Optimizing Spatial Filters by Minimizing Within-Class Dissimilarities in Electroencephalogram-Based Brain–Computer Interface

Mahnaz Arvaneh, *Student Member, IEEE*, Cuntai Guan, *Senior Member, IEEE*, Kai Keng Ang, *Member, IEEE*, and Chai Quek, *Senior Member, IEEE*

***Abstract*— A major challenge in electroencephalogram (EEG)-based brain–computer interfaces (BCIs) is the inherent nonstationarities in the EEG data. Variations of the signal properties from intra and inter sessions often lead to deteriorated BCI performances, as features extracted by methods such as common spatial patterns (CSP) are not invariant against the changes. To extract features that are robust and invariant, this paper proposes a novel spatial filtering algorithm called Kullback–Leibler (KL) CSP. The CSP algorithm only considers the discrimination between the means of the classes, but does not consider within-class scatters information. In contrast, the proposed KLCSP algorithm simultaneously maximizes the discrimination between the class means, and minimizes the within-class dissimilarities measured by a loss function based on the KL divergence. The performance of the proposed KLCSP algorithm is compared against two existing algorithms, CSP and stationary CSP (sCSP), using the publicly available BCI competition III dataset IVa and a large dataset from stroke patients performing neuro-rehabilitation. The results show that the proposed KLCSP algorithm significantly outperforms both the CSP and the sCSP algorithms, in terms of classification accuracy, by reducing within-class variations. This results in more compact and separable features.**

***Index Terms*— Brain–computer interface, common spatial patterns, EEG, nonstationary.**

I. 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 [1]–[5]. Through motor imagery or movement intentions, brain activities can be voluntarily decoded to control signals. Thus, BCIs enable users with severe motor disabilities to use their brain signals for communication and control [3], [4]. Furthermore, BCI has been used as a rehabilitation tool in restoring motor functions of patients with moderate to severe stroke impairments [6], [7]. In such a system, BCI could guide brain plasticity by demanding close attention to a motor task or by requiring the activation or deactivation of specific brain 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, such as functional magnetic resonance imaging (fMRI), functional near-infrared spectroscopy (fNIRS), etc., [4]. 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, or by artifacts, such as swallowing or blinking, among other reasons [8]. Variations in the EEG signal can lead to deteriorated BCI performances as most machine learning algorithms implicitly assume stationary data [9], [10].

Recently, several algorithms have been proposed to ameliorate the nonstationary effects in BCI applications. These approaches can be divided into two main groups, namely, the approaches adapting the model to the changes [10]–[19], and the approaches making the model robust and invariant against the changes [20]–[27].

The research studies on the former group showed that the BCI performance can be improved even by using simple adaptive methods, such as bias adaptation [10], [11]. Some studies chose adapting the classifier [12], [13], while some focused on the feature extraction [14]–[16], or the operational frequency bands [17]. One promising approach is covariate shift adaptation providing unsupervised adaptation to shifts in the feature distributions [15], [16]. Another work proposed adaptive classifiers based on the expectation-maximization method [12]. In addition, some recent studies successfully used techniques for co-adaptive learning of user and machine [18], [19].

Most of the algorithms in the latter group focused on extracting invariant features by regularizing the common spatial patterns (CSP) algorithm [22]–[24]. For example, the invariant common spatial patterns (iCSP) algorithm used extra measurements, such as electrooculogram (EOG) or

Manuscript received May 22, 2012; accepted January 5, 2013. Date of publication January 29, 2013; date of current version February 13, 2013. This work was supported by the Agency for Science, Technology and Research, Singapore.

M. Arvaneh is with the School of Computer Engineering, Nanyang Technological University, 639798 Singapore, and also with the Institute for Infocomm Research, Agency for Science, Technology and Research, 138632 Singapore (e-mail: stuma@i2r.a-star.edu.sg).

C. Guan and K. K. Ang are with the Institute for Infocomm Research, Agency for Science, Technology and Research, 138632 Singapore (e-mail: ctguan@i2r.a-star.edu.sg; kkang@i2r.a-star.edu.sg).

C. Quek is with the School of Computer Engineering, Nanyang Technological University, 639798 Singapore (e-mail: ashcqek@ntu.edu.sg).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2013.2239310

electromyogram (EMG) to improve the CSP features to be invariant against muscular or ocular artifacts [23]. Stationary CSP (sCSP) is another algorithm, which regularized CSP by penalizing the variations between covariance matrices [24]. There are also some works that improved the model by extracting the stationary part of the EEG signal before applying the CSP algorithm [25]–[27].

Despite various studies and recent advances, dealing with nonstationary changes in EEG-based BCIs has remained a challenging issue. This paper belongs to the latter group 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 algorithm is a feature extraction method that computes spatial filters maximizing the discrimination of the two classes [28], [29]. 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.

Motivated by this issue, 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. Since a Kullback–Leibler (KL) [30], [31] based term is defined to measure the within-class dissimilarities, the proposed algorithm is called KLCSP. In order to evaluate the performance of the proposed KLCSP algorithm, two datasets are used: the publicly available dataset IVa from BCI competition III [33] and a large dataset, including 132 sessions collected from stroke patients [6]. The KLCSP results are compared with the results obtained using the CSP and the sCSP algorithms, and some quantitative analysis and visualization techniques are provided to better understand the efficiency of the proposed algorithm.

The remainder of this paper is organized as follows. Section II describes the CSP algorithm and its extension, the proposed KLCSP algorithm, in detail. The applied datasets and the performed experiments are explained in Section III. Section IV presents the experimental results, followed by discussions in Section V. Finally, Section VI concludes this paper.

II. Methodology

A. Common Spatial Patterns

For classification of motor imagery tasks, spatial filters are widely used to find meaningful patterns from the noisy EEG data. Among different spatial filtering algorithms, CSP is so far the most commonly used algorithm in EEG-based BCI [28], [29]. It linearly transforms the band-pass filtered EEG data to a spatially filtered space, such that the variance of one class is maximized while the variance of the other class is minimized.

Since band-passed EEG measurements have approximately zero means, the normalized covariance matrix can be estimated as

$$\Sigma = \frac{\mathbf{X}\mathbf{X}^{\mathrm{T}}}{\mathrm{trace}(\mathbf{X}\mathbf{X}^{\mathrm{T}})} \tag{1}$$

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, respectively, T denotes the transpose operator, and trace $(\mathbf{X})$ gives the sum of the diagonal elements of $\mathbf{X}$.

