IETE Journal of Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tijr20

A New Fast Approach for an EEG-based Motor Imagery BCI Classification

Mohammad Ali Amirabadi & Mohammad Hossein Kahaei

To cite this article: Mohammad Ali Amirabadi & Mohammad Hossein Kahaei (2020): A New Fast Approach for an EEG-based Motor Imagery BCI Classification, IETE Journal of Research, DOI: 10.1080/03772063.2020.1816221

To link to this article: https://doi.org/10.1080/03772063.2020.1816221

Published online: 13 Sep 2020.

IETE JOURNAL OF RESEARCH
https://doi.org/10.1080/03772063.2020.1816221

A New Fast Approach for an EEG-based Motor Imagery BCI Classification

Mohammad Ali Amirabadi and Mohammad Hossein Kahaei

School of Electrical Engineering, Iran University of Science and Technology (IUST), Tehran 1684613114, Iran

ABSTRACT
Nowadays, Brain Computer Interface has an important role in the life quality of paralyzed people. However, this technique is mainly affected by the quality of the recorded signal in each trial. This problem could be solved by rejecting low-quality trials. But developing the processing based on the recorded signal from the brain, which is a mixture of the target signal plus noise and artifact, would not be favourable in situations that all trials have low quality. This paper solves this problem by presenting a new fast algorithm for separating recorded source signals. Results indicate the improvement in classification accuracy of the proposed method compared with the well-known state of the art works.

KEYWORDS
Brain Computer Interface; Fast independent component analysis; Support vector machine; Joint diagonalization; Regularized ℓ1-norm optimization

1. INTRODUCTION

Brain Computer Interface (BCI) is a one-way information flow that translates recorded brain signals to command signals [1,2]. This system mainly focuses on the way of helping paralyzed people to move without taking care of their disorders. The BCI does not use the normal output pathways of peripheral nerves and muscles [3], it only records the brain signals to prepare commands for control of external devices [4,5]. In motorized prostheses, movement paralysis is the main consideration. BCI provides a substitute form of communication for people with motorized impairments and helps them to send commands by using their brain activities.

Fortunately, disabled people can generate different mental states, i.e. they can do motor imagery (MI) [6]. The movement of different parts of the body triggers different parts of the brain. In fact, subjects' ability to modulate brain activities enables BCI to detect and fulfil subjects' intentions [7], allowing paralyzed people to control a robotic arm [4]. In MI, the user is asked to make a special motion, such as the left hand. These imaginations activate some parts of the motor cortex. According to the activated area, BCI determines the imagined movement [8].

The Electroencephalography (EEG) signal often appears in response to external stimulation in the form of an electrical potential difference created by the neurons [8]. EEG is a non-invasive way to detect modulated brain signals [7,9], and suitable for real-time applications. The sensorimotor rhythm appears as a power change in a specific frequency range in the sensorimotor area at the moment of motion imaging. For this reason, EEG signal changes in conjunction with different movement imaginations can be used in BCI.

EEG signal is very weak because it is recorded from the scalp; besides, noise and artifacts, such as blinking or scalp movement, affect it. Therefore, feature extraction from the EEG signal and accurate classification of different MIs are difficult [10]. Recorded EEG signal is a mixture of many independent source signals including neural oscillations, event potential, spectral perturbations, and artifacts from eye movements, muscle activities, drifts, and the electrodes. Separating source signals could help to solve the problem. Independent Component Analysis (ICA) is a well-known method for identifying independent source signals from a recorded mixture signal [11].

Common Spatial Pattern (CSP) is a spatial filter for oscillatory EEG components [12]. It is the most commonly used technique for feature extraction and very effective in MI classification [10]. This technique seeks to find spatial filters that maximize variance for one class and minimize variance for another class [13]. CSP usually uses all available channels (brain areas) to estimate the covariance matrix. In a specific MI task, each person has a different activated channel, and choosing the proper channel is a major challenge [14]. Actually, the recorded EEG signal has a wide frequency-range, and the CSP processes the whole range at once. However, for each person the activated channel is in a special frequency range. One solution for solving this problem is dividing the frequency band into some sub-bands and then deploying the CSP

on each band and finally selecting the appropriate subband [15]. However, this solution is not promising, and some better improvements have been applied in order to remove its challenges. Among them, adjusting CSP weights based on Tikhonov regularization [16], or based on an l_1-norm regularization [17] have shown better performance. The idea of adjusting weights related to some phenomena, which is discussed in details in the following, and actually the aim of this paper is to present a better weight adjustment technique.

The CSP needs the covariance matrix of data. It is better to repeat a specific trial several times to increase accuracy. Therefore, one covariance matrix would be obtained in each repetition. EEG signal corresponding to the same task is a stationary process. Therefore, the total covariance matrix is the average of all the covariance matrices of the experiments. In EEG, averaging with equal weights over all trials does not seem to be a good idea, because different trials are contaminated by different amounts of user in-concentration, eye blink, or muscle movement artifacts.

The distraction and fatigue of subjects in the long data collection process often produce mislabelled trials. Locating the artifact data segments within a single trial is a topic of some recent works. Two spatial filters namely, CSP and ICA are widely used in MI BCI. CSP adopts a supervised algorithm that needs plenty of labelled data to find a projection matrix for maximizing the differences between the variances of two-class EEG data. Besides, the selected data are suggested to have strong de-synchronization/synchronization. So, the CSP method requires high-quality training data with accurate labels [18]. The main disadvantage of BCI is its sensitivity to the quality of the recorded signal in each trial because, in addition to the target signal, EEG is composed of many other undesired source signals. There are many investigations considered this issue. Recently [19] tried to solve this problem; it first called trials contaminated by noise and artifact as the low-quality trials, and presented a solution for rejecting the low-quality trials. All the processing of this work was done on the recorded mixture signal from the scalp, not on the target source signal. So, there appears another problem; what would happen if the noise and artifact sources would be dominant in all the trials. With the previous method, in this situation, all the trials would be considered as low quality and rejected. The solution to this problem is developing all the processing on the target source signal. Accordingly, this paper concerns this issue and extends the previous work. Actually in this paper selection is done on the source signals which are separated by a new fast ICA algorithm. To the best of the authors' knowledge, this is the first time that a fast approximate joint diagonalization is used in ICA-based BCI applications. In comparison, the obtained improvements in the experimental results over the previous work confirm the sophistication of the above arguments.

