IJST, Transactions of Electrical Engineering, Vol. 36, No. E2, pp 147-161
Printed in The Islamic Republic of Iran, 2012

AN ITERATIVE SPATIO-SPECTRAL DISCRIMINANT SCHEME FOR EEG CLASSIFICATION*

D. FATTAHI[1**] AND R. BOOSTANI[2]

[1]Biomedical Eng. Group, Faculty of Electrical and Computer Engineering, Shiraz University, Shiraz, I. R. of Iran
Email: fattahi.d@gmail.com

[2]CSE & IT Dept, School of Electrical and Computer Engineering, Mollasadra Street, Shiraz, I. R. of Iran, Postal code: 71348-51154

Abstract– Brain Computer Interface (BCI) systems still suffer from lack of accuracy in real-time applications. This problem emerges from isolated optimization, and in some occasions from mismatching of feature extraction and classification stages. To unify optimization of both stages, this paper presents a novel scheme to integrate them and simultaneously optimize under a unit criterion. The proposed method iteratively estimates both spatio-spectral filters and classifier weights under a non-linear form of Fisher criterion. In order to validate the introduced method, two standard EEG sets, one containing 118 EEG signals and the other 29, were employed to demonstrate its spatial resolution capability. Experimental results on both datasets reveal the superiority of the proposed scheme in terms of enhancing the classification performance simultaneously with speeding up the optimization process, compared to the conventional methods.

Keywords– BCI, EEG feature extraction, spatio-spectral filtering, EEG classification

1. INTRODUCTION

Brain Computer Interface (BCI) systems are basically designed to translate imagery thoughts into meaningful commands in the form of curser movement or activate an electric device [1]. An optimistic vision to the future of this field is to enable an individual driving a car or triggering accessible hardware just by imagination. BCIs can provide an external communication channel for enabling the patients with neurological disorders, such as Amyotrophic Lateral Sclerosis (ALS) or severe Multiple Sclerosis (MS) to facilitate their life [2-3]. By developing signal processing techniques, especially in the field of Blind Source Separation (BSS), researchers have tried to deploy these schemes in a vast variety of BCI applications [4]. The main objective arising from this idea is that source signals are more informative, independent, and have better signal to noise ratio (SNR) than the scalp EEG signals. The foundation of BSS is established on the fact that the recorded signals (here, scalp EEG channels) are a linear combination of statistically uncorrelated sources that are spatially distributed inside the brain. In other words, the EEG sources are spatially and spectrally filtered while passing through different layers of head such as brain tissue, Cerebral-Spinal Fluid (CSF), skull and scalp. Thus, estimation of EEG sources can be carried out by solving an inverse problem, in which the recorded EEG channels should be applied to the inverse filters [2, 4]. Common Spatial Pattern (CSP) [5-6] is one of the pioneer methods used to estimate EEG sources by a set of spatial filters based on maximizing an energy criterion. To increase the elicited sources quality, some studies pass the recorded EEG signals through band-pass filters to remove the

*Received by the editors August 25, 2012; Accepted December 15, 2012.
**Corresponding author

redundant frequencies [6-7]. However, the discriminant frequency bands vary from one subject to another one. Therefore, it is necessary to optimize the subject-dependent spectral filters, as is done for the spatial filters. A well-known method toward satisfying the aforementioned objective is Common Spatio-Spectral Patterns (CSSPs) [8], which simultaneously embedded a first-order temporal filter into CSP to equip it with spectral features. Although first-order finite impulse response (FIR) filters are stable and have low complexity, they include several side-lobes that affect the quality of the selected frequency bands. Dornhege *et al.* [9] later proposed the common sparse spatio-spectral patterns (CSSSPs) algorithm to enhance the flexibility of FIR filters. Incidentally, to avoid the over-fitting problem, they incorporated a regularization term to the objective function for balancing the tradeoff between the sparsity and accuracy [9]. On the other hand, some studies focused on the content of frequency domain revealed by Discrete Fourier Transform (DFT) to some extent, where instead of the temporal filters, the frequency filters are estimated. In this way, Wu *et al.* suggested an iterative spatio-spectral pattern learning (ISSPL) algorithm [10] in which the spatial and spectral filters are optimized using two different objective functions. The main flaw of this approach is the lack of matching between the spectral and spatial filters estimation. In other words, due to the lack of unification through optimizing the stages, the resulted accuracy is not still convincing. Recently, Christoforou *et al.* [11] and Nasihatkon *et al.* [12] proposed two general frameworks, where the spatial and spectral filters were simultaneously optimized by a certain target function. Nevertheless, both methods suffer considerably from a heavy computational complexity.

In this research, an integrated scheme is proposed that iteratively estimates both spectral and spatial filters. This successive learning leads to the achievement of high classification accuracy along with a tolerable computational cost compared to the rival methods.

The rest of this paper is structured as follows: In Section 2, a purposive survey about the pros and cons of the spatial filtering approaches is presented. Then, the proposed iterative spatio-spectral scheme is introduced, and the other similar approaches are briefly discussed. The details of the implementation and the employed datasets are described in Section 3. Next, the experimental and comparative results are presented in Section 4 in which the methods are compared and discussed from different aspects. Finally, the paper is concluded in Section 5.

2. METHODS

a) Common Spatial Pattern

The basic idea of standard CSP (termed as Fukunaga and Koontz transform) was first introduced in [5]. This spatial scheme was then repeatedly employed to enhance the performance of BCI systems [7]. CSP method tries to find linear spatial filters in order to maximize the energy of filtered signals belonging to a certain class over summation of other classes. Assume that u is the spatial filter, $X^{(j)}$ and y_j represent j-th trial of multichannel EEGs and its corresponding class label, respectively. A single channel source can be found by spatially filtered scalp EEG signals, i.e. $u^T X$, and energy of the source can be written as $u^T(XX^T)u$. Hence, the average of this energy for each specific class c is equal to $u^T R_c u$, in which R_c is determined as follows:

$$R_c = \frac{1}{n_c} \sum_{\substack{j=1 \\ y_j=c}}^{n} X^{(j)} X^{(j)^T}, \quad c = 1,2,\ldots \tag{1}$$

where n_c is the number of trials belonging to the class c and n is the total number of trials. The CSP algorithm finds u as a spatial filter such that the following target function is maximized:

$$\hat{J}_c(u) = \frac{u^T . R_c . u}{u^T . R . u} \quad (2)$$

where $R = \sum_c R_c$ and u represents the spatial filter. Evidently $\hat{J}_c$ is in the form of a *Rayleigh* quotient, solution to maximization of which is given by the generalized eigenvalue problem:

$$R_c u = \lambda R u \quad (3)$$

In the traditional CSP, the eigenvector corresponding to largest eigenvalue obtained from Eq. (3) is the desired spatial vector. In many cases, instead of one eigenvector, several eigenvectors corresponding to the largest eigenvalues in Eq. (3) are chosen. Therefore, several spatial filters and consequently, several sources are estimated for each movement.