The CSP algorithm projects $\mathbf{X}$ to spatially filtered $\mathbf{Z}$ as

$$\mathbf{Z} = \mathbf{W}\mathbf{X} \tag{2}$$

where the rows of the projection matrix $\mathbf{W}$ are the spatial filters. $\mathbf{W}$ is generally computed by simultaneous diagonalization of the average covariance matrices from the both classes. This is equivalent to solving the eigenvalue decomposition problem

$$\Sigma_1\mathbf{W}^{\mathrm{T}} = \Sigma_2\mathbf{W}^{\mathrm{T}}\Lambda \tag{3}$$

where Σ_1 and Σ_2 are, respectively, the average covariance matrices of each class; and 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 $\mathbf{W}$, corresponding to the m largest and the m smallest eigenvalues, are generally used as the most discriminative filters. Subsequently, the variances of the spatially filtered EEG data (possibly after a normalization and a log-transformation) are used as the features [29].

The CSP algorithm, in computing the projection matrix $\mathbf{W}$, can be formulated as an optimization problem [35] given by

$$\begin{aligned} \min_{\mathbf{w}_i} \quad & \sum_{i=1}^{i=m} \mathbf{w}_i\Sigma_2\mathbf{w}_i^{\mathrm{T}} + \sum_{i=m+1}^{i=2m} \mathbf{w}_i\Sigma_1\mathbf{w}_i^{\mathrm{T}} \\ \text{Subject to:} \quad & \mathbf{w}_i(\Sigma_1+\Sigma_2)\mathbf{w}_i^{\mathrm{T}} = 1 \quad i = \{1,2,\ldots,2m\} \\ & \mathbf{w}_i(\Sigma_1+\Sigma_2)\mathbf{w}_j^{\mathrm{T}} = 0 \quad i,j = \{1,2,\ldots,2m\} \;\; i \neq j \end{aligned} \tag{4}$$

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. Formulating the CSP algorithm as a quadratically constrained quadratic programming (QCQP) problem in (4) enables us to penalize the within-class dissimilarities in CSP by adding a penalty term (see Section II-B).

B. Minimizing Within-Class Dissimilarities in CSP Filters

The CSP filters are learned using the average covariance matrices. Hence, they actually model 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 [30], [31].

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

$$D(P_1(i)|P_2(i)) = \sum_i P_1(i) \ln\left(\frac{P_1(i)}{P_2(i)}\right). \quad (5)$$

The KL divergence provides a nonnegative measure, which is zero if and only if $P_1 = P_2$. As shown in (5), the KL divergence evaluates the dissimilarity between two distributions via the logarithm of their ratio weighted by the occurrence probability. This means that KL is not a sort of punctual difference between two distributions but rather a probability divergence.

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) [25]. Following this assumption, 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. Based on the maximum entropy principle, the most prudent model for modeling the distribution of the EEG trials that is consistent with zero mean and a covariance matrix is Gaussian [32].

The KL divergence between multivariate Gaussian distributions, $N_0(\mu_0, \Sigma_0)$ and $N_1(\mu_1, \Sigma_1)$, has a closed-form expression

$$D(N_0|N_1) = 0.5\Big[(\mu_1 - \mu_0)^{\mathrm{T}}\Sigma_1^{-1}(\mu_1 - \mu_0) + \mathrm{trace}(\Sigma_1^{-1}\Sigma_{\mathbf{0}}) - \ln\left(\frac{\det(\Sigma_0)}{\det(\Sigma_1)}\right) - d\Big] \quad (6)$$

where det and d denote the determinant function and the dimensionality of the data, respectively. So $D(N(0, \Sigma_\omega^t)|N(0, \Sigma_\omega))$ measures the dissimilarity of the distribution of the t^{th} epoch in class ω from the average distribution in class ω, where Σ_ω^t and Σ_ω, respectively, denote the average covariance matrices of the t^{th} epoch and the whole data belonging to class ω. Subsequently, minimizing the average within-class dissimilarities of the spatially filtered data is equivalent to minimizing the loss function

$$L\left(\begin{bmatrix}\mathbf{w}_1\\ \mathbf{w}_2\\ \cdots\\ \mathbf{w}_{2m}\end{bmatrix}\right) = L(\mathbf{w}) = \frac{1}{2}\sum_{\omega=1}^{2}\frac{1}{N_\omega}\sum_{t=1}^{N_\omega}\phi(\mathbf{w}, \Sigma_\omega^t, \Sigma_\omega) = \frac{1}{2}\sum_{\omega=1}^{2}\frac{1}{N_\omega}\sum_{t=1}^{N_\omega} D(N(0, \mathbf{w}\Sigma_\omega^t\mathbf{w}^T)|N(0, \mathbf{w}\Sigma_\omega\mathbf{w}^T)) \quad (7)$$

where $\mathbf{w}=\begin{bmatrix}\mathbf{w}_1\\ \mathbf{w}_2\\ \cdots\\ \mathbf{w}_{2m}\end{bmatrix}$ is a matrix containing the first and the last m spatial filters, and N_ω denotes the number of epochs belonging to class ω (i.e., similar to the loss function proposed in [27]).

Adding the proposed loss function (7) to the CSP optimization function (4) results in spatial filters that simultaneously maximize the between-classes distance of the powers (i.e., variances), and minimize the within-class dissimilarities of the powers. Hence, the following optimization problem is proposed to obtain the optimized spatial filters:

$$\begin{aligned}&\min_{\mathbf{w}_i}\ (1-r)\left(\sum_{i=1}^{i=m}\mathbf{w}_i\mathbf{C}_2\mathbf{w}_i^T + \sum_{i=m+1}^{i=2m}\mathbf{w}_i\mathbf{C}_1\mathbf{w}_i^T\right) + rL\left(\begin{bmatrix}\mathbf{w}_1\\ \mathbf{w}_2\\ \cdots\\ \mathbf{w}_{2m}\end{bmatrix}\right)\\ &\text{Subject to: } \mathbf{w}_i(\mathbf{C}_1+\mathbf{C}_2)\mathbf{w}_i^{\mathrm{T}} = 1 \quad i=\{1,2,\ldots,2m\}\\ &\quad \mathbf{w}_i(\mathbf{C}_1+\mathbf{C}_2)\mathbf{w}_j^{\mathrm{T}} = 0 \quad i,j=\{1,2,\ldots,2m\}\ \ i\neq j\end{aligned} \quad (8)$$

where $r\,(0 \le r \le 1)$ is a regularization parameter to control the discrimination between and the similarity within the training classes. Each epoch contains ν consecutive trials from the same class. In this paper, the best subject-specific r and ν values are selected from small predefined sets by cross-validation. Since the new KL divergence-based term is used, we abbreviate the proposed algorithm as KLCSP.

