu_0)^{\mathrm{T}}\Sigma_1^{-1}(\mu_1-\mu_0) + \mathrm{trace}(\Sigma_1^{-1}\Sigma_0) - \ln\left(\frac{\det(\Sigma_0)}{\det(\Sigma_1)}\right) - d\Big]$$

- where det and *d* denote the determinant function and the dimensionality of the data, respectively.
- So $D(N(0,\sum_w^t)|N(0,\sum_w)$ measures the dissimilarity of the distribution of the *t-th* epoch in class *ω* from the average distribution in class *ω*
- Minimizing the average within-class dissimilarities of the spatially filtered data is equivalent to minimizing the loss function

$$L\left(\begin{bmatrix} \mathrm{w}_1 \\ \mathrm{w}_2 \\ \cdots \\ \mathrm{w}_{2m} \end{bmatrix}\right) = L(\mathrm{w}) = \frac{1}{2}\sum_{\omega=1}^{2}\frac{1}{N_\omega}\sum_{t=1}^{N_\omega}\phi(\mathrm{w},\Sigma_\omega^t,\Sigma_\omega)$$
$$= \frac{1}{2}\sum_{\omega=1}^{2}\frac{1}{N_\omega}\sum_{t=1}^{N_\omega}D(N(0,\mathrm{w}\Sigma_\omega^t\mathrm{w}^T)|N(0,\mathrm{w}\Sigma_\omega\mathrm{w}^T))$$

- Where $\mathrm{w}=\begin{bmatrix} \mathrm{w}_1 \\ \mathrm{w}_2 \\ \cdots \\ \mathrm{w}_{2m} \end{bmatrix}$ is a matrix containing the first and the last *m* spatial filters, and *Nω* denotes the number of epochs belonging to class *ω*

Method (KLCSP)

- Adding the proposed loss function to the CSP optimization function results in spatial filters that simultaneously maximize the between-classes distance and minimize the within-class dissimilarities of the powers.
- Hence, the following optimization problem is proposed to obtain the optimized spatial filters:

$$\min_{\mathbf{w}_i} \ (1-r)\left(\sum_{i=1}^{i=m} \mathbf{w}_i \mathbf{C}_2 \mathbf{w}_i^T + \sum_{i=m+1}^{i=2m} \mathbf{w}_i \mathbf{C}_1 \mathbf{w}_i^T\right) + rL\left(\begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \cdots \\ \mathbf{w}_{2m} \end{bmatrix}\right)$$

$$\text{Subject to: } \mathbf{w}_i(\mathbf{C}_1+\mathbf{C}_2)\mathbf{w}_i^{\mathrm{T}}=1 \quad i=\{1,2,\ldots,2m\}$$

$$\mathbf{w}_i(\mathbf{C}_1+\mathbf{C}_2)\mathbf{w}_j^{\mathrm{T}}=0 \quad i,j=\{1,2,\ldots,2m\} \ \ i\neq j$$

- where *r (*0 ≤ *r* ≤ 1*)* is a regularization parameter to control the discrimination between and the similarity within the training classes.
- Each epoch contains *v* consecutive trials from the same class.
- In this paper, the best subject-specific *r* and *v* values are selected from small predefined sets by cross-validation.

Experiment (Data description)

- *Dataset IVa from BCI Competition III*
  - This publicly available dataset comprised EEG data from five healthy subjects recorded using 118 channels.
  - The subjects were instructed to perform one of two motor imagery tasks: right hand or foot.
  - 280 trials were available for each subject, whereby 168, 224, 84, 56, and 28trials formed the training sets for subjects aa, al, av, aw, and ay, respectively.
  - Subsequently, the remaining trials formed the test sets.
- *Neuro-Rehabilitation Dataset*
  - This dataset comprised a total of 132 sessions EEG data recorded from 11 hemiparetic stroke patients.
  - Each patient underwent 12 motor imagery-based BCI with robotic feedback neuro-rehabilitation sessions recorded over one month
  - The EEG data were acquired using 25 channels.
  - The experimental paradigm is shown in Fig. 1.