2. METHOD

2.1 Description of BCI IV 1 and BCI IV 2a Datasets

The dataset 2a from BCI competition IV is an EEG data recorded from 9 subjects for 4 different MI tasks, namely the imagination of movement of the left hand, right hand, both feet, and tongue. The data record was done for each subject at two 288 trial sessions. Subjects sat in a comfortable armchair looking at a monitor. At the beginning ($t = 0s$), a fixation cross appeared on the monitor with an acoustic warning. After two seconds ($t = 2s$), an arrow pointing either to the left, right, down, or up appeared. At this time subjects performed the MI task. They should carry out the MI task until the disappearance of the fixation cross. Twenty-two electrodes were used to record EEG signals. All signals were recorded monopolarly with reference to left mastoid and ground of right mastoid. The signals were sampled with 250 *Hz* and bandpass-filtered between 0.5 *Hz* and 100 *Hz*.

The dataset 1 from BCI competition IV is an EEG data recorded from 59 channels of 4 subjects for 4 different MI tasks, namely the imagination of movement of the left hand, right hand, and feet. The recording experiment of this dataset was performed in two sessions named calibration and evaluation. In the calibration session, each subject chose two MI tasks from three MI tasks. In this session, each task was started by showing a visual cue, and the task last for 4 *s*. Totally, each subject did 200 tasks balanced between two MI tasks. After each task, there was a 4 *s* break. In the evaluation session, each task was started by a soft voice command, and the task last between 1.5 − 8 *s*. After each task, there was a break between 1.5- 8*s*. For computational efficiency, the 100*Hz* version of the dataset is used. Figure 1 shows the EEG waveforms of the 1st channel of 1st, 50^{th}, and 100^{th} trial of subjects a, b, c, d, and e, for BCI competition IV 1 dataset.

2.2 Problem Statement and Ideology of the Proposed Method

In this study, a novel strategy is proposed to solve the mentioned problems. The rationale is to recognize low-quality EEG trials. Rejecting low-quality trials in an EEG-based MI BCI was previously investigated in [19]. However, in situations that all trials are contaminated

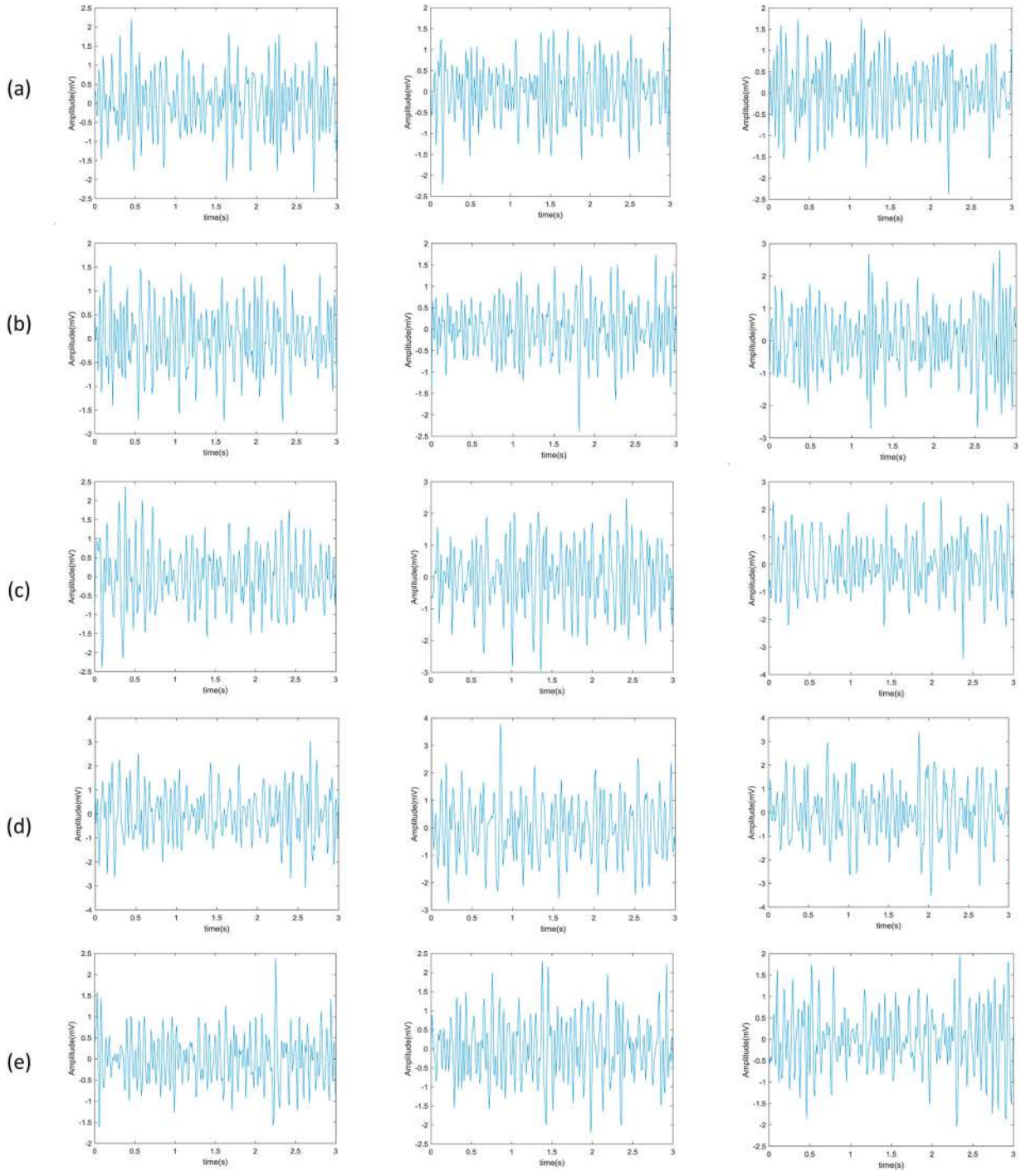

Figure 1: EEG waveform of the 1st channel of trial 1(left column), trial 50 (middle column), and trial 100 (right column), of a. subject a, b. subject b, c. subject c, d. subject d, e. subject e, for BCI competition IV 1 dataset

by noise or artifact, it is difficult to distinguish trials based on the recorded EEG signal. In the following, this paper tries to discuss this issue. As explained earlier, the recorded signals in EEG-based MI BCI are mixtures of some different source signals. Therefore, a solution for the mentioned problem is to use the original source signals, which could be done by separating the mixed signals properly using many algorithms such as ICA. Here appears the second issue that this paper wants to discuss. Assuming the source signal to be $\boldsymbol{S}$, the recorded signal would be $\boldsymbol{X} = \boldsymbol{AS}$, where $\boldsymbol{A}$ is mixture matrix. The ICA multiplies $\boldsymbol{A}^{-1}$ by $\boldsymbol{X}$ to find $\boldsymbol{S}$. But matrix inversion of high dimensional matrices has heavy processing and in low-cost applications comprises the main cost.