Since the eigenvectors obtained from Eq. (3) are orthogonal, the sources elicited by these eigenvectors are uncorrelated, and to some extent are independent at each given time frame [12-14]. Regarding this point, CSP can be considered as a general optimization problem, with the fitness function introduced in Eq. (2) and an additional optimization constraint. This limitation should impose on the uncorrelated sources, such that for $i \neq j$ it can be written as a zero correlation constraint in the form of $\mathbf{u}_i^T \mathrm{R}_c \mathbf{u}_j = 0$. Moreover, the optimization needs another constraint to certify the scale of the filters remains constant, such as $\mathbf{u}^T\mathbf{u} = 1$.

Thus, the CSP's algorithm can be rewritten as a new sequential optimization problem with two additional constraints. The spatial filters are estimated by optimizing the objective function mentioned in Eq. (2), in a subspace where the obtained sources remain uncorrelated.

For better conveying of the proposed idea, its procedure is depicted in Fig. 1. As seen, the first constraint provides uncorrelated sources, and second ensures the scale of the obtained filters remains constant.

Input:

- A set of training data in the form of $(X^{(j)}, y_j)$, $j = 1,2,\dots,n$.
- d: number of filters per movement.

Output:

- A set of spatial filters $\mathbf{u}_{ck}$.

For *Movement* $c \leftarrow 1 \dots M$ **do**

  For $k \leftarrow 1 \dots d$ **do**

    // obtain the k-th filter of the movement c as:

    $$\mathbf{u}_{ck} \leftarrow \arg\max_{\mathrm{u}} \frac{\mathbf{u}^T \mathrm{R}_c \mathbf{u}}{\mathbf{u}^T \mathrm{R}\, \mathbf{u}}$$

    subject to:

    1. $\mathbf{u}^T \mathrm{R}_c \mathbf{u}_l = 0, \qquad l = 1, \dots, k-1$
    2. $\mathbf{u}^T\mathbf{u} = 1$

  End

End

Fig. 1. CSP as a sequential filter estimation

b) Discriminative common spatial pattern (DCSP)

The standard CSP suffers from the lack of an explicit target function. The main flaw of the energy ratio criterion suggested in the CSP is that it is not designed to necessarily best separability among the classes. The new expression of CSP introduced in the previous subsection allows us to clarify this ambiguity by substitution of the energy ratio with a better discriminant target function, such as the

Fisher's discriminant criterion. Consider *e(.)* denotes the energy operand, then the log-energy of the sources (or extracted features) denoted by f_j can be formulated as:

$$f_j = \log\left(\mathrm{e}\left(\mathbf{u}^T\mathbf{X}^{(j)}\right)\right) \tag{4}$$

Therefore, the Fisher's target function can be written as:

$$\mathrm{J}_c(\mathbf{u}) = \frac{(\mu_c - \mu_{\tilde{c}})^2}{\sigma_c^2 + \sigma_{\tilde{c}}^2} \tag{5}$$

where μ_c and σ_c^2 are mean and variance of the f_j-s belonging to the class c, and $\mu_{\tilde{c}}$ and $\sigma_{\tilde{c}}^2$ are that of the other classes.

In order to keep the obtained sources uncorrelated to each other, similar to CSP, the zero correlation constraint is used in optimizing the DCSP's target function. In other words, by optimizing the function introduced in Eq. (5), spatial filters are determined to preserve un-correlated property among the sources. Hence, discriminative CSP (DCSP) [12] algorithm can be written as shown in Fig. 2.

Input:

- A set of training data in the form of $\left(X^{(j)}, y_j\right)$, $j = 1,2,\ldots,n$.
- d: number of filters per movement.

Output:

- A set of spatial filters $\mathbf{u}_{ck}$.

For *Movement* $c \leftarrow 1 \ldots M$ **do**

  For $k \leftarrow 1 \ldots d$ **do**

    // obtain the *k*-th filter of the movement *c* as:

$$\mathbf{u}_{ck} \leftarrow \arg\max_{\mathrm{u}}\left\{\mathrm{J}_c(\mathbf{u}) = \frac{(\mu_c - \mu_{\tilde{c}})^2}{\sigma_c^2 + \sigma_{\tilde{c}}^2}\right\}$$

    in which

$$f_j = \log\left(\mathrm{e}\left(\mathbf{u}^T\mathbf{X}^{(j)}\right)\right)$$

$$\mu_c = \mathrm{mean}\{f_j \mid y_j \in c\} \text{ and } \mu_{\tilde{c}} = \mathrm{mean}\{f_j \mid y_j \notin c\}$$

$$\sigma_c^2 = \mathrm{variance}\, f_j \mid y_j \in c\} \text{ and } \sigma_{\tilde{c}}^2 = \mathrm{variance}\,\{f_j \mid y_j \notin c\}$$

    subject to:

3. $\mathbf{u}^T\mathrm{R}_c\mathbf{u}_l = 0, \qquad l = 1,\ldots,k-1$
4. $\mathbf{u}^T\mathbf{u} = 1$

  End

End

Fig. 2. DCSP algorithm

The Fisher's criterion tries to optimize the spatial filters by maximizing between classes scattering and minimizing within classes scattering. However, the CSP's target function just focuses on between classes separating. Therefore, it is expected that DCSP will achieve more discriminant sources rather the standard CSP.

Since the Fisher's discriminant criterion is optimal for the classes with Gaussian distribution, for non-Gaussian features, it is necessary to re-distribute the scattered features to the normal one by a suitable transform. Applying the Jarque-Bera normality test [21] to the EEG features, especially in the BCI applications, the distribution of log-energy of features is nearer to the normal distribution rather than that of energy-based features [12]. Therefore, log-energy provides discriminant features with normal distribution that is an optimal case for the Fisher criteria. Moreover, experimental results were led to a

considerable improvement in terms of accuracy and stability in DCSP compared to the standard version of CSP [12].

c) Discriminative (and iterative) spatio-spectral pattern learning (DSSPL)

In this section, an iterative procedure is proposed which merges learning of discriminant feature extraction and classifier stages simultaneous to the estimation of the spatio-spectral filters, all in one optimization package. The proposed algorithm has two major steps that iteratively follow each other as explained in the forthcoming stages:

1. Spatial filters estimation:

In this step, d spatial filters are estimated by applying the DCSP to the spectrally pre-filtered signals. At the beginning, the multichannel EEGs are projected to the frequency domain using *DFT*, and then passing through d initial spectral filters. Assume that $X^{(j)}$ is the signal matrix of the *j*-th trial, in which the recorded EEG signal of each specific channel is arranged in the corresponding row. Let $\mathbf{F} = \mathcal{DFT}\{\mathbf{X}\}$, L be the number of frequency bins, and $\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_L$ are arranged in the columns of $\mathbf{F}$. Assume that the DFT coefficients of the initialized spectral filter are denoted by $\phi_1, \phi_2, \dots, \phi_L$, and $\mathbf{\Phi}$ is the diagonal matrix of these coefficients, i.e. $\mathbf{\Phi} = \mathrm{diag}(\phi_1, \phi_2, \dots, \phi_L)$. Thus, the DFT of the multichannel EEGs filtered by the *k-th* spectral filter can be termed as $\mathbf{F}\mathbf{\Phi}_k (k = 1 \dots d)$. Hence, the energy of signal $\mathbf{F}$ filtered by *k*-th spatio-spectral filter pair of $\mathbf{\Phi}_k$ and $\mathbf{u}_k$ is determined as follows:

$$\begin{aligned} \mathrm{e}\left(\mathbf{u}_k^T \mathbf{F}\mathbf{\Phi}_k\right) &= \mathbf{u}_k^T \left(\mathbf{F}\Phi_k {\Phi_k}^H \mathbf{F}^H\right)\mathbf{u}_k \\ &= \mathbf{u}_k^T (\mathbf{F}\,|\mathbf{\Phi}_k|^2 \mathbf{F}^H)\mathbf{u}_k \\ &= \mathbf{u}_k^T \left(\sum_{i=1}^{L} |\phi_{k,i}|^2 \mathbf{F}_i {\mathbf{F}_i}^H\right) \mathbf{u}_k \end{aligned} \tag{6}$$

This energy of the spatio-spectrally filtered signals should be considered as f_j in Eq. (5), and then the d spatial filters are estimated using DCSP. However, the DCSP's average zero correlation constraint must be corrected to support the initial spectral filters in a linear form. To address this objective, the following zero correlation is proposed:

$$\sum_{\substack{j=1 \\ y_j = c}}^{n} \mathbf{u}_k^T \mathbf{F}^{(j)} \mathbf{\Phi}_k \mathbf{\Phi}_{k-1} \mathbf{F}^{(j)} \mathbf{u}_{k-1} = 0 \tag{7}$$

In other words, each estimated source should be uncorrelated to the former obtained sources, while each spatial filter has a unique conjunction spectral filter. Note that the constraint proposed in Eq. (7) has a linear form with respect to $\mathbf{u}_k$.

2. Learning spectral filters jointly with the classifier:

Due to the linearity of spatial and spectral filtering procedures, both filters can be iteratively trained, such that a spectral iteration is followed by a spatial one and this procedure successively continues until the termination criterion is met. That is to say, contrary to the first step, we can first apply the already optimized spatial filters to the raw data, and then update the spectral filters based on the resultant spatially filtered data.

In Eq. (6), energy of the spatio-spectrally filtered signal was expressed as a term of spatial and spectral filters. We can denote $|\phi_{k,i}|^2$ by $\beta_{k,i}$ in Eq. (6) and then:

$$f(\mathbf{F}; \beta_k, \mathbf{u}_k) = \mathrm{e}(\mathbf{u}_k^T \mathbf{F}\mathbf{\Phi}_k)$$

$$= \mathbf{u}_k^T(\sum_{i=1}^{L} \beta_{k,i}\mathbf{F}_i\mathbf{F}_i^H)\mathbf{u}_k = \sum_{i=1}^{L} \beta_{k,i}\mathbf{u}_k^T\mathbf{F}_i\mathbf{F}_i^H\mathbf{u}_k$$
$$= \sum_{i=1}^{L} \beta_{k,i}.z_{k,i}, \qquad for\ k = 1 \dots d \tag{8}$$

where $f(\mathbf{F}; \beta_k, \mathbf{u}_k)$ is the feature obtained by the *k*-th filter pair, and $z_{k,i}$ is:

$$z_{k,i} = \mathbf{u}_k^T\mathbf{F}_i\mathbf{F}_i^H\mathbf{u}_k \tag{9}$$

The feature vector $\mathbf{f}$ is then formed by all $k = 1 \dots d$ features:

$$\mathbf{f} = [f(\mathbf{F}; \beta_1, \mathbf{u}_1) \dots f(\mathbf{F}; \beta_d, \mathbf{u}_d)]^H \tag{10}$$

The resultant vector $\mathbf{f}$ needs to be classified by a suitable classifier. A linear classifier is often employed for its good generalization ability and to avoid the potential over-fitting problem, while nonlinear classifiers are a potential tent to be over-fitted [4].

Several linear classifiers are suggested to handle EEG features including Fisher Linear Discriminant Analysis (FLDA), Linear Neural Network, Logistic Regression (LR), and Support Vector Machines (SVMs) [15]. Since the Fisher's discriminant criterion was used in optimizing the spatial filters, to establish a better match between the spatio-spectral optimization and the classifier optimizer, FLDA classifier is chosen for preserving the unity of objective function throughout the whole optimization process. For every linear classifier, a set of w weights and b bias should be estimated such that:

$$O(\mathbf{f}) = \mathbf{w}^H.\mathbf{f} + b$$

Considering Eqs. (8) and (10) results in:

$$O(\mathbf{f}) = \mathbf{w}^H.[f(\mathbf{F}; \beta_1, \mathbf{u}_1) \dots f(\mathbf{F}; \beta_d, \mathbf{u}_d)]^H + b$$
$$= \sum_{k=1}^{d}\sum_{i=1}^{L} w_k.\beta_{k,i}\ .\ z_{k,i} + b \tag{11}$$

where O is a one dimensional scalar value and in the case of two-class problem, $O(\mathbf{f})$ is passed through a sign function and $\mathbf{f}$ is classified according to the sign of O. Let us denote:

$$\widetilde{w}_{k,i} = w_k.\beta_{k,i} \tag{12}$$

Then Eq. (11) can be rewritten as:

$$O(\mathbf{f}) = \sum_{k=1}^{d}\sum_{i=1}^{L} \widetilde{w}_{k,i}\ .\ z_{k,i} + b = \widetilde{\mathbf{w}}.\mathbf{z} + b \tag{13}$$

