MULTILINEAR GENERALIZATION OF COMMON SPATIAL PATTERN

Qibin Zhao[1], *Liqing Zhang*[1*], *Andrzej Cichocki*[2]

[1]Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
[2]Lab for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Japan

ABSTRACT

The Common Spatial Patterns (CSP) algorithm has been widely used in EEG classification and Brain Computer Interface (BCI). In this paper, we propose a multilinear formulation of the CSP, termed as TensorCSP or Common Tensor Discriminant Analysis (CTDA) for high-order tensor data. As a natural extension of CSP, the proposed algorithm uses the analogous optimization criteria in CSP and a new framework for simultaneous optimization of projection matrices on each mode based on tensor analysis theory is developed. Experimental results demonstrate that our proposed algorithm is able to improve classification accuracy of multi-class motor imagery EEG.

Index Terms— EEG, Tensor, Common Spatial Pattern, Brain Computer Interface

1. INTRODUCTION

Tensors (also known as n-way arrays or multidimensional arrays) are used in a variety of applications ranging from neuroscience and psychometrics to chemometrics [1, 2, 3]. From a viewpoint of data analysis, tensor analysis is very attractive because it takes into account spatial and temporal correlations between variables more accurately than 2D matrix factorizations, and it usually provides sparse common factors or hidden components with physiological meaning and interpretation. In most applications, especially in neuroscience (EEG, fMRI), the standard PARAFAC and Tucker models were used[4, 5].

In order to implement a reliable BCI system, an effective discrimination of different mental states from EEG recordings is very important. To this end, Common Spatial Patterns (CSP) [6, 7] has proven to be very powerful in determining spatial filters which extract discriminative brain rhythms for the motor imagery (MI) based BCI. In the most of previous works on EEG classification, EEG data is represented as a matrix in high-dimensional space. However, more information such as frequency were not considered. In this paper, motivated by the successes of the tensor LDA[8] and tensor subspace analysis[9], we extend CSP algorithm to high-order tensor and propose a novel tensor-based CSP algorithm, called common tensor discriminant analysis (CTDA), for EEG classification in MI based BCI. The experimental results demonstrate the effectiveness of the proposed algorithm and tensor representations of EEG signals.

2. TENSOR ALGEBRA

Tensors are multidimensional arrays which transform linearly under coordinate transformations. The order of a tensor $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times \ldots N_M}$ is M. An element of $\mathcal{X}$ is denoted by $\mathcal{X}_{n_1 n_2 \ldots n_M}$, where $1 \leq n_i \leq N_i$ and $1 \leq i \leq M$. We introduce the following tensor operations relevant to this paper.

The outer product of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_M}$ and another tensor $\mathcal{Y} \in \mathbb{R}^{J_1 \times J_2 \times \ldots J_N}$ is $\mathcal{X} \circ \mathcal{Y}$ defined by

$$(\mathcal{X} \circ \mathcal{Y})_{i_1 i_2 \ldots i_M j_1 j_2 \ldots j_N} = \mathcal{X}_{i_1 i_2 \ldots i_M} \mathcal{Y}_{j_1 j_2 \ldots j_N}. \tag{1}$$

The contraction of a tensor is obtained by equating two indices and summing over all values of the repeated indices. Contraction reduces the tensor order by 2. For example, given two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$; the outer product of $\mathbf{x}$ and $\mathbf{y}$ is $\mathbf{Z} = \mathbf{x} \circ \mathbf{y}$; and the contraction of $\mathbf{Z}$ is $\mathbf{Z}_{ii} = \mathbf{x}^T \mathbf{y}$, where the repeated indices imply summation. The value of $\mathbf{Z}_{ii}$ is the inner product of $\mathbf{x}$ and $\mathbf{y}$. In general, for tensors $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times \ldots \times N_M \times K_1 \times K_2 \times \ldots \times K_L}$ and $\mathcal{Y} \in \mathbb{R}^{N_1 \times N_2 \times \ldots \times N_M \times P_1 \times P_2 \times \ldots \times P_Q}$, the contraction on the tensor product $\mathcal{X} \circ \mathcal{Y}$ is

$$\begin{aligned} &[[\mathcal{X} \circ \mathcal{Y}; (1:M)(1:M)]]_{k_1 \ldots k_L p_1 \ldots p_Q} = \\ &\quad \sum_{n_1=1}^{N_1} \cdots \sum_{n_M=1}^{N_M} \mathcal{X}_{n_1 \ldots n_M k_1 \ldots k_L} \mathcal{Y}_{n_1 \ldots n_M p_1 \ldots p_Q}. \end{aligned} \tag{2}$$