The proposed KLCSP algorithm is a nonlinear optimization problem, and due to the equality constraints it is a nonconvex optimization problem. There are several numerical optimization methods, such as augmented Lagrangian or sequential quadratic programming (SQP) that can be applied to find local minima of this problem. In this paper, the fmincon solver available in MATLAB based on SQP method was used [36], [37]. It is noted that all the $2m$ KLCSP filters are computed simultaneously. In addition, the spatial filters obtained from the CSP algorithm were used as the initial point. The motivation behind using the CSP filters for initialization is that when the regularization part is removed from the proposed cost function, the CSP filters are the global solution. Thus, starting from the CSP filters, the solver searches for filters improving the cost function compared to the CSP filters. Our experimental results on two datasets (see Section IV) also confirmed that although the KLCSP filters obtained from this initialization may not be the global solution, they significantly outperformed the CSP filters in terms of the classification accuracy.

C. KLCSP Versus sCSP

The sCSP algorithm [24] is another extension of the CSP algorithm proposed to make the CSP filters robust against the within-class variations and nonstationarities. In the sCSP algorithm, the dissimilarities between the average EEG trials and each epoch are measured by subtracting their covariance matrices. Then, to make the resultant difference matrix positive and subsequently usable in the Rayleigh quotation, the signs of its negative eigenvalues are flipped. As mentioned in [24], the proposed approximation is an upper bound for the initial nonflipped difference matrix, making the sCSP optimization directly solvable as a generalized eigenvalue problem. However, since the flipped difference matrices are different from the nonflipped ones, the sCSP algorithm may not lead to the optimal filters.

In contrast, the proposed KLCSP algorithm measures the nonstationarities based on the KL divergence that has a known interpretation for measuring the dissimilarities between two

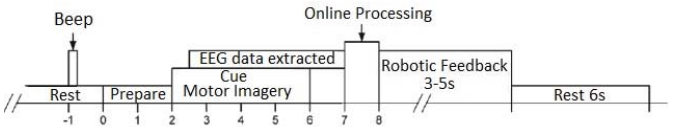

Fig. 1. Timing of each repeat in the neuro-rehabilitation sessions.

distributions [31]. It does not impose any approximations in measuring the dissimilarities ignoring the manifold structure of the covariance matrices. In addition, the constraints in the proposed KLCSP optimization function lead to diagonal covariance matrices in both the classes. Therefore, unlike sCSP, the projected signals by the obtained KLCSP filters are uncorrelated in both the classes. The proposed KLCSP algorithm considers the correlation between the projected signals to achieve a better discrimination. However, the KLCSP algorithm requires solving a nonconvex optimization problem yielding solutions that may be sub-optimal.

In Section III, these algorithms are compared in terms of the classification accuracy, and the advantages and disadvantages of them will be further discussed in Section V.

III. Experiments

A. Data Description

In this paper, the EEG data from two datasets were used. These two datasets are described as follows.

1) Dataset IVa from BCI Competition III [33]: This publicly available dataset comprised EEG data from five healthy subjects recorded using 118 channels. During the recording sessions, 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 28 trials formed the training sets for subjects *aa*, *al*, *av*, *aw*, and *ay*, respectively. Subsequently, the remaining trials formed the test sets. Using this dataset, the performance of the proposed KLCSP algorithm can be evaluated against different sizes of training data recorded from a large number of channels.

2) Neuro-Rehabilitation Dataset [6]: This large 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 (refer NCT00955838 in ClinicalTrials.gov) [6]. The EEG data were acquired using 25 channels. The experimental paradigm is shown in Fig. 1. In each repeat, 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. Finally, the patient was asked to rest for 6 s. 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 phase before the rehabilitation phase to train the online classifier.

In this paper, the classification problem involved distinguishing between the motor imagery stage and the rest stage. Therefore, each session comprised 160 motor imagery actions of the affected hand and 160 rest conditions. In this paper, the first 160 single-trials of each session were considered as the training set, and the second 160 single-trials were considered as the test set. Since variability and diversity in the rest class are more pronounced than motor imagery classes, this dataset is a proper choice to investigate the efficiency of the proposed algorithm.

B. 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 whereby the selected time segment was used by the winner of the BCI competition IV dataset IIa [34]. In this paper, a single band-pass filter from 8 to 30 Hz was used for filtering the EEG data, since this single frequency band includes the range of frequencies that are mainly involved in performing motor imagery. The filtering was performed using a fifth-order Butterworth filter. Thereafter, 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. Note that, in this paper, we did not reject any trials or electrodes.

IV. Evaluation

A. Selecting the Parameters in KLCSP

In the proposed KLCSP algorithm, two parameters are required to be optimally selected, namely, the regularization parameter r and the number of trials in each epoch ν. In this paper, the best subject-specific r and ν were selected from the sets of $r \in \{0.1, 0.2, \ldots, 0.9\}$ and $\nu \in \{1, 5, 10\}$, respectively, where the five-fold cross-validation was performed for the different values of r and ν on the train data and the ones resulting in the minimum error were chosen. After choosing the best subject-specific r and ν, the KLCSP filters were trained using the whole train data, and then evaluated on the test data.

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. In addition, there was no overlap between the epochs. Following this issue, during the five-fold cross-validation, after randomly selecting the evaluation trials, the remaining trials were ordered based on the time that they were recorded, and then each epoch was constructed using the ν 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.

To have a fair comparison, the same procedure as described above was applied to find the best subject-specific parameters of the sCSP algorithm (i.e., the regularization parameter and the number of trials in each epoch).

TABLE I
TEST CLASSIFICATION ACCURACIES OF DATASET IVA, BCI COMPETITION III, OBTAINED BY CSP, SCSP, AND THE PROPOSED KLCSP FILTERS

	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

B. Performance Comparison

In the first experiment, we compared the proposed KLCSP algorithm with the standard CSP and the sCSP algorithms using the dataset IVa from BCI competition III. Table I presents the classification accuracies on the test data obtained by CSP, sCSP, and KLCSP. The results showed that the proposed KLCSP algorithm outperformed the CSP and the sCSP algorithms by an average of 4.14% and 0.77%, respectively.

In the second experiment, we evaluated the proposed KLCSP algorithm using the neuro-rehabilitation dataset. This large dataset was recorded over 132 sessions from 11 stroke patients. The first half of each session was considered as the training set, and the second half was considered as the test set. Table II reports the average classification accuracies of the test sets from the neuro-rehabilitation dataset obtained by CSP, sCSP, and the proposed KLCSP algorithm. The results showed that the proposed KLCSP algorithm yielded the mean (median) accuracy of 73.43% (72.50%), whereas the CSP and the sCSP algorithms yielded the mean (median) accuracies of 68.69% (67.81%), and 69.86% (68.75%), respectively. Interestingly, for some patients, the average improvements achieved by the KLCSP algorithm are substantial (e.g., P034 by 8.9 and 7.4% average improvement against the corresponding CSP and sCSP results, respectively).