In this paper, both rejecting the low-quality trials by separating independent components of the recorded signal and reducing the high computation of matrix inversion in ICA are addressed. Generally speaking, the diagonalization of a matrix is the best way to reduce the complexity of its inversion. Because the inverse of a diagonal matrix is just the inverse of its diagonal elements. However, the diagonalization of a matrix is not so easy. Therefore, this paper presents an approximate joint diagonalization algorithm [20]. Generally speaking, the diagonalization of symmetric matrixes is easier than non-symmetric matrixes. Therefore, instead of $\boldsymbol{X}$, its covariance matrix is used for diagonalization. So, the proposed method would be faster than common ICA algorithms, because, in addition to using diagonalization before matrix inversion, the proposed fast diagonalization algorithm (see Appendix A) reduces the complexity of diagonalization.

The block diagram of the proposed method is shown in Figure 2. The first block records EEG signals from the brain for each trial. The recorded signals are mixtures of different source signals. The second block transforms them back to the original source signals by using a fast ICA algorithm. At the third block, a fast diagonalization algorithm measures criteria for within trial qualities. According to these criteria, this block devotes sparse weights to each trial and calculates a weighted averaged covariance matrix of the recorded signal. At the fourth block, the obtained covariance matrix is fed to a CSP filter. The output of this block is used as a feature for classification by Support Vector Machine (SVM) in the last block. In this section, the above processing is explained in detail.

2.3 Mathematical Analysis of the Proposed Method

Consider $\boldsymbol{X}^k = \boldsymbol{A}\boldsymbol{S}^k$ to be the recorded signal at trial k; $k = 1, \ldots, K$, where $\boldsymbol{X}^k = [\boldsymbol{x}_1^k, \boldsymbol{x}_2^k, \ldots, \boldsymbol{x}_N^k]$; $x_i^k \in R^M$. The goal is to estimate both $\boldsymbol{A}$ and $\boldsymbol{S}$ from $\boldsymbol{X}$. Let the within-trial covariance matrix be $\boldsymbol{C}^k$. Theoretically diagonalization process of this matrix would be $\boldsymbol{C}^k = E(\boldsymbol{X}^k \boldsymbol{X}^{k^T}) = \boldsymbol{A}E(\boldsymbol{S}^k \boldsymbol{S}^{k^T})\boldsymbol{A}^T = \boldsymbol{A}\boldsymbol{Q}^k\boldsymbol{A}^T$. Because the source signals are independent, and the cross-correlation terms that form the off-diagonal part of $\boldsymbol{Q}^k$ are zero. When more than two matrices are to be diagonalized, exact diagonalization may be possible if the matrixes possess a certain common structure. Otherwise only approximate joint diagonalization could be used. An efficient algorithm for approximate joint diagonalization is Fast Frobenius Diagonalization [19], which is based on the second-order approximation of a cost function for the simultaneous diagonalization problem. The Fast Frobenius Diagonalization algorithm tries to find matrix $\boldsymbol{V}$ (where $\boldsymbol{V} = \boldsymbol{A}^{-1}$) that diagonalizes the given covariance matrix in the following form:

$$\boldsymbol{Q}^k = \boldsymbol{V}\boldsymbol{C}^k\boldsymbol{V}^T, \tag{1}$$

Fast Frobenius Diagonalization Algorithm iteratively finds an approximate solution for the following optimization problem:

$$\min_{\boldsymbol{V}} \sum_{k=1}^{K} \sum_{i \neq j} ((\boldsymbol{V}\boldsymbol{C}^k\boldsymbol{V}^T)_{ij})^2 \tag{2}$$

In each iteration, matrix $\boldsymbol{V}$ is updated in the following form:

$$\boldsymbol{V}_{n+1} \leftarrow (I + \boldsymbol{W}_n)\boldsymbol{V}_n \tag{3}$$

where $\boldsymbol{I}$ denotes the identity matrix, $\boldsymbol{W}_{\mathrm{n}}$ is the update matrix, constrained to have zeroes on the main diagonal, and n is the iteration number.

Implementing Algorithm 1 derives the diagonalized covariance matrix ($\boldsymbol{Q} = \{\boldsymbol{Q}^1, \ldots, \boldsymbol{Q}^K\}$). Therefore, the covariance matrix of source signals could be obtained by multiplying $\boldsymbol{Q}^{-1}$ by $\boldsymbol{X}$. Algorithm 2 summarizes this step (see Appendix B) [20].

In the next step, the total covariance matrix is estimated by a weighted averaging over within trial covariance

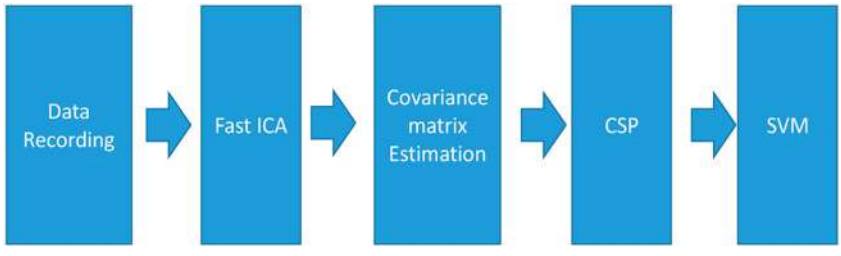

Figure 2: Block diagram of the proposed method

matrices of the separated source signals. The assigned weights conserve the following constraints:

$$C_1 : \boldsymbol{w}^T 1 = 1; C_2 : \boldsymbol{w}^T \boldsymbol{e}_k \geq 0, \tag{4}$$

where $\boldsymbol{w}$ is the weight vector. Actually, low-quality trials should be rejected by assigning weights obtained by solving the following l_1-norm optimization problem [19]:

$$\begin{aligned} &\min_{\boldsymbol{w}} \frac{1}{tr(\boldsymbol{D})} \boldsymbol{D}\boldsymbol{w}_1 \\ &S.t.\ \boldsymbol{w}^T 1 = 1; \boldsymbol{w}^T \boldsymbol{e}_k \geq 0, \end{aligned} \tag{5}$$

where $\boldsymbol{D} = diag[p_1, p_2, \ldots, p_K]$, and $p_k$is a scalar related to the quality of each trial. The underlying assumption behind the diagonalization is that the residue resulting from the diagonalization with respect to a low-quality trial is large. Considering $\boldsymbol{E}^k$ as the diagonalization residue, p_k can take the form of $p_k = \boldsymbol{E}^k{}_F$, where $\boldsymbol{E}^k$ could be derived by Algorithm 1.