When the contraction is conducted on all indices except the i-th index on the tensor product of $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{N_1 \times N_2 \times \ldots \times N_M}$, we denote this procedure as

$$\begin{aligned} & [[\mathcal{X} \circ \mathcal{Y}; (\bar{i})(\bar{i})]] \\ = \; & [[\mathcal{X} \circ \mathcal{Y}; (1:i-1, i+1:M)(1:i-1, i+1:M)]] \\ = \; & \sum_{n_1=1}^{N_1} \cdots \sum_{n_{i-1}=1}^{N_{i-1}} \sum_{n_{i+1}=1}^{N_{i+1}} \cdots \sum_{n_M=1}^{N_M} \mathcal{X}_{n_1 \ldots n_M} \mathcal{Y}_{n_1 \ldots n_M} \\ = \; & \mathrm{mat}_i(\mathcal{X}) \mathrm{mat}_i^T(\mathcal{Y}) = \mathcal{X}_{(i)} \mathcal{Y}_{(i)}^T, \end{aligned} \tag{3}$$

where $\mathrm{mat}_d(\mathcal{X})$ or $\mathcal{X}_d$ denotes the mode-d matricizing of $\mathcal{X}$, and $[[\mathcal{X} \circ \mathcal{Y}; (\bar{i})(\bar{i})]] \in \mathbb{R}^{N_i \times N_i}$.

The mode-d product of a M order tensor $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times \ldots \times N_M}$ and a matrix $\mathbf{U} \in \mathbb{R}^{J \times N_d}$ is $\mathcal{X} \times_d \mathbf{U} \in \mathbb{R}^{N_1 \times \ldots \times N_{d-1} \times J \times N_{d+1} \times \ldots \times N_M}$ defined by

$$\begin{aligned} \mathcal{X} \times_d \mathbf{U} &= \sum_{n_d=1}^{N_d} (\mathcal{X}_{n_1 n_2 \ldots n_M} \mathbf{U}_{j n_d}) \\ &= [[\mathcal{X} \circ \mathbf{U}; (d)(2)]]. \end{aligned} \tag{4}$$

3. COMMON TENSOR DISCRIMINANT ANALYSIS

The CTDA tries to find the most discriminative tensor subspace. By projecting the data points into the tensor subspace, different classes have the most differing power ratios whereas the sum of total power

*Corresponding author: Liqing Zhang, Email: zhang-lq@cs.sjtu.edu.cn

remains constant. Here the power is calculated by the variance in the time domain.

The CTDA algorithm is trained on labeled tensor data, i.e., we have a set of M order tensor samples $\mathcal{X}_i \in \mathbb{R}^{N_1 \times \ldots \times N_M}, i = 1, \ldots, n$. We can define the mode-M covariance tensor (i.e., high order covariance) as

$$\mathcal{R} = \frac{1}{n} \sum_{i=1}^{n} [[\mathcal{X}_i \circ \mathcal{X}_i; (M)(M)]], \tag{5}$$

where $\mathcal{R} \in \mathbb{R}^{N_1 \times \ldots \times N_{M-1} \times N_{M-1} \ldots \times N_1}$ is a $2(M-1)$ order tensor which has a symmetric length on first $M-1$ and last $M-1$ mode.

The idea of CTDA is to find $M-1$ filters $\mathbf{W}_k, k = 1, \ldots M-1$ which can simultaneously diagonalize two covariance tensor. Similar to the CSP algorithm in multi-class case, $\mathbf{W}_k^{(c)}, k = 1, \ldots, M-1; c = 1, \ldots, C$ are calculated respectively according to the criteria defined as:

$$\begin{aligned} &\mathcal{R}^{(c)} \times_1 \mathbf{W}_1^{(c)T} \ldots \times_{M-1} \mathbf{W}_{M-1}^{(c)}{}^T \times_M \mathbf{W}_{M-1}^{(c)} \ldots \\ &\qquad \times_{2(M-1)} \mathbf{W}_1^{(c)} = \mathcal{D}^{(c)}, \quad \text{and} \\ &\left(\sum_{c=1}^{C} \mathcal{R}^{(c)} \right) \times_1 \mathbf{W}_1^{(c)T} \ldots \times_{M-1} \mathbf{W}_{M-1}^{(c)}{}^T \times_M \mathbf{W}_{M-1}^{(c)} \ldots \\ &\qquad \times_{2(M-1)} \mathbf{W}_1^{(c)} = \mathcal{I}, \end{aligned} \tag{6}$$

where $\mathcal{R}^{(c)}$ denotes a covariance tensor of the c-th class tensor data calculated by Eq.(5), $\mathcal{D}^{(c)}$ and $\mathcal{I}$ are two superdiagonal tensors. $\mathbf{W}_k^{(c)}, k = 1, \ldots M-1$ are projection matrices on the k-th mode for simultaneously diagonalizing c-th class covariance tensor $\mathcal{R}^{(c)}$ and total covariance tensor $\sum_{c=1}^{C} \mathcal{R}^{(c)}$. This procedure is illustrated as Fig.1 with four order covariance tensor $\mathcal{R}^{(c)}$ and two projection matrices of $\mathbf{W}_1$ and $\mathbf{W}_2$. After calculating the $\mathbf{W}_k^{(c)}$ for each class c, we obtain the $\mathbf{W}_k = [\mathbf{W}_k^{(1)}, \ldots, \mathbf{W}_k^{(C)}], k = 1, \ldots, M-1$, which are optimal projection matrices on k-th mode for discrimination and the tensor product of corresponding projection vectors on each mode $\mathbf{W}_k|_{k=1}^{M-1}$ represents a projection tensor.