Fig. 2 depicts scatter plots of the classification accuracies obtained from the neuro-rehabilitation dataset. The first two sub-figures from the left side, respectively, compare the classification results of the sCSP and the proposed KLCSP algorithms against the CSP results. The classification results obtained using the proposed KLCSP algorithm and the sCSP algorithm are then compared in the last sub-figure. Each plotted point on the sub-figures indicates the classification accuracy obtained from one of the 132 sessions. The classification accuracies of the sessions belonging to a same patient were plotted with a same color and mark. As the classification accuracies were plotted, the points above the diagonal line mean the algorithm of the y-axis performed better than the one of the x-axis.

The results in Fig. 2 showed that the sCSP and the proposed KLCSP algorithms, respectively, outperformed the CSP algorithm in 95 and 112 sessions over the total 132 sessions, and for the rest of the sessions, the CSP results were only slightly better than the results obtained by sCSP and KLCSP. Interestingly, the last sub-figure showed that in 103 over 132 sessions, the proposed KLCSP algorithm outperformed the sCSP algorithm. With a closer look at the KLCSP results, it is realized that the biggest improvements were achieved by those subjects with CSP performances less than 75%.

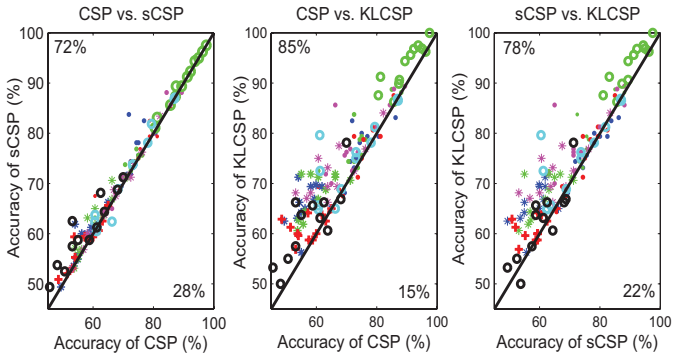

Fig. 2. Comparison of the classification accuracies using scatter plots for the neuro-rehabilitation dataset. Each plotted point on the sub-figures indicates the classification accuracy obtained from one of the 132 sessions. The accuracies of the sessions belonging to a same patient were plotted with a same color and mark.

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. The first three rows of this table compare the average(median) classification accuracies of the different groups obtained by CSP, sCSP, and the proposed KLCSP filters, respectively. Finally, the last three rows show the statistical wilcoxon signed-rank test results between CSP, sCSP, and KLCSP in the different groups. The results showed that the sCSP algorithm improved the classification accuracy of the subjects with moderate and poor CSP performances, although the improvements were not statistically significant. In contrast, the proposed KLCSP algorithm improved the results of all the groups of the subjects, including those with poor, moderate, and high CSP performances, and interestingly the improvements for the subjects with moderate or poor CSP performances were statistically significant. Comparing the sCSP and KLCSP results also revealed that for the subjects with poor CSP performances the proposed KLCSP algorithm significantly outperformed the sCSP algorithm.

C. Impact of KLCSP on Subjects With Different Qualities

To better understand the impact of the KLCSP algorithm on different subjects, each part of the KLCSP cost function, given in (8), was quantified for the dataset IVa from BCI competition III. The proposed KLCSP cost function comprises two parts. The first part, $\sum_{i=1}^{i=m}\mathbf{w}_i\mathbf{C}_2\mathbf{w}_i^T+\sum_{i=m+1}^{i=2m}\mathbf{w}_i\mathbf{C}_1\mathbf{w}_i^T$, is inversely related to the discrimination between the mean powers of the two classes. Thus, in this paper, $(\sum_{i=1}^{i=m}\mathbf{w}_i\mathbf{C}_2\mathbf{w}_i^T+\sum_{i=m+1}^{i=2m}\mathbf{w}_i\mathbf{C}_1\mathbf{w}_i^T)^{-1}$ is referred to as the between-classes distance. The second part, $L([\begin{smallmatrix}\mathbf{w}_1\\ \cdots\\ \mathbf{w}_{2m}\end{smallmatrix}])$ defined in (7), is directly related to the variations within the classes, and referred to as the within-class dissimilarities. In discrimination of the samples around the decision boundary, the amount of the within-class dissimilarities as well as the separation of the mean powers of the two classes is crucial.

Table IV presents the percentage changes of the between-classes distances and the within-class dissimilarities when the CSP filters were replaced by KLCSP filters, for the train sets of the dataset IVa. In this table, the qualities of the subjects

TABLE II
AVERAGE CLASSIFICATION ACCURACIES OF NEURO-REHABILITATION DATASET OBTAINED BY CSP, SCSP, AND THE PROPOSED KLCSP FILTERS

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

TABLE III
PERCENTAGE CHANGES OF BETWEEN-CLASSES DISTANCES AND WITHIN-CLASS DISSIMILARITIES, WHEN CSP FILTERS ARE REPLACED BY KLCSP FILTERS, FOR THE TRAIN SET OF DATASET IVA

	aa	*al*	*av*	*aw*	*ay*
Quality of subject	Poor	Good	Poor	Good	Moderate
Changes of between-classes dist. (%)	–7	–5.5	–7.1	–3.1	–7.9
Changes of within-class diss.(%)	-50	–6	–80	–21	–31

The p-value denotes the wilcoxon signed-rank test, and the bold values denote the significance with 1% level.

TABLE IV
OVERVIEW OF ALL THE RESULTS. GROUPING WAS PERFORMED BASED ON THE CSP ERROR RATE

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**

The qualities of the subjects were defined based on the CSP error rates. Dist. and diss. denote distance and dissimilarities, respectively.

were defined based on their CSP error rates (see Table I). To compute the aforementioned variables, the number of trials in each epoch was set to 1.

Table IV showed that the KLCSP algorithms were more valuable for subjects with poor initial CSP performances (e.g., *aa*, *av*), since compared to the CSP filters the KLCSP filters substantially reduced the within-class dissimilarities of these subjects, while the between-class distances were slightly decreased. The reason could be the fact that users with poor CSP performances often have noisy and nonstationary signals; thus CSP may fail to produce stable and discriminative signals. In contrast, KLCSP can alleviate the destructive effects of nonstationary and artifact-corrupted trials resulting in a better discrimination of the two classes. On the other hand, the subjects who performed well with CSP filters benefited less from applying the KLCSP filters, as the KLCSP filters only slightly reduced their within-class dissimilarities (e.g., *aa*). This makes sense, since these subjects have already well-separated and stable signals using the standard CSP filters. So there is no room to improve by KLCSP.