The ideal covariance matrix could be the equal weight average of within-trial covariance matrices. Considering this assumption, l_1-norm optimization changes to the following regularized l_1-norm optimization [19]:

$$\begin{aligned} &\min_{\boldsymbol{w}} \frac{\alpha}{tr(\boldsymbol{D})} \boldsymbol{D}\boldsymbol{w}_1 + \frac{1}{2} \frac{1}{\sum_k \boldsymbol{Q}^k{}_F^2} \sum_k \left(\frac{1}{K} - w_k\right) \boldsymbol{Q}^k{}_F^2 \\ &S.t.\ \boldsymbol{w}^T 1 = 1; \boldsymbol{w}^T \boldsymbol{e}_k \geq 0, \end{aligned} \tag{6}$$

where α is the regularization parameter. After some mathematical manipulation (6) becomes as

$$\begin{aligned} &\min_{\boldsymbol{w}} \frac{\alpha}{tr(\boldsymbol{D})} \boldsymbol{D}\boldsymbol{w}_1 + \frac{1}{2} \frac{1}{tr(\boldsymbol{G})} \left(\boldsymbol{w} - \frac{1}{K}\right)^T \boldsymbol{G} \left(\boldsymbol{w} - \frac{1}{K}\right) \\ &S.t.\ \boldsymbol{w}^T 1 = 1; \boldsymbol{w}^T \boldsymbol{e}_k \geq 0, \end{aligned} \tag{7}$$

Where

$$\boldsymbol{G} = \begin{bmatrix} tr[\boldsymbol{Q}^1 \boldsymbol{Q}^{1\boldsymbol{T}}] & \cdots & tr[\boldsymbol{Q}^1 \boldsymbol{Q}^{K\boldsymbol{T}}] \\ \vdots & \ddots & \vdots \\ tr[\boldsymbol{Q}^K \boldsymbol{Q}^{1\boldsymbol{T}}] & \cdots & tr[\boldsymbol{Q}^K \boldsymbol{Q}^{K\boldsymbol{T}}] \end{bmatrix}.$$

The projected gradient method is known as a solution for convex optimization problems, which is extensively investigated over the last decades [21]. Alternating Direction Method for Multipliers (ADMM) algorithm is used to blend the decomposability of dual ascent with the superior convergence properties of the method of multipliers. The ADMM solves the following optimization problem [22]:

$$\begin{aligned} &\min_{\boldsymbol{w},\boldsymbol{v}} f(\boldsymbol{w}) + g(\boldsymbol{v}) \\ &s.t.\ \boldsymbol{A}\boldsymbol{w} + \boldsymbol{B}\boldsymbol{v} = \boldsymbol{c}, \end{aligned} \tag{8}$$

where $\boldsymbol{w} \in \boldsymbol{R}^n$ and $\boldsymbol{v} \in \boldsymbol{R}^m$, $\boldsymbol{A} \in \boldsymbol{R}^{p \times n}$, $\boldsymbol{B} \in \boldsymbol{R}^{p \times m}$, $\mathbf{c} \in \mathbf{R}^{\mathrm{p}}$, and f and g are convex functions. The Lagrangian form of the above problem is as follows:

$$\begin{aligned} L_\rho(\boldsymbol{w}, \boldsymbol{v}, \boldsymbol{y}) = f(\boldsymbol{w}) + g(\boldsymbol{v}) + \boldsymbol{y}^T(\boldsymbol{A}\boldsymbol{w} + \boldsymbol{B}\boldsymbol{v} - \boldsymbol{c}) \\ + \left(\frac{\rho}{2}\right) \boldsymbol{A}\boldsymbol{w} + \boldsymbol{B}\boldsymbol{v} - \boldsymbol{c}_2^2, \end{aligned} \tag{9}$$

where $\rho > 0$. The ADMM can solve the above problem by the following iterations:

$$\boldsymbol{w}^{n+1} = \min_{\boldsymbol{w}} L_\rho(\boldsymbol{w}, \boldsymbol{v}^n, \boldsymbol{y}^n) \tag{10}$$

$$\boldsymbol{v}^{n+1} = \min_{\boldsymbol{v}} L_\rho(\boldsymbol{w}^{n+1}, \boldsymbol{y}^n, \boldsymbol{v}) \tag{11}$$

$$\boldsymbol{y}^{n+1} = \boldsymbol{y}^n + \rho(\boldsymbol{A}\boldsymbol{w}^{n+1} + \boldsymbol{B}\boldsymbol{v}^{n+1} - \boldsymbol{c}) \tag{12}$$

Therefore, in order to solve (7) using ADMM, first should write its Lagrangian equivalent. The Lagrangian equivalent of (7) is in the following form:

$$\begin{aligned} \frac{\alpha}{tr(\boldsymbol{D})} \boldsymbol{D}\boldsymbol{w}_1 + \frac{1}{2} \frac{1}{tr(\boldsymbol{G})} \left(\boldsymbol{w} - \frac{1}{K}\right)^T \boldsymbol{G} \left(\boldsymbol{w} - \frac{1}{K}\right) \\ + l_{C_1}(\boldsymbol{w}) + l_{C_2}(\boldsymbol{w}), \end{aligned} \tag{13}$$

where l_C is the indicator function. Furthermore, comparing (7) with (8), the following equivalencies would exit:

$$\begin{aligned} f(\boldsymbol{w}) = \frac{\alpha}{tr(\boldsymbol{D})} \boldsymbol{D}\boldsymbol{w}_1 \\ + \frac{1}{2} \frac{1}{tr(\boldsymbol{G})} \left(\boldsymbol{w} - \frac{1}{K}\right)^T \boldsymbol{G} \left(\boldsymbol{w} - \frac{1}{K}\right) + l_{C_1}(\boldsymbol{w}) \end{aligned} \tag{14}$$

$$g(\boldsymbol{v}) = l_{C_2}(\boldsymbol{v}). \tag{15}$$

Considering (10), and (13) obtains:

$$\boldsymbol{w}^{n+1} = \min_{\boldsymbol{w}} L_\rho(\boldsymbol{w}, v^n, y^n) \tag{16}$$