Experiments (Data description)

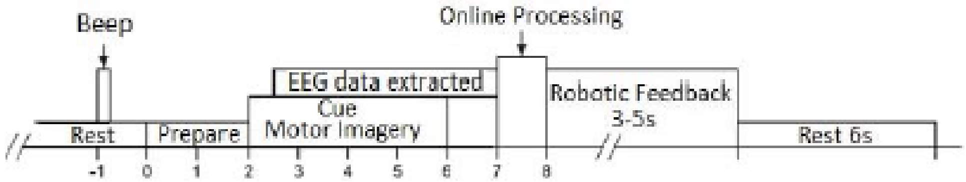


Fig. 1. Timing of each repeat in the neuro-rehabilitation sessions.

- The patient was first prepared with a visual cue for 2 s, then a “go” cue would instruct the patient to perform motor imagery of the impaired hand.
- If the voluntary motor intent was detected within the 4 s action period, the strapped MIT-Manus robot would assist the patient in moving the impaired limb toward the goal.
- There was a total of 160 repeats in each session (1 repeat means a complete run from preparation cue to the rest stage).
- There was a dedicated calibration(training) phase before the rehabilitation phase to train the online classifier.
- The classification problem involved distinguishing between the motor imagery stage and the rest stage.
- First 160 single-trials were considered as the training set, and the second 160 single-trials were considered as the test set.

Experiments (Data Processing)

- The performance of the proposed KLCSP algorithm was evaluated on the abovementioned datasets, and compared with the CSP and the sCSP algorithms.
- For each dataset, the EEG data from 0.5 to 2.5 s after the visual cue were used
- A bandpass filter from 8 to 30 Hz was used for filtering the EEG data, since this frequency band includes the range of frequencies that are mainly involved in performing motor imagery.
- The spatially filtered signals were obtained using the first and the last two spatial filters of (s/KL)CSP, $m = 2$.
- Finally, the variances of the spatially filtered signals were applied as the inputs of the LDA classifier.

Evaluation (Selecting the Parameters in KLCSP)

- In the KLCSP, two parameters are required to be optimally selected, the regularization parameter *r* and the number of trials in each epoch *v*.
- The best subject-specific *r* and *v* were selected from the sets of $r \in \{0.1, 0.2, \ldots, 0.9\}$ and $v \in \{1, 5, 10\}$, respectively.
- Five-fold cross-validation was performed for the different values of *r* and *v* on the train data and the ones resulting in the minimum error were chosen.
- To consider the changes over the time, in all the experiments, each epoch was constructed by a set of consecutive trials from the same class.
- It is noted that by selecting different numbers of trials in each epoch, nonstationarities and variations in different time-scales are taken into account.
- Considering a small number of trials in each epoch results in focusing on trial-by-trial changes, such as muscular artifacts,
- while increasing the number of trials shifts the focus into slower changes, such as variations of task involvement or fatigue.

Evaluation (Performance Comparison)

- In the first experiment, we compared the proposed KLCSP algorithm with the standard CSP and the sCSP algorithms using the first dataset.
- Table I presents the classification accuracies on the test data obtained by CSP, sCSP, and KLCSP

	aa	*al*	*av*	*aw*	*ay*	Mean ± Std
CSP	68.75	98.21	66.83	90.17	84.92	81.78 ± 13.6
sCSP	76.78	98.21	71.93	91.51	87.3	85.15 ± 10.7
KLCSP	79.46	98.21	69.89	91.96	90.07	85.92 ± 11.2