Comparing the results in Tables I and IV may raise a question regarding the subject *av*. Table IV showed that the KLCSP filters reduced the within-class dissimilarities of the train data for the subject *av* to one fifth of the CSP's amount, while the between-class distance was slightly decreased. Thus, one may expect to see a big improvement in the test classification accuracy of this subject using the KLCSP filters. However, KLCSP yielded only an improvement of 3.06%. Although 3% improvement is still remarkable in BCI applications, there is a need to investigate why KLCSP could not get more than 69.9% test accuracy for this subject despite substantial improvement of the discrimination in the train data. One possible reason may be some nonstationarities in the test data that could not be captured from the train data (e.g., changing the strategy of performing the mental tasks in the test session, among others). As a result, since KLCSP only uses the train data, it would not be able to truly deal with such nonstationarities. This issue will be investigated more in the future.

TABLE V
DETAILS OF THE TWO SELECTED SESSIONS FOR FURTHER ANALYSIS. THE PARAMETERS WERE SELECTED USING FIVE-FOLD CROSS-VALIDATION ON THE TRAIN DATA (ACC: ACCURACY)

Patient's Code	Session No.	Impaired Hand	Classification Acc. CSP	KLCSP	Selected Parameters α	ν
P007	7	Right	81.25	91.25	0.5	5
P037	4	Right	66.25	85.625	0.3	1

D. Toward Understanding the Merits of KLCSP

In the previous sections, we showed quantitative evidences indicating the proposed KLCSP can significantly improve the classification accuracy in EEG-based BCIs. In this section, we provide more analysis and visualizations to better understand the nature and the impact of our proposed algorithm on nonstationary changes in the EEG signals and the feature space.

The analysis was conducted with two sessions selected from the neuro-rehabilitation dataset, since they achieved two of the largest improvements in terms of the classification accuracy. Table V provides more details about the selected sessions, including the patient's code, the session number (as 12 sessions were recorded from each patient, it can be from 1 to 12), the test classification accuracies obtained by CSP and the proposed KLCSP algorithm, the regularization value r and the number of trials in each epoch ν selected by five-fold cross-validation on the train data.

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. Since the

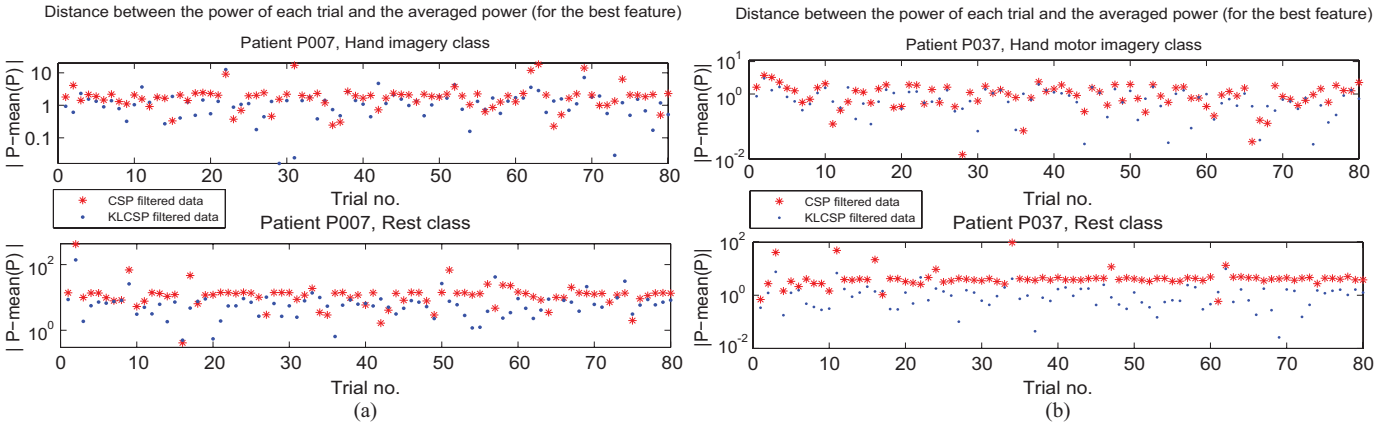

Fig. 3. Distance between the power of each trial and the average power of the corresponding class after filtering by the best CSP and the proposed KLCSP filters. (a) Patient P007, session 7. (b) Patient P037, session 4. In band-passed EEG trials, the power of a trial is equivalent to the variance of that trial. For better visualization, the y-axis were plotted in log scale.