Differentiating the objective function of (16) with respect to $\boldsymbol{w}$ and finding its root obtains:

$$\begin{aligned} \boldsymbol{w}^{n+1} = \left(\rho \frac{\boldsymbol{G}}{tr(\boldsymbol{G})} + \boldsymbol{I}\right)^{-1} \\ \times \left[z^n - \boldsymbol{y}^n + \rho\left(\frac{\boldsymbol{G}1}{Ktr(\boldsymbol{G})} - \frac{\alpha}{tr(\boldsymbol{D})} \boldsymbol{D}1 - \xi 1\right)\right]. \end{aligned} \tag{17}$$

Substituting (17) to constraint C_1 in (4), and after some mathematical manipulations obtains:

$$\xi = \left(\rho 1^T\left(\frac{\rho \boldsymbol{G}}{tr(\boldsymbol{G})} + \boldsymbol{I}\right)^{-1} 1\right)^{-1}\left[1^T\left(\frac{\rho \boldsymbol{G}}{tr(\boldsymbol{G})} + \boldsymbol{I}\right)^{-1} \times \left[z^n - \boldsymbol{y}^n + \rho\left(\frac{\boldsymbol{G}1}{Ktr(\boldsymbol{G})} - \frac{\alpha}{tr(\boldsymbol{D})}\boldsymbol{D}1\right)\right] - 1\right]. \tag{18}$$

Considering (11), and (13) obtains:

$$\begin{aligned}\boldsymbol{v}^{n+1} &= \min_{\boldsymbol{v}} l_{C_1}(\boldsymbol{v}) + \frac{1}{2\rho}\boldsymbol{v} - (\boldsymbol{w}^{n+1} + \boldsymbol{y}^n)_2^2 . \\ &= \min_{\boldsymbol{w}} \frac{\alpha}{tr(\boldsymbol{D})}\boldsymbol{D}\boldsymbol{w}_1 \\ &\quad + \frac{1}{2}\frac{1}{tr(\boldsymbol{G})}\left(\boldsymbol{w} - \frac{1}{K}\right)^T \boldsymbol{G}\left(\boldsymbol{w} - \frac{1}{K}\right) \\ &\quad + \frac{1}{2\rho}v^n - \boldsymbol{w} - y^n{}_2^2 + \xi(\boldsymbol{w}^T 1 - 1).\end{aligned} \tag{19}$$

According to the definition of the indicator function, $\boldsymbol{v}$ should be in the domain of C_2; i.e. $\boldsymbol{v} \geq 0$. Accordingly,$\boldsymbol{w}^{n+1} + \boldsymbol{y}^n$ should be positive to minimize the objective function. In fact $\mathbf{w}^{n+1} + \mathbf{y}^n$ should be projected to the domain of C_2, i.e.

$$\boldsymbol{v}^{n+1} = \boldsymbol{P}_{C_2}(\boldsymbol{w}^{n+1} + \boldsymbol{y}^n). \tag{20}$$

Considering (12), can write

$$\boldsymbol{y}^{n+1} = \boldsymbol{y}^n + \boldsymbol{w}^{n+1} - \boldsymbol{v}^{n+1}. \tag{21}$$

Therefore, the target sparse weights can be obtained by choosing initializing points of $\mathbf{w}^0, \mathbf{v}^0, \mathbf{y}^0$ and then repeating (17), (20), (21).

3. EXPERIMENTAL RESULTS

This section presents the results of deploying the proposed methods on two available well-known datasets, namely BCI competition IV 1 and 2a. The conducted experiment of BCI competition IV is the recording of EEG signals during different MI tasks. The purpose is to classify the recorded EEG signals of different MI tasks. According to section 2, the following CSP based methods are used for feature extraction:

(1) Proposed ICA-CSP with equal weights ($w_k = 1/K$). The idea is to give equal weights to trials.
(2) Proposed ICA-CSP with quality related weights ($w_k = \eta E_F^{k^{-1}}$, where η is a normalization factor). The idea is to give a small weight to a large residue (low quality) trial.
(3) Proposed ICA-CSP with sparse weights. The idea is to give sparse weights to trials.

Furthermore, in order to clarify the improvement of the proposed methods, they are compared with some well-known related methods presented in [16,17,19,23–25]. The following CSP based methods are presented in these references for feature extraction:

(1) CSP with the sparse weights [19]. The idea is to give sparse weights to trials, this method was presented in [19].
(2) CSP with equal weights (CSP) [23]. The idea is to use the simple form of CSP, this method was presented in [23].
(3) CSP with weights adjusted based on Tikhonov regularization (TRCSP) [16]. The idea is to use Tikhonov regularization for adjusting weights, this method was presented in [16].
(4) CSP with weights adjusted based on an l_1-norm regularization (WLCSP) [17]. The idea is to use l_1-norm optimization for adjusting weights, this method was presented in [17].
(5) CSP with filter banks (FBCSP) [24]. The idea is to use a filter bank with equal frequency bandwidths to divide the frequency range of $0-40$ *Hz* into 8 equal parts to compose 9 different non-overlapping frequency bands with 4 *Hz* bandwidth. This method was presented in [24].
(6) CSP with filter banks (FBCSP) [25]. The idea is to use a filter bank with different frequency bandwidths to divide the frequency range of $8-32$ *Hz* into 8 parts with centre frequencies of 8, 9.57, 10.09, 11.89, 15.75, 21.71, 22, 27.27 *Hz*. This method was presented in [25].

In order to have a fair comparison, the log-variance of the output of these CSP based methods is used as the feature vector. The features are classified by SVM Toolbox of MATLAB software (version 2015a). The Radial Basis Function kernel with unit kernel parameter is used for classification. From the SVM Toolbox, the svmtrain(.) function is used for derivation the model of the training data. The inputs of this function are labels and extracted features of the training data. Then the svmclassify(.) function of SVM Toolbox is used for predicting the labels of the testing data. Inputs of this function are the obtained model and the extracted features of testing data. However, new versions of MATLAB software, the fitcsvm(.) and predict(.) functions are substituted for svmtrain(.) and svmclassify(.) functions, respectively. The inputs of these two functions are the same as mentioned the previous functions.