Where

$$\widetilde{\mathbf{w}} = \left[\widetilde{w}_1^{(1)} \dots \widetilde{w}_1^{(L)} \dots \widetilde{w}_d^{(1)} \dots \widetilde{w}_d^{(L)}\right]$$
$$\mathbf{z} = \left[z_1^{(1)} \dots z_1^{(L)} \dots z_d^{(1)} \dots z_d^{(L)}\right] \tag{14}$$

Afterwards, the FLDA classifier tries to find the $\widetilde{w}_{k,i}$ weights such that the following optimization function be maximized:

$$J_c(\widetilde{\mathbf{w}}) = \frac{(\mu_c - \mu_{\tilde{c}})^2}{\sigma_c^2 + \sigma_{\tilde{c}}^2} \tag{15}$$

μ_c and σ_c^2 are mean and variance of the O_j-s belonging to the class c, and $\mu_{\tilde{c}}$ and $\sigma_{\tilde{c}}^2$ are that of the other classes. Note that $J_c(\widetilde{\mathbf{w}})$ is just a function of weights $\widetilde{w}$ and is independent of b (the constant value does not affect the variance and is eliminated from numerator by subtraction). Thus, we can set b to every arbitrary value (here $b = 0$ for the simplicity) and estimate $\widetilde{w}$ by maximizing the $J_c(\widetilde{\mathbf{w}})$ in Eq. (15).

When the classifier weights are updated, we can say one iteration is finished and the d spatial filters and $\widetilde{\mathbf{w}}$ are obtained. The initial spectral filters for the subsequent iteration can be obtained from $\widetilde{\mathbf{w}}$ with regard to Eq. (12). Since the scale and the sign of spectral coefficients have no significance, we can simply set the spectral filters to be normalized $\widetilde{\mathbf{w}}$:

$$\boldsymbol{\beta}_k = \frac{[\widetilde{w}_k^{(1)} \dots \widetilde{w}_k^{(L)}]}{|| \ [\widetilde{w}_k^{(1)} \dots \widetilde{w}_k^{(L)}] \ ||_2} \tag{16}$$

To avoid over-fitting problem, a Gaussian smoothing window is applied to the obtained filters by Eq. (16). In fact, the smoothing process acts as a generalization term. Length of the window can be found by cross validation (e.g. $L/20$ or $L/30$).

At the first iteration, the initial spectral filters can be set to cover a broad frequency range of EEGs, e.g., 7-30 Hz. As mentioned previously, this iterative procedure repeats till the termination criterion is met. This criterion can be chosen as the relative change between two consecutive iterations is less than a preset threshold, or the number of iterations exceeds than a predefined threshold.

All that remains is to adjust the decision bias b, i.e., the point along the one-dimensional subspace separating the projected features of two classes. Finding this value is brought up after termination of filters learning. The value of b can be chosen to minimize the training error, or can be estimated using a priori statistical assumption [15-16]. Due to diminishing the computational cost, the second way is chosen. By assuming the normal distribution for classifier outputs of both classes, the following structure is suggested in this study to estimate the bias value of b:

$if \quad \mu_c \geq \mu_{\tilde{c}}$:

$$b = \frac{(\mu_c - \sigma_c) + (\mu_{\tilde{c}} + \sigma_{\tilde{c}})}{2}$$

else:

$$b = \frac{(\mu_{\tilde{c}} - \sigma_{\tilde{c}}) + (\mu_c + \sigma_c)}{2} \tag{17}$$

The suggested threshold considers the deviations of means and standard deviations of the two classes. In other words, statistical characteristics of the train set for estimation of b is employed.

To show concatenated parts of this algorithm together, the structure of this algorithm is illustrated in Fig. 3, which iteratively finds the spatio-spectral filters in conjunction with the classifier weights. After estimating the $\widetilde{w}$ and b, the training phase is finished. For the test phase, $\mathbf{z}$ vectors and O values should be calculated for each EEG signal using Eqs. (9), (13) and (14), and then classification is carried out according to the O's sign.

d) Comparison with common spatio-spectral pattern (CSSP)

CSSP [8] is one of the widely used spatio-spectral filter estimation methods in BCI applications. This method assumes that frequency distortion of EEG sources in passing through head tissues can be modeled by a second order FIR filtering. In turn, this filtering is equal to adding the signal by its weighted delayed version. Thus, to obtain the source signals, we can filter the scalp signals by a set of spatial filters on the EEG signals in addition to its delayed version. If we denote the j-th source by z^j, then:

$$z^j = u^{(0)} X^j + u^{(\tau)} \delta^\tau X^j \tag{18}$$

where $u^{(0)}$ and $u^{(\tau)}$ are the spatial filters corresponding to the signal and the delayed signal, and δ^τ is an operator which postpones the starting point of the signal for τ seconds, i.e.

$$\delta^\tau X_{(t)} = X_{(t-\tau)}$$

Input:

- A set of training data in the form of $(X^{(j)}, y_j)$, $j = 1,2,\dots,n$.
- d: number of spatio-spectral filter-pairs per movement.
- A set of initial spectral filters $\mathbf{\Phi}0_k$, $k = 1 \dots d$.

Output:

- A set of the spatio-spectral filters $(\mathbf{\Phi}_{ck}, \mathbf{u}_{ck})$, $k = 1 \dots d$.
- A set of classifier weights $\widetilde{\mathbf{w}}$ and b.

For $j = 1,2,\dots,n$

$F^{(j)} = DFT\{X^{(j)}\}$

End

$\mathbf{\Phi} = \mathbf{\Phi}0$

For *Movement* $c \leftarrow 1 \dots M$ **do**

Repeat

For $k \leftarrow 1 \dots d$ **do**

// obtain the k-th spatial filter as:

$$\mathbf{u}_{ck} \leftarrow \arg\max_{\mathbf{u}} \left\{ \mathbb{J}_c(\mathbf{u}) = \frac{(\mu_c - \mu_{\bar{c}})^2}{\sigma_c^2 + \sigma_{\bar{c}}^2} \right\}$$

in which

$$f(\mathbf{F}; \mathbf{\Phi}_k, \mathbf{u}_k) = \mathbf{u}_k^T \left(\sum_{i=1}^{L} |\phi_{k,i}|^2 \mathbf{F}_i \mathbf{F}_i^H \right) \mathbf{u}_k$$

$\mu_c = \text{mean}\{f^{(j)} \,|y_j \in c\}$ and $\mu_{\bar{c}} = \text{mean}\{f^{(j)} \,|y_j \notin c\}$

$\sigma_c^2 = \text{variance}\{f^{(j)} \,|y_j \in c\}$ and $\sigma_{\bar{c}}^2 = \text{variance}\{f^{(j)} \,|y_j \notin c\}$

subject to:

1. $\sum_{\substack{j=1 \\ y_j = c}}^{n} \mathbf{u}_k^T \mathbf{F}^{(j)} \mathbf{\Phi}_k \mathbf{\Phi}_{k-1} \mathbf{F}^{(j)} \mathbf{u}_{k-1} = 0$, $\quad l = 1, \dots, k-1$
2. $\mathbf{u}^T \mathbf{u} = 1$

End

Calculate $z_k{}^{(i)} = \mathbf{u}_k^T \mathbf{F}_i \mathbf{F}_i{}^H \mathbf{u}_k$, $\quad i = 1 \dots L$

Calculate $\mathbf{z} = [z_1^{(1)} \dots z_1^{(L)} \;\dots\; z_d^{(1)} \dots z_d^{(L)}]$

// obtain the classifier weights as:

$$\widetilde{\mathbf{w}} \leftarrow \arg\max_{\mathbf{u}} \left\{ \mathbb{J}_c(\widetilde{\mathbf{w}}) = \frac{(\mu_c - \mu_{\bar{c}})^2}{\sigma_c^2 + \sigma_{\bar{c}}^2} \right\}$$

in which:

$O = \widetilde{\mathbf{w}}.\mathbf{z} + b$

$\mu_c = \text{mean}\{O^{(j)} \,|y_j \in c\}$ and $\mu_{\bar{c}} = \text{mean}\{O^{(j)} \,|y_j \notin c\}$

$\sigma_c^2 = \text{variance}\{O^{(j)} \,|y_j \in c\}$ and $\sigma_{\bar{c}}^2 = \text{variance}\{O^{(j)} \,|y_j \notin c\}$

// obtain the initial spectral filters for next iteration as:

$$(\mathbf{\Phi}_k)^2 = \frac{[\widetilde{w}_k^{(1)} \dots \widetilde{w}_k^{(L)}]}{\| \, [\widetilde{w}_k^{(1)} \dots \widetilde{w}_k^{(L)}] \, \|_2}$$

Until termination criterion is satisfied

Obtain the value of b from (16).

End

Fig. 3. DSSPL algorithm

We can rewrite the Eq. (18) as follows:

$$z^j = [u^{(0)} \quad u^{(\tau)}] \begin{bmatrix} X^j \\ \delta^\tau X^j \end{bmatrix} = \hat{u}\hat{X}^j \tag{19}$$

CSSP tries to find the $\hat{u}$ vector using the CSP's target function, and in the same way solve it by the eigenvalue decomposition. In this regard, finding a set of $\hat{u}$ spatial filters covers estimation of the joint spatio-spectral filters. The value of τ can be obtained by line search leading to proper classification accuracy.

CSSP is a simple, fast and efficient procedure, but it still suffers from the mentioned problematic fitness function of CSP. In addition, CSSP uses a simple FIR spectral filter, which cannot cover all of the frequency distortions.

The elicited features by this method are categorized in a separate step. The employed classifier weights may be optimized in a different condition from the CSP's criterion. The separate optimization can decrease the procedure's performance.

In contrast, the proposed method (DSSPL) has a more efficient merit function (as described in II-B), uses comprehensive spectral filters, and a more coordinated integrated classifier.

e) Comparison with iterative spatio-spectarl pattern learning (ISSPL)

ISSPL [10] uses an iterative scheme to optimize spatial filters, spectral filters and classifier weights. In the first step, the spatial filters are obtained for each class using traditional CSP, such that the energy ratio of the corresponding class source over other classes is maximized. For example, for a two-class problem, we have to obtain two different sets of spatial filters and two sets of sources consequently. Next, energy of the obtained sources is fed to a SVM classifier. If the energy of both classes is equally maximized (as is done in ISSPL) the classifier cannot separate them. To address this problem, ISSPL tries to optimize the SVM classifier weights with positive sign for the first class and negative sign for another one. Then, for each class the spectral filters are obtained from average absolute value of the classifier weights (similar to Eq. (16)). These filters are used as initial spectral filters in the next iteration. The algorithm of ISSPL is described briefly in Fig. 4.

Although both ISSPL and DSSPL methods use an iterative algorithm, there are many essential differences between them. At first, the ISSPL uses traditional CSP in spatial filter estimation, which is very sensitive to outliers and noisy samples, while the DSSPL uses an improved version of CSP (DCSP).

In addition, DSSPL uses a fixed criterion and target function to optimize spatial filters, spectral filters and classifier weights. In contrast, ISSPL optimizes these three items under different criterions (such as CSP and SVM target function). Because of iterative structure in both of the methods, these three steps (optimizing spatial filters, spectral filters and classifier weights) are executed successively and are not independent. Thus, the mentioned unity in DSSPL can improve the final classification rate, e.g., in ISSPL, the CSP tries to optimize the spatial filters and remove the irrelevant components. However, in the next step, optimizing the spectral filters and classifier weights using the SVM criterion leads to the removal of other irrelevant components which are different from those that were removed by CSP in the previous step. This inconsistency can be continued in the next iterations and consequently causes lower performance in ISSPL.

Another difference emerges from the computational cost point of view. Although the ISSPL uses a simple CSP merit function to optimize the spatial filters, it has to estimate more spatio-spectral filters than ISSPL. In other words, for a two-class problem, ISSPL estimates two sets of spatio-spectral filters and classifier weights, while in DSSPL it diminishes to one set, which leads to a lower run time in practice.

Input:
- A set of training data in the form of $(X^{(j)}, y_j)$, $j = 1,2,\dots,n$.
- d: number of spatio-spectral filter-pairs per movement.
- A set of initial spectral filters $\boldsymbol{\Phi}0_k$, $k = 1 \dots d$.

Output:
- A set of the spatio-spectral filters $(\boldsymbol{\Phi}_{ck}, \mathbf{u}_{ck})$, $k = 1 \dots d$.
- A set of classifier weights $\tilde{\mathbf{w}}$ and b.