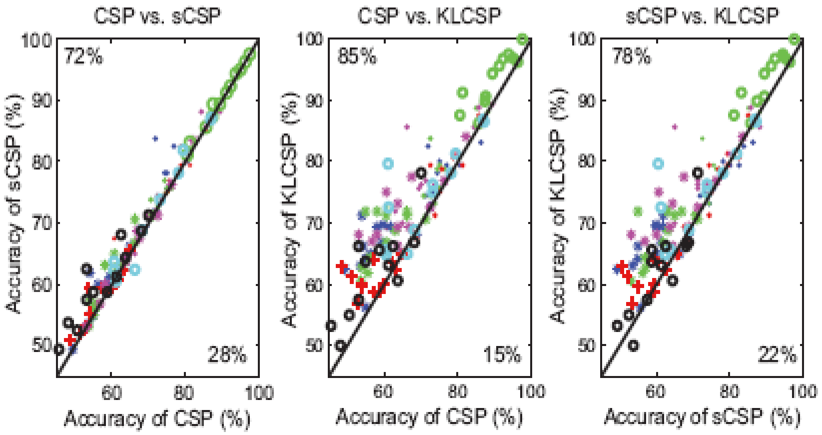
- The results showed that the proposed KLCSP algorithm outperformed the CSP and the sCSP algorithms by an average of 4.14% and 0.77%.
- Table II reports the average classification accuracies of the test sets from the neuro-rehabilitation dataset obtained by CSP, sCSP, and KLCSP.

Patient's Code	P003	P005	P007	P010	P012	P029	P034	P037	P044	P047	P050	Mean ± Std
CSP	57.5	71.32	89.80	65.44	57.02	60.07	56.81	74.17	73.54	74.89	75.00	68.69±11.94
sCSP	60.57	71.67	90.21	65.85	58.02	61.69	58.30	74.53	74.74	77.08	75.78	69.86±11.55
KLCSP	62.5	75.03	93.58	72.28	61.13	67.83	65.74	79.17	75.26	76.77	78.69	73.45±10.38

- The results showed that the KLCSP algorithm yielded the mean (median) accuracy of 73.43% (72.50%), whereas the CSP and the sCSP algorithms yielded the accuracies of 68.69% (67.81%), and 69.86% (68.75%).

Evaluation (Performance Comparison)

- Fig. 2 depicts scatter plots of the classification accuracies obtained from the neuro-rehabilitation dataset.



- Each plotted point on the sub-figures indicates the classification accuracy obtained from one of the 132 sessions.
- The points above the diagonal line mean the algorithm of the y-axis performed better than the one of the x-axis.
- The sCSP and the proposed KLCSP algorithms, respectively, outperformed the CSP algorithm in 95 and 112 sessions over the total 132 sessions,
- Interestingly, the last sub-figure showed that in 103 over 132sessions, the proposed KLCSP algorithm outperformed the sCSP algorithm.

Evaluation (Performance Comparison)