Table 1: Classification accuracy for different values of α for each subject in datasets BCI IV 1, and BCI IV 2a

	α							
Subject	100	50	10	5	1	0.2	0.02	0
Dataset IV 2a								
1	0.7295±0.022	0.707±0.020	0.707±0.025	0.707±0.028	0.707±0.018	0.707±0.027	0.707±0.022	0.707±0.024
2	0.7003±0.063	0.628±0.067	0.628±0.067	0.627±0.069	0.657±0.059	0.652±0.061	0.672±0.070	0.728±0.061
3	0.6519±0.044	0.639±0.049	0.643±0.049	0.632±0.048	0.620±0.041	0.626±0.041	0.643±0.041	0.661±0.046
4	0.6422±0.039	0.593±0.035	0.592±0.031	0.587±0.049	0.632±0.035	0.644±0.035	0.654±0.042	0.632±0.036
5	0.5880±0.066	0.521±0.069	0.513±0.061	0.540±0.071	0.589±0.061	0.581±0.062	0.603±0.069	0.596±0.063
6	0.6736±0.039	0.615±0.037	0.615±0.046	0.621±0.041	0.651±0.034	0.669±0.034	0.660±0.043	0.665±0.043
7	0.6928±0.055	0.666±0.059	0.671±0.055	0.678±0.051	0.704±0.051	0.700±0.051	0.724±0.059	0.704±0.059
8	0.6265±0.054	0.591±0.059	0.590±0.054	0.599±0.059	0.631±0.051	0.631±0.059	0.627±0.059	0.650±0.051
9	0.6541±0.039	0.653±0.033	0.652±0.046	0.647±0.030	0.646±0.032	0.647±0.037	0.687±0.033	0.703±0.046
Ave	0.6621 ± 0.047	0.624±0.046	0.623 ± 0.048	0.627 ± 0.049	0.649±0.042	0.651±0.045	0.664±0.048	0.672±0.046
Dataset IV 1								
a	0.839 ± 0.124	0.834±0.121	0.849±0.119	0.778±0.118	0.688±0.108	0.698±0.115	0.702±0.125	0.682±0.109
b	0.741 ± 0.064	0.741±0.067	0.726±0.063	0.762±0.069	0.761±0.075	0.772±0.070	0.617±0.062	0.576 ±0.075
c	0.736 ± 0.091	0.736±0.093	0.741±0.090	0.716±0.088	0.576±0.099	0.525±0.096	0.496±0.097	0.597±0.088
d	0.763 ± 0.103	0.763±0.098	0.768±0.105	0.733±0.115	0.755±0.099	0.691±0.098	0.495±0.101	0.514±0. 106
e	0.949 ± 0.049	0.949±0.054	0.949±0.046	0.945±0.061	0.955±0.051	0.96±0.054	0.909±0.059	0.874±0.053
f	0.865 ± 0.107	0.865±0.099	0.845±0.101	0.740±0.095	0.845±0.101	0.860±0.102	0.833±0.097	0.792±0.102
g	0.908 ± 0.065	0.913±0.072	0.933±0.071	0.853±0.069	0.737±0.073	0.702±0.061	0.54±0.059	0.535±0.067
Ave	0.829±0.086	0.829±0.086	0.830±0.085	0.79 ± 0.087	0.759 ± 0.084	0.744±0.085	0.656±0.085	0.653±0.085

Table 1 shows the classification accuracy for different values of α (sparsity parameter) for each subject in datasets BCI IV 1 and IV 2a. 10-fold cross validation is done, accordingly, results are reported as mean ± standard deviation. α is the sparsity parameter, increasing α promotes sparsity. As seen, the α value related to the best accuracy of each dataset is different, actually, the dataset that was recorded better (with fewer low quality trials) has smaller α. In dataset IV 1, the accuracy decreases with increasing sparsity, which means that there are few low quality trials in this dataset. However, in the dataset IV 2a, increasing sparsity results in better performance, which means that this dataset has more low quality trials. In addition, it can be seen that different subjects in a dataset have the same α value related to their best accuracy, and a clear relationship could be seen between α and accuracy. Accordingly, $\alpha = 0.2$ and $\alpha = 10.2$ would be used respectively for classification accuracies in datasets BCI IV 1 and IV 2a, in Tables 2 and 3.

Table 2 compares the proposed methods with each other and with [19], because it is the main reference of this paper, and the improvements compared with this reference should be clarified and discussed deeply. Results of the classification accuracy, for 10-fold cross validation and $\gamma = 10^{-5}$, are reported in Table 2 as mean ± standard deviation. The proposed ICA-CSP with sparse weight has the highest classification accuracy. The proposed ICA-CSP with quality related weight shows a little better classification accuracy compared with proposed ICA-CSP with equal weight. This shows the weight coefficients obtained by residue matrices are designed in a sophisticated manner. However, as sparse weight performs better than non-sparse weight (determined by error matrices obtained from joint diagonalization), so, in a future work, the error matrices should be used for designing weights in a more sophisticated way. Compared with [19] (CSP with the sparse weight), the proposed ICA-CSP with sparse weight shows better

Table 2: Classification accuracy of the proposed methods and the main reference of this paper ([19]) for each subject in datasets BCI IV 1, and BCI IV 2a

	Method			
Subject	CSP with the sparse weight [19]	Proposed ICA-CSP with sparse weight	Proposed ICA-CSP with quality related weight	Proposed ICA-CSP with equal weight
Dataset IV 2a				
1	0.7339±0.018	0.7295±0.022	0.7035±0.025	0.7050±0.026
2	0.6491±0.044	0.7003±0.063	0.6377±0.052	0.616 ±0.043
3	0.6272±0.034	0.6519±0.044	0.6294±0.024	0.6044 ±0.027
4	0.6461±0.047	0.6422±0.039	0.5874±0.048	0.5786 ±0.042
5	0.5803±0.116	0.5880±0.066	0.5076±0.059	0.5205 ±0.048
6	0.6706±0.054	0.6736±0.039	0.5679±0.051	0.5795±0.056
7	0.6992±0.0493	0.6928±0.055	0.6657±0.057	0.6661±0.058
8	0.6283±0.063	0.6265±0.054	0.6019±0.070	0.5950±0.068
9	0.6507±0.047	0.6541±0.039	0.6472±0.043	0.6503±0.038
Ave	0.6539 ± 0.052	0.6621 ± 0.047	0.6165 ± 0.048	0.6129 ± 0.045
Dataset IV 1				
a	0.8544±0.104	0.8544±0.112	0.7877±0.093	0.8344±0.102
b	0.7466±0.079	0.7527±0.064	0.7377±0.068	0.7327±0.063
c	0.7316±0.090	0.7816±0.091	0.7461±0.104	0.7616±0.098
d	0.7683±0.142	0.8072±0.103	0.7872±0.103	0.7872±0.101
e	0.9494±0.061	0.9644±0.049	0.9544±0.037	0.9444±0.040
f	0.890 ± 0.066	0.910±0.107	0.930 ± 0.075	0.890±0.081
g	0.9188±0.0645	0.9183±0.079	0.9133±0.072	0.9183±0.076
Ave	0.8370 ± 0.087	0.8555 ± 0.087	0.8366 ± 0.079	0.8384 ± 0.080