For $j = 1,2,\dots,n$
$F^{(j)} = DFT\{X^{(j)}\}$
End
$\boldsymbol{\Phi} = \boldsymbol{\Phi}0$
For *Movement* $c \leftarrow 1 \dots M$ **do**
Repeat
For $k \leftarrow 1 \dots d$ **do**
Obtain $\mathbf{u}_{ck}$ and $\mathbf{u}_{\bar{c}k}$ using CSP algorithm
Obtain the classifier weights $\tilde{\mathbf{w}}$ and b by optimizing the SVM criteria.
Obtain the spectral filters $\boldsymbol{\Phi}_{ck}$ and $\boldsymbol{\Phi}_{\bar{c}k}$ from average of $\tilde{\mathbf{w}}$-s.
End
Until termination criterion is satisfied
End

Fig. 4. ISSPL algorithm

f) Comparison with second order bilinear discriminant analysis (SOBDA)

Suppose that $X^{(j)}$ and $y_j \in [-1,1]$ represent *j*-th trial of multichannel EEGs and its corresponding class label, respectively. SOBDA [11] defines the following discriminant function:

$$f(X;\theta) = C\,Trace(\boldsymbol{U}^T X \boldsymbol{V}) + (1-C)Trace(\Lambda \boldsymbol{A}^T X \boldsymbol{B}\boldsymbol{B}^T X^T \boldsymbol{A}) + w0 \quad (20)$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are spatial and spectral filters for the linear term, $\boldsymbol{A}$ and $\boldsymbol{B}$ are spatial and spectral filters in the quadratic term, Λ is diagonal matrix of the weights corresponding to each filter-pair and each class, *w0* is the bias, and *C* is a regularization term between the mentioned two major parts. It results in positive and negative values for trials with $y_j = +1$ and $y_j = -1$, respectively.

The discriminant function in Eq. (20) has two major parts; first, a linear term which contains amplitude of the signals, and second, a quadratic term which contains power of the signals. In other words, the linear term captures phase-locked event-related potentials and synchronous discriminant features in the EEG signal, while the quadratic term elicite asynchronous and second-order statistics of the signal.

If we integrate all of the SOBDA's parameters in one vector denoted as $\theta = \{\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{A}, \boldsymbol{B}, \Lambda, w0\}$, the SOBDA tries to optimize the θ using Logistic Regression (LR) criteria. The parameter of this curve is estimated by the following likelihood function:

$$L(\theta) = -\sum_{j=1}^{n} \log\left(1 + e^{-y_j f(X;\theta)}\right) \quad (21)$$

Where *n* is number of trials, y_j is the label of *j*-th sample, $L(\Theta)$ is the regression function, and $f(X,\Theta)$ is the discriminant function described in Eq. (20).

The SOBDA uses both first and second order components to estimate the spatio-spectral filters with the objective of achieving better classification accuracy. However, it suffers from some drawbacks, which

causes high computational complexity. The first drawback comes back to its optimization algorithm, which is not iterative, and it is difficult to optimize four sets of filters together. This problem would be more critical if the number of sources increase. In addition, in the quadratic part of Eq. (20), the spatial and spectral filters are obtained per each class, hence, two different sets of filters for a two-class problem should be estimated. Thus, it can be said that SOBDA is a time consuming and complex algorithm because it has to estimate a large number of filters.

3. MATERIALS AND IMPLEMENTATION CRITERIA

a) Implementation details

To evaluate how the proposed algorithm acts effectively, four considerable schemes in spatial and spatio-spectral filtering field including CSP [7], CSSP [8], ISSPL [10] and SOBDA [11], were implemented and all of them were applied to the two different BCI datasets with different spatial resolution.

In all of the experiments, EEG signals were first filtered between 7-30 Hz in the preprocessing stage or the initial spectral filters were selected such that the frequency interval of 7-30 Hz was equally covered. The number of spatial or spatio-spectral filters (depending on the method) ranged from 1 to 5, and for each method, that number of filters providing the best classification accuracy is chosen. To fairly compare the methods in terms of accuracy and time processing, the same optimization algorithm for all of the implementations is chosen: a specific non-linear optimization method which uses interior-point algorithm with a combination of line search and trust region steps [17-18], [23].

In CSP and CSSP methods the extracted features are classified by SVM and FLD classifiers, while the three other methods (ISSPL, SOBDA and DSSPL) are naturally equipped with their own classifier. Among the compared methods, just ISSPL and the proposed method (DSSPL) benefit from an iterative learning scheme. Maximum number of iterations for both methods was set to 3. In the SOBDA implementation, all of the free parameters were found by line search and cross validation, according to what was suggested in the original paper [11].

b) Datasets

Two standard EEG datasets with different spatial resolution are employed to assess the proposed method compared to the mentioned competitive approaches. The datasets include *Graz BCI* [19] and *IVa (BCI Competition III)* [23] that are separately described as follows:

Graz BCI Dataset: Three normal subjects (S1, S2, and S3) whose age ranged between 25 and 35, were trained to concentrate on five mental tasks including; movement imagination of left hand, right hand, tongue, foot along with an arithmetic task. Each subject was required to perform these imaginations for 3.5 s in an 8-second paradigm. The signals are recorded from 29 gold electrodes according to the 10-20 standard recording system. The signals first were filtered between 0.5 to 30 Hz and sampled at 256 Hz [19]. In this study, only two-classes (out of five classes) are investigated including left and right imagery movements. After removing trials with a high level of artifact noises, totally 228, 223 and 188 trials remained for S1, S2 and S3 subjects, respectively. These trials almost equally belonged to left and right hand imagery movements.

Dataset IVa in BCI Competition III: The EEG data were recorded from five healthy subjects labeled as “aa”, “al”, “av”, “aw”, and “ay”. According to the international 10-20 recording system, 118 channels were placed on their scalp. Signals were down sampled to 100 Hz and filtered between 0.5-30 Hz. During

each trial, subjects were required to perform either of two motor imagery tasks for 3.5 s: right-hand and right-foot imagery movements. The total of 140 trials were collected for each subject and each task [23].

4. RESULTS

a) Classification rate

Since we collect enough number of samples, ten times ten fold cross validation is used to determine the classification accuracy. The accuracy of applying DSSPL and the other four rival methods is shown in Tables1 and 2 for *Graz BCI* and *BCI competition* datasets, respectively. The contents of the tables are arranged as *mean* ± *standard deviation* of the classification accuracy. As mentioned before, each method selects a number of filters (ranging from 1 to 5), leading to its highest performance.