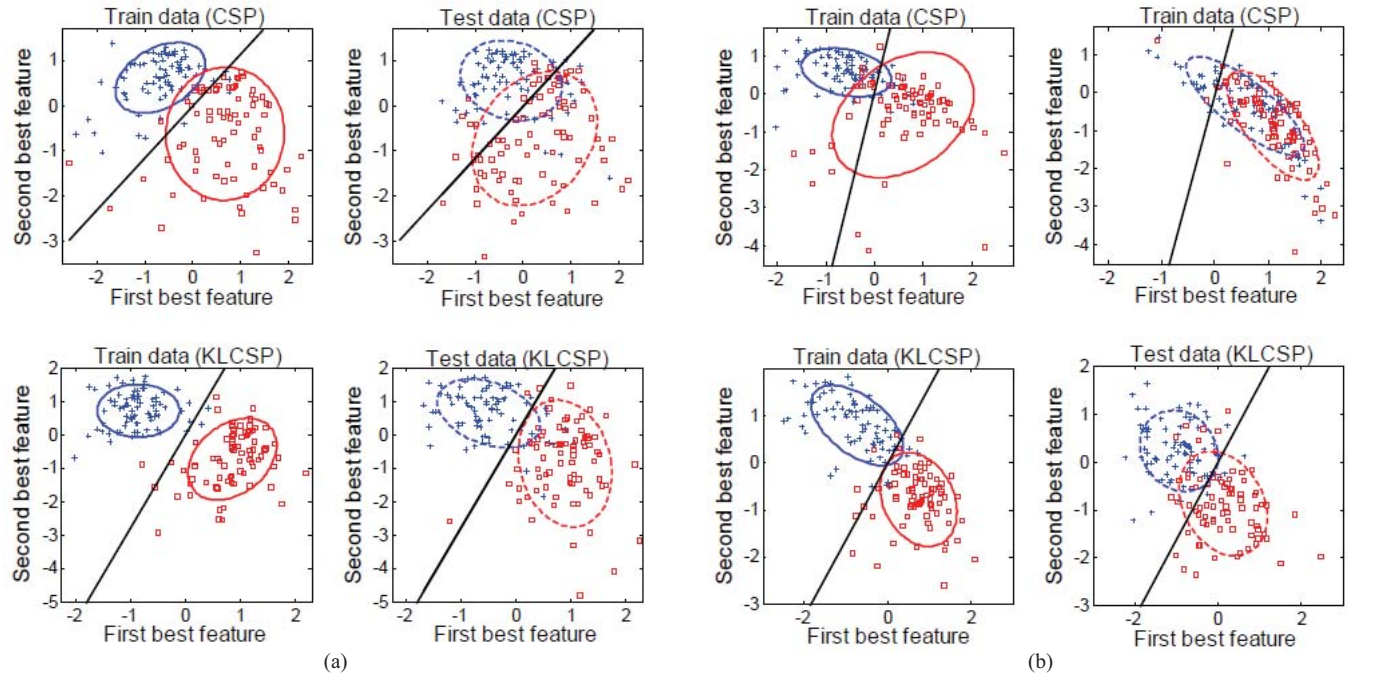

Fig. 4. Distributions of two best features obtained by CSP and the proposed KLCSP filters. (a) Patient P007, session 7. (b) Patient P037, session 4. The first row shows the features obtained by CSP filters and the second row shows the features obtained by the proposed KLCSP filters. The best features obtained using the fisher score on the train data. The blue crosses 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. The features were plotted after normalization.

EEG signals are centered, the power of each trial is equivalent to the variance of that trial. So in this part, the EEG data were filtered using the best CSP and KLCSP filters (the best filters were defined by the fisher score of the corresponding features in the train set), and the powers of the filtered signals were calculated. High trial-to-trial variations can yield high variations between the powers. On the other side, decreasing the dissimilarities can be interpreted by decreasing the distance between the powers. Therefore, this figure gives us an insight about the variabilities within each class after CSP and KLCSP filtering. Based on Fig. 3, shorter distances between the powers of the trials and the average power of the corresponding class indicate more similarities.

From Fig. 3, specifically the y-axes, 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 that the patients were allowed to do almost any other mental tasks than the impaired hand motor imagery, this class has high variations. Apart from this issue, as can be seen, the distances between the powers of the trials and the average power in the KLCSP filtered trials are mostly smaller than the CSP ones. With a closer look, we can see that the proposed KLCSP algorithm was able to efficiently damp most of the big variations (those trials with large deviation from the mean power), such as trials 3, 62, and 63 in the hand motor imagery class of Patient P007, or trials 3, 11, and 34 in the rest class of Patient P037.

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. Moreover, the features were plotted after the normalization. The blue crosses 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. Comparing the distributions of the train features extracted from CSP and the proposed KLCSP algorithm clearly reveals that the KLCSP features were more compact and thus more separable. Furthermore, 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. This shows that most of the nonstationary changes in the test data of this patient could be successfully captured by the proposed KLCSP filters, that may be due to a constant topography between the train and the test nonstationarities.

To better explain the performance differences between the CSP and the KLCSP algorithms, Fig. 5 compares some examples of the spatial filters. In general, this figure showed that the CSP filters presented large weights in several unexpected locations from the neurophysiological point of view. For P007 and P037, although the CSP filters captured the relevant patterns, i.e., dipole-like activations 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, as it was adversely affected by artifacts in the electrodes F3, F8, and FT8. On the contrary, the proposed KLCSP algorithm penalized the nonstationarities and the artifact-corrupted channels, and extracted filters that are neurophysiologically more relevant, with strong weights over the relevant motor cortex areas and smooth weights over the other areas.

Artifacts, such blinking and other muscle movements, produce voltage changes with much higher amplitude than the endogenous brain activity. Thus, artifact-corrupted trials usually have excessive dominant powers in some channels, resulting in CSP filters with big spatial weights for those channels [28] (see the *aa*s CSP filter in Fig. 5 as an example). However, in the KLCSP algorithm, the distribution of an artifact-corrupted trial is compared with the distribution of the average EEG trials from the same class using the proposed loss function (7). Since there are big differences between the powers of the affected channels in these two distributions, the KLCSP algorithm attenuates the spatial weights of those channels to minimize the loss function.

V. Discussion

The experimental results demonstrated the effectiveness of the proposed KLCSP algorithm. The experimental results showed that the proposed KLCSP optimization with the CSP filters as the initial point yielded spatial filters outperforming the sCSP algorithm, although the obtained KLCSP filters were not guaranteed to be the global solution. In particular, the KLCSP results of the subjects who had poor CSP performances (i.e., CSP error rate more than 30%) were significantly

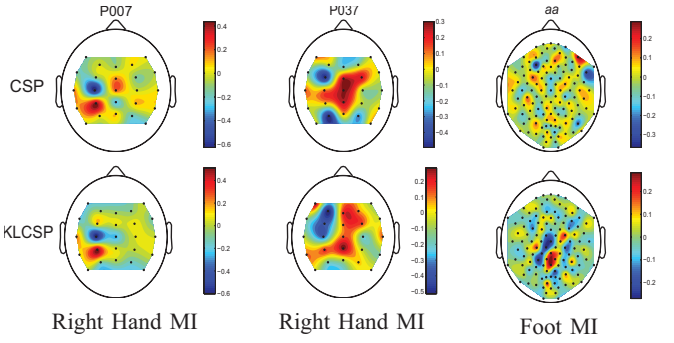

Fig. 5. Electrode weights for the corresponding filters obtained by CSP and KLCSP, for subjects P007 (performing right hand motor imagery), P037 (performing right hand motor imagery), and *aa* (performing foot motor imagery).

better than the corresponding sCSP results. Moreover, compared to the dataset IVa, the performance difference between KLCSP and sCSP in the neuro-rehabilitation dataset was more salient. This can be due to the fact that the neuro-rehabilitation dataset was more contaminated by noise, nonstationarities, and artifact-corrupted trials. All these results suggest that the proposed penalty term in the KLCSP algorithm could be more successful in capturing the nonstationarities and variations. However, it must be noted that the sCSP algorithm is computationally more efficient than the proposed KLCSP algorithm.

Regarding the computational complexity, the most time consuming part of the proposed KLCSP algorithm is related to finding the optimization parameters, r and ν, by cross-validation. However, this time can be substantially reduced using parallel computing. Moreover, in the lack of time, one can suffice to use a fixed value for the parameter r (i.e., number of trials in each epoch). In our experiments, in 47% of the cases, 5 was selected as the r value, while in 28% and 25% of the cases 10 and 1 were selected, respectively. Averagely, fixing the r value to 5 yielded the classification results around 0.8% less accurate than the classification results of using cross-validation in selecting r. So if it is required to reduce the computation time, using five trials in each epoch are recommended. The other fact is that when the number of channels is increased, the KLCSP algorithm requires more time to find the optimal filters. Nevertheless, using less than 30 channels and fixing the number of trials in each epoch to five, our experiments showed that it is truly possible to compute the KLCSP filters in a few minutes break between the calibration and feedback phases.