Table 3: Classification accuracy of the proposed ICA-CSP with sparse weight method and the well-known state of the art methods for each subject in datasets BCI IV 1

	Method						
Subject	CSP with the sparse weight [19]	CSP [23]	TRCSP [16]	WLCSP [17]	FBCSP [24]	FBCSP [25]	Proposed ICA-CSP with sparse weight
a	0.8544±0.104	0.61611±0.177	0.834 ± 0.061	0.74778±0.085	0.743 ± 0.107	0.61722 ± 0.180	0.8544±0.112
b	0.7466±0.079	0.79778±0.100	0.787 ± 0.108	0.79778±0.089	0.758 ± 0.121	0.86389 ± 0.052	0.7527±0.064
c	0.7316±0.090	0.76833±0.084	0.773 ± 0.084	0.77833±0.079	0.751 ± 0.127	0.85222 ± 0.073	0.7816±0.091
d	0.7683±0.142	0.77944±0.157	0.794 ± 0.133	0.84333±0.095	0.893 ± 0.072	0.89944 ± 0.084	0.8072±0.103
e	0.9494±0.061	0.96444±0.053	0.964 ± 0.047	0.97444±0.051	0.938 ± 0.033	0.97 ± 0.034	0.9644±0.049
f	0.89±0.066	0.58667±0.153	0.798 ± 0.136	0.89±0.083	0.696 ± 0.082	0.88389 ± 0.057	0.91±0.107
g	0.9188±0.0645	0.738±0.091	0.913 ± 0.068	0.908±0.096	0.767 ± 0.106	0.792 ± 0.129	0.9183±0.079
Ave	0.8370 ± 0.087	0.750 ± 0.116	0.838±0.091	0.8486 ± 0.082	0.792±0.093	0.839±0.087	0.8555 ± 0.087

classification accuracy in both datasets, and this clarifies the superiority of the proposed method.

Considering the classification accuracy for each subject, again the proposed ICS-CSP with sparse weight performs better and proves that there exist low quality trials in each dataset. So, the assumed ideology in section 2.1 is correct. For example, subject 2 shows more than 7% accuracy improvement using the proposed ICS-CSP with sparse weight comparing with proposed ICA-CSP with equal weight.

It has worth to mention that all the quality of a trial is somehow dependent on the subject that is performing the MI tasks. For example, subject 9 shows the same accuracy at each of the proposed methods, and no trial was rejected. This indicates that this subject did the MI tasks well, and there is almost no low quality trial in recording its dataset.

Comparing the results over datasets IV 1 and IV 2a, by using the proposed ICS-CSP with sparse weight, dataset IV 2a shows more classification accuracy improvement, which indicates that dataset IV 2a more low-quality trials than IV a. For example, in dataset BCI IV 1, subject g has no classification accuracy improvement, and the maximum classification accuracy improvement in this dataset is 2%. But in dataset BCI IV 2a, subject 9 has 9% classification accuracy improvement.

Table 3 presents a comparison between the proposed ICA-CSP with sparse weight with some of the well-known related methods presented in [16,17,23–25]. The explanation about the details of these methods is out of the purpose of this paper, and the interested reader may refer to [16,17,23–25]. The classification accuracy is deployed on dataset BCI IV 1. As seen, the proposed ICA-CSP with sparse weight has the best performance among all the referred methods. It becomes more interesting while considering that these are well-known state of the art methods in BCI MI classification task. This proves the main idea of this paper that separating the mixed signals and then removing the low quality trials by providing sparse weights would really improve classification accuracy in different datasets and different users. As seen, after the proposed ICA-CSP with sparse weight, WLCSP has better performance and after that TRCSP and CSP with the sparse weight have better performance than other methods. Because, these methods also solve an optimization problem to find the optimum weights, and they find it. However, they do not separate different mixed source signals; that's why they could not achieve the performance of the proposed ICA-CSP with sparse weight method. As seen, the FBCSP methods have better performance than CSP method. Because, the effectiveness of CSP depends on the subject-specific frequency band. However, FBCSP tries to find more effective subject-specific frequency bands, and improves the CSP performance. It should be noted that [25] divided the frequency bands in a more sophisticated manner than [24] and could achieve a better performance.

4. CONCLUSION

The classification accuracy of MI BCI could be degraded by low-quality trials. So, rejecting such trials is a promising solution for improving classification accuracy. But processing should be done on the source signals, not on the recorded signals because in the worst case scenario where all the trials are low quality, processing the recorded signal results in rejecting all the trials. In this paper, a new fast approach was presented for rejecting low-quality trials. Actually, a new fast ICA algorithm was presented for separating the source signals, and then the rejection was developed on the source signal. Results indicated the improvement of classification accuracy of the proposed methods in comparison with some of the well-known related works.

REFERENCES

1. S. Soman, and B. K. Murthy, "Using brain computer interface for synthesized speech communication for the physically disabled," *Procedia. Comput. Sci.*, Vol. 46, pp. 292–8, 2015.

2. P. Gargava, K. Sindwani, and S. Soman. "Controlling an arduino robot using Brain Computer Interface," in *Proc. 3rd Int. Conf. Reliab. Infocom Technol. Opt.*, 2014, pp. 1–5.

3. D. J. McFarland, and J. R. Wolpaw, "Brain–computer interface use is a skill that user and system acquire together," *PLoS Biol.*, Vol. 16, no. 7, pp. e2006719, 2018.

4. D. M. Brandman, et al., "Rapid calibration of an intracortical brain–computer interface for people with tetraplegia," *J. Neural Eng.*, Vol. 15, no. 2, pp. 026007, 2018.

5. H. Pant, S. Soman, and M. Sharma. "Twin neural networks for efficient eeg signal classification," in *2018 Int. Joint Conf. Neural Netw*, pp. 1–7, 2018.

6. E. A. Mohamed, M. Z. Yusoff, A. S. Malik, M. R. Bahloul, D. M. Adam, and I. K. Adam, "Comparison of EEG signal decomposition methods in classification of motor-imagery BCI," *Multimed. Tools. Appl.*, Vol. 77, no. 16, pp. 21305–27, 2018.

7. M. C. Corsi, M. Chavez, D. Schwartz, L. Hugueville, A. N. Khambhati, D. S. Bassett, and F. de Vico Fallani, "Integrating EEG and MEG signals to improve motor imagery classification in brain–computer interface," *Int. J. Neural Syst.*, Vol. 29, no. 01, pp. 1850014, 2019.