Table 1. Classification results of applying csp, cssp, isspl, sobda and dsspl algorithms to eeg signals of graz bci dataset

Subjects	CSP+SVM	CSP+FLD	CSSP+SVM	CSSP+FLD	ISSPL	SOBDA	DSSPL
S1	71.4±8.1	73.1±6.4	79.7±7.0	78.3±6.6	84.5±5.9	88.6±5.4	89.5±5.0
S2	81.2±7.0	80.1±5.5	89.6±5.2	88.7±5.1	90.0±5.0	92.6±4.8	90.2±4.1
S3	72.1±7.5	72.2±7.1	77.9±7.9	78.1±7.0	83.2±6.0	87.3±5.8	88.3±5.3

Table 2. Classification results of applying csp, cssp, isspl, sobda and dsspl algorithms to bci competition iii dataset

Subjects	CSP+SVM	CSP+FLD	CSSP+SVM	CSSP+FLD	ISSPL	SOBDA	DSSPL
aa	72.3±5.0	72.7±5.5	74.0±4.8	74.4±5.0	88.2±4.0	92.9±3.9	93.7±3.5
al	89.6±4.1	89.2±3.5	90.6±2.1	91.0±2.0	96.2±2.2	99.0±0.2	97.8±0.6
av	74.9±7.1	75.3±6.0	73.8±5.8	74.1±6.0	77.0±5.5	78.1±5.6	81.6±3.6
aw	86.3±5.2	87.0±5.5	91.3±4.0	90.2±4.4	93.6±4.0	95.2±3.3	95.0±2.5
ay	90.5±4.8	88.1±3.9	89.4±5.1	88.9±5.5	93.0±3.2	94.9±3.0	94.2±3.4

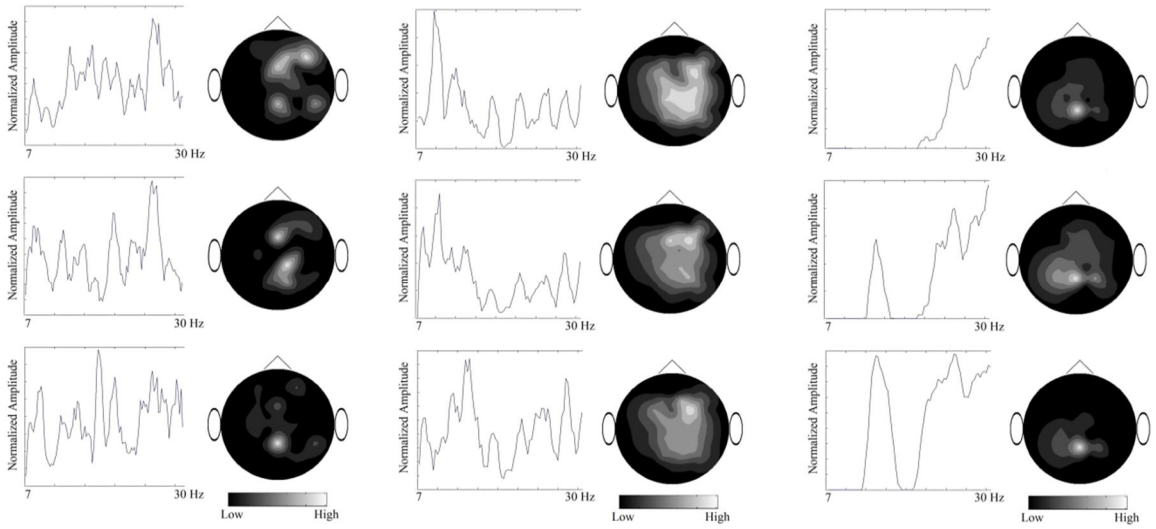
As we can see, the obtained results by the proposed method are highly superior to those of standard CSP and CSSP. This drastic difference is resulted by considering more spectral components that involve more discriminant frequency intervals in the decision making process. Moreover, DSSPL acts more precisely than ISSPL because it considers the objective function of DCSP instead of CSP, in addition to unifying the whole optimization steps under a certain criterion.

The SOBDA uses both first and second order components to estimate the spatio-spectral filters, which leads to achieving high classification accuracy. Accuracy of DSSPL and SOBDA are almost equivalent; however, DSSPL seems to be more proper in hard cases, while SOBDA is superior in good cases. Our definition of the good cases is those subjects who provide separable features and accordingly result in a higher classification rate in all of the methods (S2, al, aw, ay). In contrast, hard cases refer to those who have lower classification rate (S1, S3, aa, av). The produced results show DSSPL provides lower standard deviation around the mean classification accuracy. Consequently, it can be claimed that DSSPL is more robust and stable compared to SOBDA.

Another advantage of DSSPL that improves its accuracy originates from average zero correlation constraint applied to estimating the spatio-spectral filters. This constraint leads to minimization of the mutual information among the elicited features. Most of the other methods use this constraint just for estimating the spatial filters.

b) Spatio-spectral filters

Unity of the optimization function in the proposed method is one of its advantages compared to the other similar approaches such as ISSPL. For further explanation, the updating procedure of first spatial and the corresponding spectral filters estimated from the subject *S3* are shown in Fig. 5. As seen, column (a) shows the spatio-spectral filters obtained after each of the three iterations for DSSPL. In addition, for ISSPL these filters are shown in columns (b) and (c) for the left and right hand imagery movements, respectively.



(a) (b) (c)

Fig. 5. Updating procedure of spatial and corresponding spectral filters for DSSPL in column (a) and for ISSPL in columns (b) and (c)

It is obvious that spatial and spectral patterns are entirely different for these two methods which originated from applying two different optimization functions in filter estimation stage. DSSPL tries to achieve discriminative spatial and frequency bands, while ISSPL aims to find spatio-spectral patterns providing higher energy ratio for each class, separately.

In DSSPL, after three iterations, spatial filters tend to focus on parietal lobe of the head. It confirms the physiological assumption that says discriminative information of movement imageries arise from this region of head [20]. However, in ISSPL, no clear pattern can be seen in the obtained spatial filters at each epoch. This deficiency comes back to utilizing different objective functions in the training phase of spatial and spectral filters.

c) Computational cost

All of the mentioned algorithms were implemented in *MATLAB (R2010a version)* on *Windows 7* using "*Intel(R) Core(TM)2 Duo T9300@ 2.50GHz 2.5GHz - 3.00GB RAM*" platform. Approximation of processing time for estimating 5 filters (or 5 filter-pairs) in all of the methods is summarized in Table 3. In this study, subject *aa* from *BCI Competition III* dataset was chosen, while just 1-time 1-fold of the 10 times 10 fold cross validation process was considered.

Although the processing time is not a precise gauge for assessment of an algorithm complexity, some studies use it as a rough indicator instead of the computational order. This is due to the fact that the solution cannot be expressed as a closed form for some methods such as the proposed one in this study; therefore, the only way to compare the algorithms in terms of complexity is the calculation of their running time.

Table 3. Processing time of applying CSP, CSSP, ISSPL, SOBDA and DSSPL algorithms to the subject AA in BCI competition III dataset

Method	**CSP+SVM CSP+FLD**	**CSSP+SVM**	**CSSP+FLD**	**ISSPL**	**SOBDA**	**DSSPL**
Processing Time	less than 1 min	less than 2 min		14 min	21 min	8 min

In CSP, spectral filtering is applied as a preprocessing step, and CSSP algorithm needs to estimate a few parameters in determining its spectral filters. Thus, as expected and shown in Table 3, they are executed in a short time. In contrast, in the other three methods, spectral filters are optimized to provide a high resolution, which causes high computational cost accordingly.