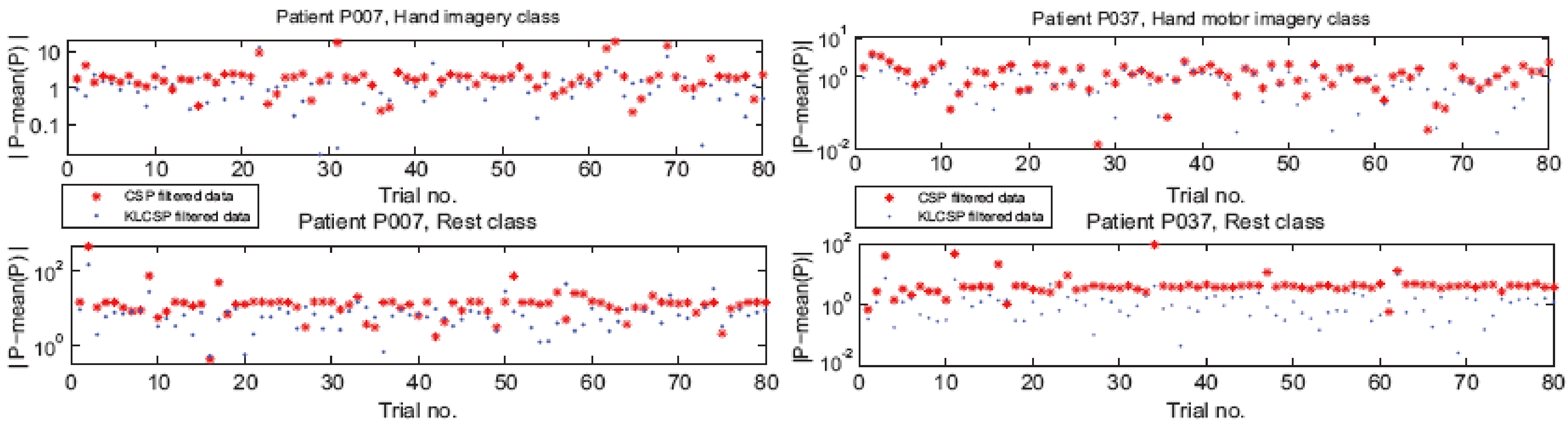
- Collecting all the results of the two aforementioned datasets and dividing them to three groups based on their CSP error rates.
- Table III investigates the performance of KLCSP and sCSP on the different BCI users.

Error Rate	0–15	15–30	>30	All
CSP mean (Median)	90.2(89.4)	75.7(75)	59.7(60.6)	69.2(68.3)
sCSP mean (Median)	90.3(89.4)	77.1(75.6)	61.3(61.2)	70.4(69.4)
KLCSP mean (Median)	91.8(90.6)	78.2(78.1)	66.4(66.2)	73.9(72.5)
p-value (sCSP versus CSP)	0.965	0.208	0.156	0.378
p-value (KLCSP versus CSP)	0.260	**0.008**	**<0.001**	**<0.001**
p-value (KLCSP versus sCSP)	0.292	0.177	**<0.001**	**0.008**

- First three rows compare the average(median) classification accuracies of the different groups obtained by CSP, sCSP, and KLCSP filters, respectively.
- Last three rows show the statistical wilcoxon test results.
- KLCSP algorithm improved the results of all the groups of the subjects, including those with poor, moderate, and high CSP performances
- The improvements from KLCSP for the subjects with moderate or poor CSP performances were statistically significant.
- KLCSP results revealed that for the subjects with poor CSP performances the KLCSP algorithm significantly outperformed the sCSP algorithm.

Evaluation (Toward Understanding the Merits of KLCSP)

- Fig. 3 shows the distance between the power of each trial and the average power of the corresponding class in the train sets after filtering by the best CSP and KLCSP filters.