8. S. Soman, "High performance EEG signal classification using classifiability and the Twin SVM," *Appl. Soft. Comput.*, Vol. 30, pp. 305–18, 2015.

9. J. Dinarès-Ferran, R. Ortner, C. Guger, and J. Solé-Casals, "A new method to generate artificial frames using the empirical mode decomposition for an EEG-based motor imagery BCI," *Front. Neurosci.*, Vol. 12, pp. 308, 2018.

10. Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Temporally constrained sparse group spatial patterns for motor imagery BCI," *IEEE Transactions on Cybernetics*, Vol. 49, no. 9, pp. 3322–32, 2018.

11. J. R. Wessel, "Testing multiple psychological processes for common neural mechanisms using EEG and independent component analysis," *Brain Topogr.*, Vol. 31, no. 1, pp. 90–100, 2018.

12. J. Kevric, and A. Subasi, "Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system," *Biomed. Signal. Process. Control.*, Vol. 31, pp. 398–406, 2017.

13. S. H. Park, D. Lee, and S. G. Lee, "Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, Vol. 26, no. 2, pp. 498–505, 2017.

14. Y. Park, and W. Chung. "BCI classification using locally generated CSP features," in *2018 6th Int. Conf. Brain-Comput. Interface*, pp. 1–4, 2018.

15. M. R. Islam, T. Tanaka, and M. K. I. Molla, "Multiband tangent space mapping and feature selection for classification of EEG during motor imagery," *J. Neural Eng.*, Vol. 15, no. 4, pp. 046021, 2018.

16. 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. 355–62, 2010.

17. H. Wang, Q. Tang, and W. Zheng, "L1-norm-based common spatial patterns," *IEEE Trans. Biomed. Eng.*, Vol. 59, no. 3, pp. 653–62, 2011.

18. B. Zhou, X. Wu, Z. Lv, L. Zhang, and X. Guo, "A fully automated trial selection method for optimization of motor imagery based brain-computer interface," *PloS one*, Vol. 11, no. 9, pp. e0162657, 2016.

19. N. Tomida, T. Tanaka, S. Ono, M. Yamagishi, and H. Higashi, "Active data selection for motor imagery EEG classification," *IEEE Trans. Biomed. Eng.*, Vol. 62, no. 2, pp. 458–67, 2014.

20. A. Ziehe, P. Laskov, G. Nolte, and K. R. MÃžller, "A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation," *J. Mach. Learn. Res.*, Vol. 5, pp. 777–800, 2004.

21. G. Torrisi, S. Grammatico, R. S. Smith, and M. Morari, "A projected gradient and constraint linearization method for nonlinear model predictive control," *SIAM J. Control Optim.*, Vol. 56, no. 3, pp. 1968–99, 2018.

22. S. Boyd, N. Parikh, and E. Chu. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Hanover: Now Publishers Inc.

23. Y. Wang, S. Gao, and X. Gao. "Common spatial pattern method for channel selelction in motor imagery based brain-computer interface," *2005 IEEE Eng. Med. Biol. 27th annu conf.*, 2006.

24. K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan. "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *2008 IEEE Int. Joint Conf. Neural Networ.*, pp. 2390–7, 2008.

25. H. Zhang, Z. Y. Chin, K. K. Ang, C. Guan, and C. Wang, "Optimum spatio-spectral filtering network for brain–computer interface," *IEEE Trans. Neural Netw.*, Vol. 22, no. 1, pp. 52–63, 2010.

APPENDICES

APPENDIX A: ALGORITHM 1, FAST APPROXIMATE JOINT DIAGONALIZATION

```
Input C = {C^1, ..., C^K}, ;
W_1 = 0, V_1 = I;
    while error > epsilon & n < 1000
% Compute W
For k = 1:K
    z(i,j) = z(i,j)+C(i,i,k)*C(j,j,k);
    y(i,j) = y(i,j) + 0.5*C(j,j,k)*(C(i,j,k)+conj(C(j,i,k)));
end
    W(i,j) = (z(j,i)*y(j,i)-z(i,i)*y(i,j))/(z(j,j)*z(i,i)-z(i,j)^2);
    W(j,i) = ((z(i,j)*y(i,j)-z(j,j)*y(j,i))/(z(j,j)*z(i,i)-z(i,j)^2));
    [f,e] = log2(norm(W,'inf'));
    s = max(0,e-1);
    W = W/(2^s);
% Compute update
    V = (I + W)*V;
    V = diag(1./sqrt(diag(V*V')))*V;
    C = A*C*V';
for k = 1:K
    f = f + trace((V*C(:,:,k)*V')'*  V*C(:,:,k)*V')  -
trace (V*C(:,:,k)*V'.*V*C(:,:,k)*V');
end
% convergence
    error = abs(f(n)-f(n-1));
      n = n+1;
end
```

APPENDIX B: ALGORITHM 2, ICA

```
Input X
% tau – array of time-delays, default: tau = [ 0 1];
X = X - mean(x);
C = zeros(N, N, 2);
for t = 1:2
    C0 = x(:,1:T-tau(t))*x(:,1+tau(t):T)' / (T-tau(t)-1);
    C(:,:,t) = (C0+C0')/2;
end
    Q = Algorithm 1(C)
```

Authors

M. A. Amirabadi was born in Zahedan, Iran, in 1993. He received the B.Sc. degree in Optics & Laser Engineering from Malek-e-Ashtar University of Technology, Isfahan, Iran, in 2015, and the M.Sc. degree in Communication Engineering from Iran University of Science and Technology, Tehran, Iran in 2017. Now he is studying Ph.D. in Communication Engineering in Iran University of Science and Technology, Tehran, Iran. His research interests include Multimode Fiber Optic Communication, Free Space Optical Communication, and Deep Learning.

Corresponding author. Email: m_amirabadi@elec.iust.ac.ir

M. H. Kahaei received the BSc degree in electrical engineering from Isfahan University of Technology, Isfahan, Iran, in 1986, the MSc degree in adaptive signal processing from the University of the Ryukyus, Okinawa, Japan, in 1994, and the PhD degree in signal processing from Queensland University of Technology, Brisbane, Australia, in 1998. Since 1999, he has been with the School of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran, where he is currently an Associate Professor and the Head of Signal and System Modeling laboratory. His research interests include array signal processing with primary emphasis on compressed sensing, sparse optimization problems, data science, localization, tracking, DOA estimation, blind source separation, and wireless sensor networks.