The KLCSP algorithm requires estimating the single-trial covariance matrices as well as the average covariance matrices. One question can be about the estimation of the single-trial covariance matrices. Since the EEG signals are contaminated by noise and artifacts, the sample-based covariance estimation can be adversely affected. Some algorithms, such as smoothing algorithms [38], may help to get a better estimation of the single-trial covariance matrix in some cases. However, they increase the computation time and may need some other recordings. Furthermore, the KLCSP algorithm inherently aims to detect such changes in the data by comparing the

covariance matrices of each trial/epoch with the average covariance matrix of the corresponding class, and learn spatial filters that are robust against the existing variations in the train data. Nevertheless, some preprocessing algorithms, such as selecting subject-specific frequency bands or time segments can increase the quality of the trials resulting in a better general performance for KLCSP.

The other issue is that the average covariance matrix for each class can also be distorted by noisy and artifact-corrupted trials. As a result, in the CSP algorithm, the distorted estimation of the covariance matrices may lead to poor spatial filters. As our experimental results showed, the proposed KLCSP algorithm automatically tries to reduce the destructive effects by enforcing the spatial filters to be less sensitive to artifacts and within-class variations, although the average covariance matrices may not be properly estimated. By the way, it is possible to combine KLCSP with the regularized covariance estimations to take the advantages of both the algorithms. We evaluated some of such algorithms on some randomly selected subjects, i.e., a KLCSP with diagonal loading [22] or a KLCSP with minimum covariance determinant estimate [39], among others. Unfortunately, none of them yielded classification accuracies noticeably higher than the corresponding KLCSP results. Thus, the KLCSP algorithm with a single regularization can be preferable. It is simpler and computationally more efficient.

VI. Conclusion

This paper proposed a novel spatial filtering algorithm for the EEG-based BCIs, called KLCSP, to extract features that are robust and invariant against the nonstationarities inherent 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 on five healthy subjects from the publicly available BCI competition III dataset IVa as well as 11 stroke patients performing neuro-rehabilitation in a total of 132 sessions 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 improvements were particularly more significant for the subjects with poor CSP performances.

The quantitative visualization 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, and the trained model using the proposed KLCSP algorithm was able to effectively capture the nonstationarities learned from the train data. In addition, comparing the spatial filters showed that the KLCSP filters were neurophysiologically more relevant, with strong weights over the relevant motor cortex areas and smooth weights over the other areas.

We would like to emphasize that the proposed algorithm does not require any additional recordings, and it is completely data-driven. The proposed algorithm only uses the train data to make the features robust and invariant against nonstationarities and variations rather than adapting to nonstationarities happening over the test sessions. Nevertheless, online adaptation algorithms may further enhance the proposed algorithm against unseen changes happening in the test data.

Acknowledgment

The authors would like to thank the participants and the entire research team at the Tan Tock Seng Hospital Rehabilitation Center, as well as everyone involved in the data collection from the stroke patients. They would also like to thank Berlin BCI Group for providing the data set IVa from BCI competition III, as well as H. Ahmadi and T. Ward for their constructive comments.

References

[1] N. Birbaumer, "Brain-computer-interface research: Coming of age," *Clin. Neurophysiol.*, vol. 117, no. 3, pp. 479–483, 2006.

[2] J. R. Wolpaw, D. J. McFarland, and T. M. Vaughan, "Brain-computer interface research at the Wadsworth Center," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 222–226, Jun. 2000.

[3] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, Jun. 2002.

[4] E. A. Curran and M. J. Stokes, "Learning to control brain activity: A review of the production and control of EEG components for driving brain-computer interface (BCI) systems," *Brain Cognit.*, vol. 51, no. 3, pp. 326–336, Apr. 2003.

[5] J. del R. Millan, J. Mourino, M. Franze, F. Cincotti, M. Varsta, J. Heikkonen, and F. Babiloni, "A local neural classifier for the recognition of EEG patterns associated to mental tasks," *IEEE Trans. Neural Netw.*, vol. 13, no. 3, pp. 678–686, May 2002.

[6] K. K. Ang, C. Guan, K. S. G Chua, T. B. Ang, C. W. K. Kuah, C. Wang, K. S. Phua, Z. Y. Chin, and H. Zhang, "A large clinical study on the ability of stroke patients to use EEG-based motor imagery brain-computer interface," *Clin. EEG Neurosci.*, vol. 42, no. 4, pp. 245–252, Oct. 2011.

[7] G. Pfurtscheller, G. R. Muller-Putz, R. Scherer, and C. Neuper, "Rehabilitation with brain-computer interface systems," *Computer*, vol. 41, no. 10, pp. 58–65, Oct. 2008.

[8] T. M. Vaughan, "Guest editorial brain-computer interface technology: A review of the second international meeting," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 94–109, Jun. 2003.

[9] M. Krauledat, G. Dornhege, B. Blankertz, and K.-R. Müller, "Robustifying EEG data analysis by removing outliers," *Chaos Complex. Lett.*, vol. 2, nos. 2–3, pp. 259–274, 2007.

[10] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, and K.-R. Müller, "Towards adaptive classification for BCI," *J. Neural Eng.*, vol. 3, no. 1, pp. R13-–R23, Mar. 2006.

[11] C. Vidaurre, M. Kawanabe, P. von Bünau, B. Blankertz, and K.-R. Müller, "Toward unsupervised adaptation of lda for brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 587–597, Mar. 2011.

[12] Y. Li and C. Guan, "An extended em algorithm for joint feature extraction and classifcation in brain-computer interfaces," *Neural Comput.*, vol. 18, no. 11, pp. 2730–2761, 2006.

[13] C. Vidaurre, A. Schlogl, R. Cabeza, R. Scherer, and G. Pfurtscheller, "Study of on-line adaptive discriminant analysis for EEG-based brain computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 3, pp. 550–556, Mar. 2007.

[14] S. Sun and C. Zhang, "Adaptive feature extraction for EEG signal classification," *Med. Bio. Eng. Comput.*, vol. 44, no. 10, pp. 931–935, Sep. 2006.

[15] Y. Li, H. Kambara, Y. Koike, and M. Sugiyama, "Application of covariate shift adaptation techniques in brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 6, pp. 1318–1324, Jun. 2010.

[16] M. Sugiyama, M. Krauledat, and K.-R. Müller, "Covariate shift adaptation by importance weighted cross validation," *J. Mach. Learn. Res.*, vol. 8, pp. 985–1005, Dec. 2007.