In the ISSPL and SOBDA, it is necessary to estimate the spatial and spectral filter matrices two times (one for each class), which is the main reason for the long running time. ISSPL tries to estimate these filters in an iterative procedure; spatial filters using the simple CSP, and spectral filters using a parameterized version of SVM. This trick causes less executing time than SOBDA, but regarding lack of a direct relation in estimation of spectral and spatial filters (difference of optimization criteria), the classification accuracy is not convincing in ISSPL [10, 11] as shown in Tables 1 and 2.

DSSPL tries to estimate spatial and spectral filters one time for both classes (not for each class separately as in ISSPL or SOBDA). This simplicity, in addition to merging of feature extraction and classification in an iterative manner, causes less processing time than that of ISSPL and SOBDA. Moreover, the same optimization criterion which is based on maximum discrimination is used for all steps of the algorithm making it more accurate than ISSPL.

5. CONCLUSION

In this research, an iterative spatio-spectral algorithm is introduced using a combination of discriminative CSP (DCSP) and a specific parameterized version of FLD classifier. In the proposed method (DSSPL), spatial filters are estimated using DCSP, and then spectral filters and FLDA classifier weights are adjusted simultaneously. Using a fixed criterion in all optimizations leads to an integrated and coordinated scheme. In addition, incorporating an independency constraint in DSSPL leads to minimizing the mutual information among the features and consequently proving higher classification rate. Moreover, by using Fisher's discriminant criterion, just one set of spatio-spectral filters is employed for classification of the two-class signals. However, in most of the other similar methods it is necessary to find a set of filters for each class.

These properties cause efficiency of our method in high accuracy and low running time. For evaluation, the proposed method and four other similar schemes were implemented over the two highly cited BCI datasets. Experimental results show that considering computational cost and classification rate, our algorithm has higher performance overall.

For further work, other linear or non-linear discriminant criterions can be parameterized and used in the feature extraction and classification formulation. Moreover, the effect of using kernel functions in such approaches (spatio-spectral filtering) can be examined and discussed. In addition, utilization of spatio-spectral filtering for source localization can be the optimistic horizon of this approach.

***Acknowledgment*:** For provision of the data, special thanks go to: *Graz BCI team:* Dr. Obermaier and Prof. Pfurtscheller. *Berlin BCI group:* Klaus-Robert Muller, Benjamin Blankertz, and Gabriel Curio.

REFERENCES

1. Graimann, B., Allison, B. & Pfurtsheller, G. (2010). *Brain-computer interfaces, revolutionizing human-computer interaction. 2ed,* Springer, New York.

2. Rezai, R. F. (2011). *Recent advances in brain-computer interface systems.* 1st ed, Intech, Croatia.
3. Phukan, J., Pender, N. P. & Hardiman, O. (2007). Cognitive impairment in amyotrophic lateral sclerosis. Vol. 6, No. 11, pp. 994–1003.
4. Muller, K. R., Krauledat, M., Dornhege, G., Curio, G. & Blankertz, B. (2004). Machine learning techniques for brain-computer interfaces. *Biomed. Eng.*, Vol. 49, No. 1, p. 11.
5. Fukunaga, K. & Koontz, W. L. G. (1970). Application of the karhunen-love expansion to feature selection and ordering. Vol. C-19, No. 4, pp. 311–318.
6. Koles, Z. J. (1991). The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. *Electroencephalogr. Clin. Neurophysiol.*, Vol. 79, pp. 440–447.
7. Ramoser, H., Muller-gerking, J. & Pfurtscheller, G. (1999). Designing optimal spatial filters for single-trial EEG classification in a movement task. *IEEE Trans. Rehab. Eng*, Vol. 8, pp. 441–446.
8. Lemm, S., Blankertz, B., Curio, G. & Muller, K. R. (2005). Spatio-spectral filters for improved classification of single trial EEG. *IEEE Trans. Biomed. Eng*, Vol. 52, pp. 1541–1548.
9. Dornhege, G., Blankertz, B., Krauledat, M., Losch, F., Curio, G. & Muller, K. R. (2006). Combined optimization of spatial and temporal filters for improving brain-computer interfacing. *IEEE trans. Biomed. Eng.*, Vol. 53, No. 11.
10. Wu, W., Gao, X., Hong, B. & Gao, S. (2008). Classifying single-trial eeg during motor imagery by iterative spatio-spectral patterns learning (isspl). *Biomedical Engineering, IEEE Transactions on*, Vol. 55, No. 6, pp. 1733–1743.
11. Christoforou, C., Haralick, R., Sajda, P. & Parra, L. C. (2010). Second-order bilinear discriminant analysis. *Machine Learning Research*, Vol. 11, pp. 665–685.
12. Nasihatkon, B. (2008). A general framework to estimate optimal spatial and spatio-spectral filters for EEG signal classification. Master's thesis, Shiraz University.
13. Strang, G. (1988). *Linear algebra and its applications*. Brooks/Cole, pp. 347–354.
14. Parra, L. C. & Sajda, P. (2003). Blind source separation via generalized eigenvalue decomposition. *Journal of Machine Learning Research*, Vol. 4, pp. 1261–1269.
15. Duda, R. O., Hart, P. E. & Stork, D. G. (1997). *Pattern classification, 2nd*. Portola Valley, California: Wiley.
16. Bishop, C. M. (2006). *Pattern Recognition And Machine Learning*. New York: Springer.
17. Waltz, R. A., Morales, J. L., Nocedal, J. & Orban, D. (2006). An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Mathematical programming*. Vol. 107, No. 3, pp. 391-408.
18. Wachter, A. & Biegler, L. T. (2005). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Springer, Math. Program.*
19. Obermaier, B., Neuper, C., Guger, C. & Pfurtscheller, G. (2001). Information transfer rate in a five-classes brain-computer interface. *IEEE Trans. On Neural Systems and Rehabilitation Eng.*, Vol. 9, No. 3.
20. Scarabino, T., Giovanni, S., Salvolini, U., Salleand, F. D., Duvernoy, H. & Rabischong, P. (2006). *Atlas of morphology and functional anatomy of the brain*.
21. Jarque, C. M. & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, Vol. 6, No. 3, pp. 255-259.
22. Guido Dornhege, G. C., Blankertz, B. & Muller, K. R. (2004). Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms. *IEEE Trans. Biomed. Eng.*, Vol. 51, No. 6, pp. 993-1002.
23. Hatam, M. & Masnadi-shirazi, M. A. (2008). Analytical discrete optimization. *Iranian Journal of Science & Technology, Transaction B, Engineering*, Vol. 32, No. B3, pp. 249-263.