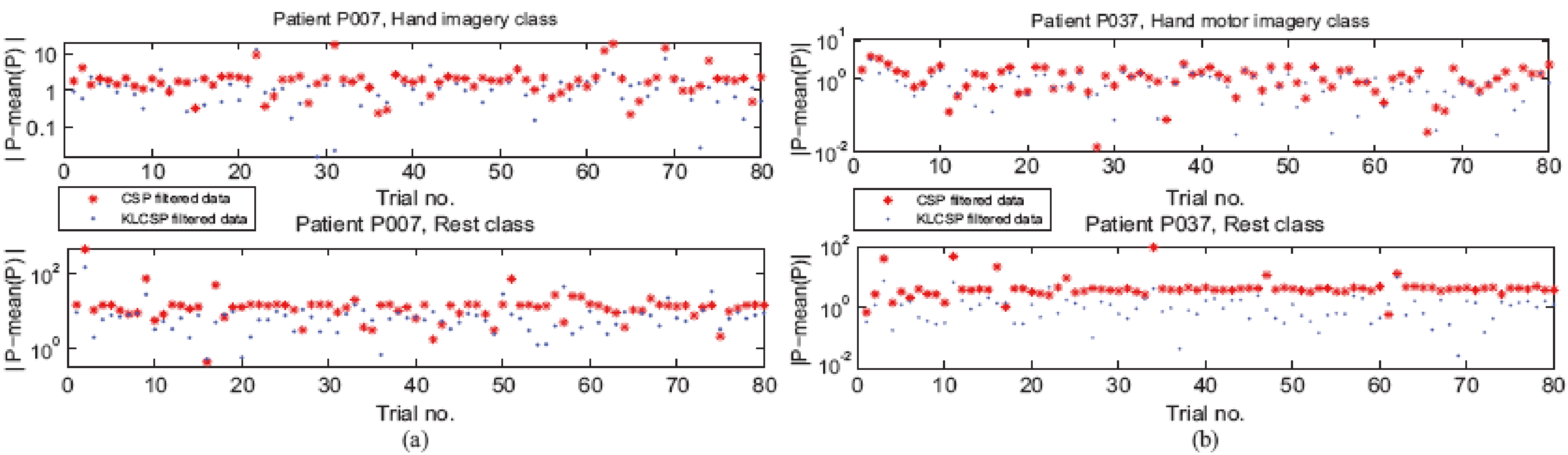


(a) (b)

- The best filters were defined by the fisher score of the corresponding features in the train set.
- Higher fisher score is obtained with maximum between class distance and minimum within class distance.
- Based on Fig. 3, shorter distances between the powers of the trials and the average power of the corresponding class indicate more similarities.

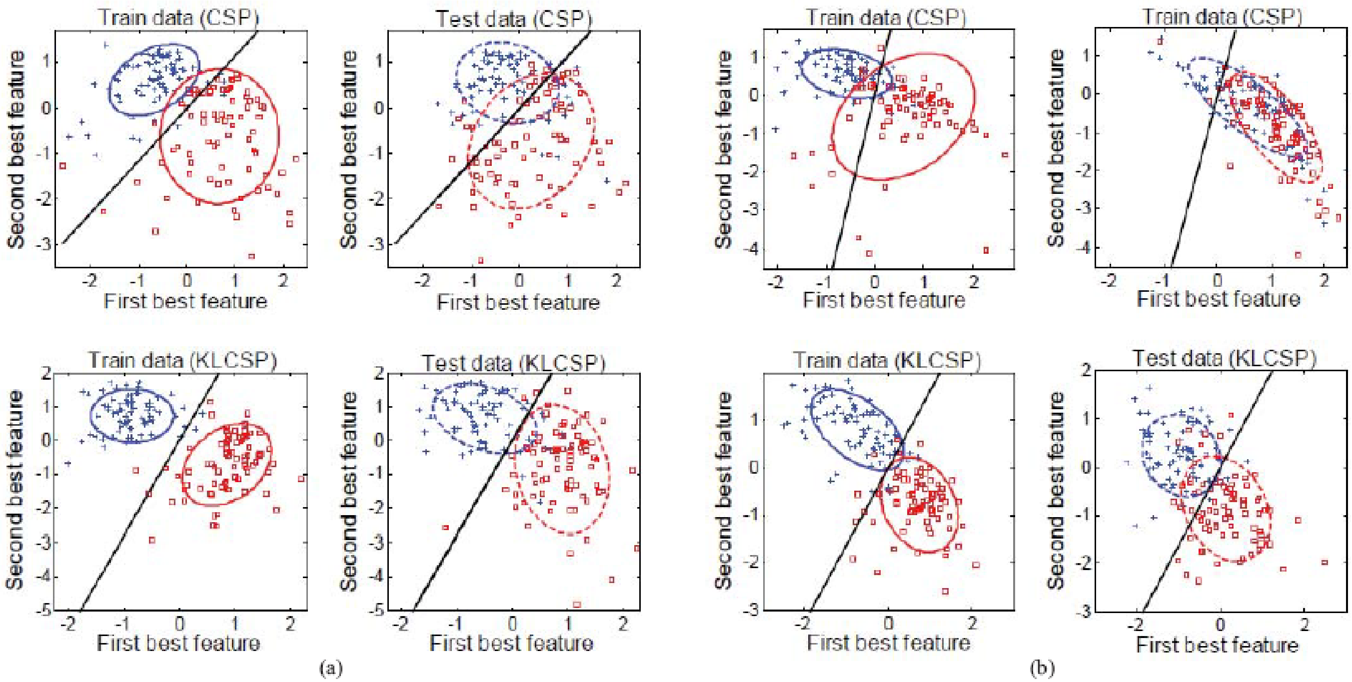
Evaluation (Toward Understanding the Merits of KLCSP)