[17] K. P. Thomas, C. Guan, C. T. Lau, V. A. Prasad, and K. K. Ang, "Adaptive tracking of discriminative frequency components in EEG for a robust brain-computer interface," *J. Neural Eng.*, vol. 8, no. 3, pp. 1–15, Apr. 2011.
[18] S. Lu, C. Guan, and H. Zhang, "Unsupervised brain computer interface based on intersubject information and online adaptation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 17, no. 2, pp. 135–145, Apr. 2009.
[19] C. Vidaurre, C. Sannelli, K.-R. Müller, and B. Blankertz, "Machine-learning-based coadaptive calibration for brain-computer interfaces," *Neural Comput.*, vol. 23, no. 3, pp. 791–816, 2011.
[20] C. Gouy-Pailler, M. Congedo, C. Brunner, C. Jutten, and G. Pfurtscheller, "Nonstationary Brain source separation for multiclass motor imagery," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 2, pp. 469–478, Mar. 2010.
[21] M. Zhong and M. Girolami, "A Bayesian approach to approximate joint diagonalization of square matrices," in *Proc. 29th Int. Conf. Mach. Learn.*, Jun. 2012, pp. 1–8.
[22] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, Feb. 2010.
[23] B. Blankertz, M. Kawanabe, R. Tomioka, F. Hohlefeld, V. Nikulin, and K.-R. Müller, "Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing," *Adv. Neural Inf. Process. Syst.*, vol. 20, pp. 1–8, Feb. 2008.
[24] W. Samek, C. Vidaurre, K.-R. Müller, and M. Kawanabe, "Stationary common spatial patterns for brain-computer interfacing," *J. Neural Eng.*, vol. 9, no. 2, p. 0260134, Feb. 2012.
[25] P. von Bünau, F. C. Meinecke, F. Király, and K.-R. Müller, "Finding stationary subspaces in multivariate time series," *Phys. Rev. Lett.*, vol. 103, no. 21, pp. 1–4, Nov. 2009.
[26] P. von Bünau, F. C. Meinecke, S. Scholler, and K.-R. Müller, "Finding stationary brain sources in EEG data," in *Proc. 27th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2010, pp. 2810–2813.
[27] W. Samek, M. Kawanabe, and C. Vidaurre, "Group-wise stationary subspace analysis—A novel method for studying non-stationarities," in *Proc. 5th Int. Brain-Comput. Inter. Conf.*, 2011, pp. 16–20.
[28] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2008.
[29] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–6, Dec. 2000.
[30] Q. V. Vincent, Y. Bin, and E. K. Robert, "Information in the nonstationary case," *Neural Comput.*, vol. 21, no. 3, pp. 688–703, Mar. 2009.
[31] S. Kullback, *Information Theory and Statistics*, London, U.K.: Peter Smith, 1978.
[32] E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.*, vol. 160, no. 4, pp. 620–630, May 1957.
[33] G. Dornhege, B. Blankertz, G. Curio, and K. R. Müller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun. 2004.
[34] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers Neurosci.*, vol. 6, pp. 1–9, Mar. 2012.
[35] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing the channel selection and classification accuracy in EEG-based BCI," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1865–1873, Jun. 2011.
[36] M. Powell, "A fast algorithm for nonlinearly constrained optimization calculations," in *Numerical Analysis*, vol. 630, G. Watson, Ed. Berlin, Germany: Springer-Verlag, 1978, pp. 144–157.
[37] M. Powell, "Variable metric methods for constrained optimization," in *Computing Methods in Applied Sciences and Engineering*, vol. 704, R. Glowinski, J. L. Lions, and I. Laboria, Eds. Berlin, Germany: Springer-Verlag, 1979, pp. 62–72.
[38] S. H. Hass, M. G. Frei, I. Osorio, B. Pasik-Duncan, and J. Radel, "EEG ocular artifact removal through ARMAX model system identification using extended least squares," *Commun. Format. Syst.*, vol. 3, no. 1, pp. 19–40, Jun. 2003.
[39] C. Croux and G. Haesbroeck, "Influence Function and Efficiency of the Minimum Covariance Determinant Scatter Matrix Estimator," *J. Multivariate Anal.*, vol. 71, pp. 161–190, Nov. 1999.

Mahnaz Arvaneh (S'11) received the B.Sc. degree in electrical engineering from K. N. Toosi University of Technology, Tehran, Iran, and the M.Sc. degree in control engineering from the Ferdowsi University of Mashhad, Mashhad, Iran, in 2005 and 2007, respectively. She is currently pursuing the Ph.D. degree with Singapore Nanyang Technological University.

She was with the R&D Department, Mega-motor Co., Tehran, from 2008 to 2009. In 2012, she was a Visiting Student with the Electrical Engineering Department, National University of Ireland, Maynooth, Ireland. Her current research interests include brain–computer interfaces, signal processing, machine learning, and pattern recognition.

Cuntai Guan (S'91–M'92–SM'03) received the Ph.D. degree in electrical and electronic engineering from Southeast University, Nanjing, China, in 1993.

He is currently a Principal Scientist with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore. He is the Department Head of Neural and Biomedical Technology, Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore. He has authored or co-authored over 150 papers in refereed journals and conferences, and holds 13 patents and pending patents. His current research interests include neural and biomedical signal processing, neural and cognitive process and its clinical applications, and brain–computer interface algorithms, systems, and applications.

Dr. Guan is an Associate Editor of the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, and *Frontiers in Neuroprosthetics*.

Kai Keng Ang (S'05–M'07) received the B.A.Sc. (First Class Hons.) and the Ph.D. degrees in computer engineering from Nanyang Technological University, Singapore.

He is currently the Brain-Computer Interface Laboratory Head and a Scientist with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore. He was a Senior Software Engineer with Delphi Automotive Systems Singapore Pte. Ltd., from 1999 to 2003, where he was involved in research on embedded software for automotive engine controllers. He has authored or co-authored several papers. His current research interests include brain–computer interfaces, computational intelligence, machine learning, pattern recognition, and signal processing.

Chai Quek (M'96–SM'10) received the B.Sc. degree in electrical and electronics engineering and the Ph.D. degree in intelligent control from Heriot-Watt University, Edinburgh, U.K.

He is currently an Associate Professor, a member of the Centre for Computational Intelligence (formerly the Intelligent Systems Laboratory), and the Assistant Chair with the School of Computer Engineering, Nanyang Technological University, Singapore. His current research interests include intelligent control, intelligent architectures, artificial intelligence in education, neural networks, fuzzy neural systems, neurocognitive informatics, and genetic algorithms.

Dr. Quek is a member of the IEEE Technical Committee on Computational Finance and Economics.