- From Fig. 3, one can easily recognize higher dissimilarities and variations within the rest class as compared with the motor imagery class.
- In the neurorehabilitation dataset, since the rest class was a “no-command” state (the patients were allowed to do almost any other mental tasks than the impaired hand motor imagery), this class has high variations.
- The distances between the powers of the trials and the average power in the KLCSP filtered trials are mostly smaller than the CSP ones.



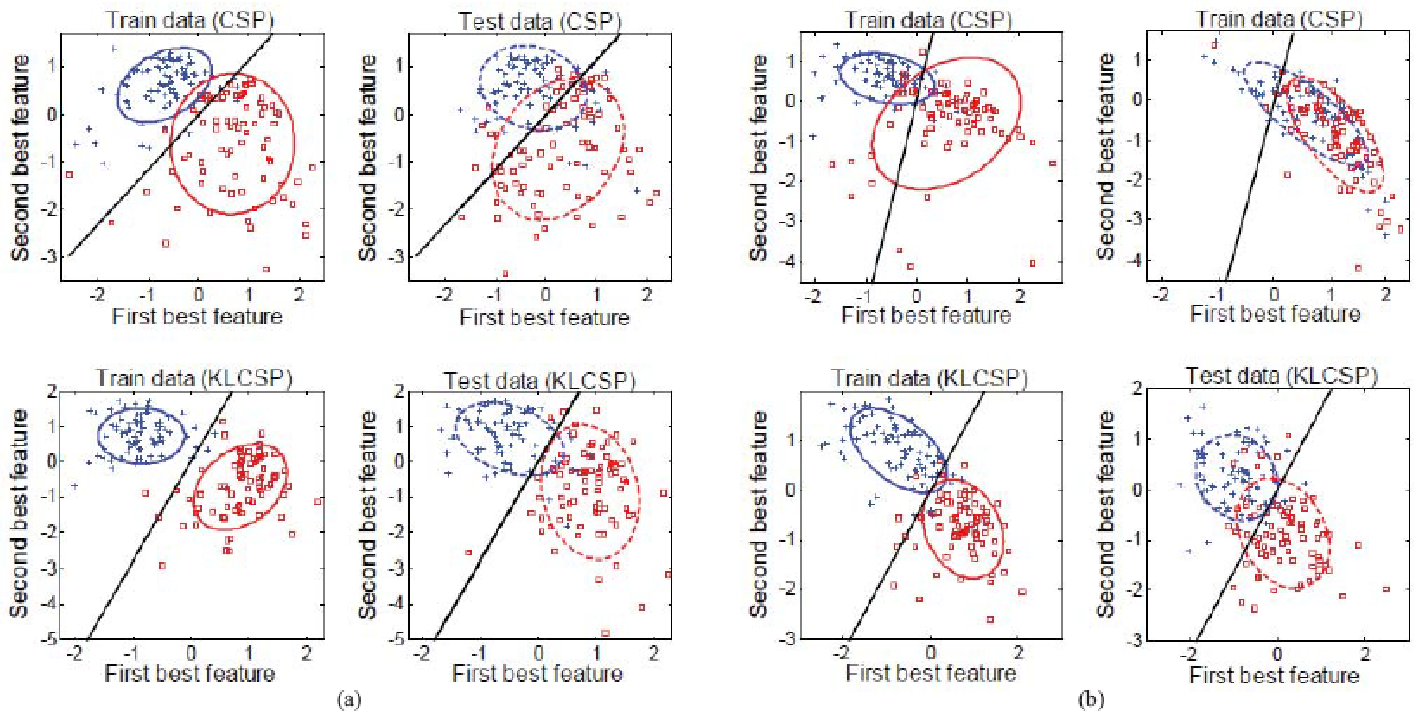
Evaluation (Toward Understanding the Merits of KLCSP)

- Fig. 4 shows the train and the test features obtained by CSP and the proposed KLCSP filters.



- It is noted that for the ease in visualization only two features which had the highest fisher scores in the train data were plotted.
- The blue and red squares denote the features of the hand motor imagery and the rest class, respectively.
- The black line represents the LDA hyperplane obtained by the train data.

Evaluation (Toward Understanding the Merits of KLCSP)



- The figure clearly reveals that the KLCSP features were more compact and thus more separable.
- transferring from the train to the test in CSP caused big shifts as well as big changes in the shape of the feature distributions.
- In contrast, the differences between the feature distributions of the train and the test sessions in KLCSP were almost limited to small shifts.

Evaluation (Toward Understanding the Merits of KLCSP)

- To better explain the performance differences between the CSP and the KLCSP algorithms, Fig. 5 compares some examples of the spatial filters.

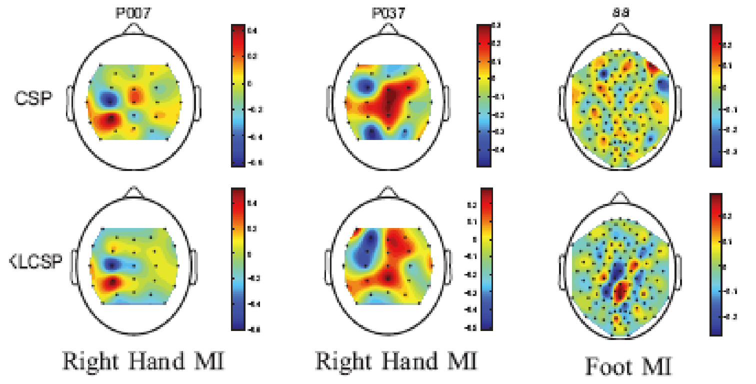


Fig. 5. Electrode weights for the corresponding filters obtained by CSP and KLCSP, for subjects P007 (performing right hand motor imagery), P037 (performing right hand motor imagery), and *aa* (performing foot motor imagery).

- For P007 and P037, although the CSP filters captured the relevant patterns over the left motor cortex, they were still affected by some nonstationarities and artifacts in some irrelevant channels.
- For *aa*, the CSP filter failed to capture the foot motor imagery pattern.
- On the contrary, the KLCSP algorithm extracted filters that are neurophysiologically more relevant, with strong weights over the relevant motor cortex areas and smooth weights over the other areas.

Conclusion

- This paper proposed a novel spatial filtering algorithm, KLCSP, to extract features that are robust and invariant against the nonstationarities in the EEG signals.
- This was achieved by defining a new criterion, that maximizes the discrimination between the classes while minimizes the within-class dissimilarities.
- Thus, a loss function was defined to measure the within-class dissimilarities based on the KL divergence, and it was imposed in the CSP optimization function.
- The experimental results demonstrated that the proposed KLCSP algorithm significantly outperformed the CSP and the sCSP algorithms by an average of 4.7% and 3.5%, respectively ($p < 0.01$).
- The results also showed that the KLCSP filtered signals had less within-class variations compared to the CSP ones.
- Moreover, plotting the feature distributions confirmed that the KLCSP features were more compact and more separable,

Thank you