According to the CSP objective functions, let $\mathbf{W}^{(c)}$represents the maximal discriminative pattern for c-th class, and $\mathbf{X}_{c;i}$ denotes the i-th training sample, which is a two order tensor (i.e., matrix), belonging to the $c \in [1, \ldots, C]$-th class. We have

$$\begin{aligned} \mathbf{D} &= \mathbf{W}^{(c)T} \mathbf{R}^{(c)} \mathbf{W}^{(c)} \\ &= \frac{1}{n_c} \sum_{i=1}^{n_c} \left[(\mathbf{W}^{(c)T} \mathbf{X}_{c;i})(\mathbf{W}^{(c)T} \mathbf{X}_{c;i})^T \right] \\ &= \frac{1}{n_c} \sum_{i=1}^{n_c} \left[\left[(\mathbf{X}_{c;i} \times_1 \mathbf{W}^{(c)T}) \circ (\mathbf{X}_{c;i} \times_1 \mathbf{W}^{(c)T}); (2)(2) \right] \right] \\ &= \frac{1}{n_c} \sum_{i=1}^{n_c} \left[\left[(\mathbf{X}_{c;i} \times_1 \mathbf{W}^{(c)T}) \circ (\mathbf{X}_{c;i} \times_1 \mathbf{W}^{(c)T}); (\bar{1})(\bar{1}) \right] \right]. \end{aligned} \tag{7}$$

Similarly, the constrain condition can also be written as:

$$\begin{aligned} \mathbf{I} &= \mathbf{W}^{(c)T} \left(\sum_{c=1}^{C} \mathbf{R}^{(c)} \right) \mathbf{W}^{(c)} \\ &= \sum_{c=1}^{C} \sum_{i=1}^{n_c} \frac{1}{n_c} \left[\left[(\mathbf{X}_{c;i} \times_1 \mathbf{W}^{(c)T}) \circ (\mathbf{X}_{c;i} \times_1 \mathbf{W}^{(c)T}); (\bar{1})(\bar{1}) \right] \right] \end{aligned} \tag{8}$$

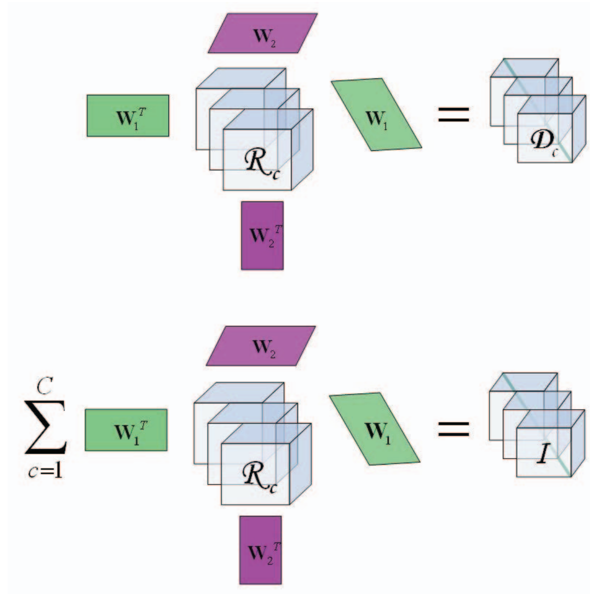


Fig. 1. The objective functions of common tensor discriminant analysis.

Based on analogy with Eq.(7) and (8), we define CTDA by replacing $\mathbf{X}_{c;i}$, $\mathbf{D}$ and $\mathbf{I}$ with $\mathcal{X}_{c;i}, \mathcal{D} \in \mathbb{R}^{H_1 \times \ldots \times H_{M-1} \times H_{M-1} \times \ldots \times H_1}$ and $\mathcal{I} \in \mathbb{R}^{H_1 \times \ldots \times H_{M-1} \times H_{M-1} \times \ldots \times H_1}$ respectively, then we obtain

$$\begin{aligned} &\frac{1}{n_c} \sum_{i=1}^{n_c} \left[\left[\left(\mathcal{X}_{c;i} \prod_{k=1}^{M-1} \times_k \mathbf{W}_k^{(c)T} \right) \circ \left(\mathcal{X}_{c;i} \prod_{k=1}^{M-1} \times_k \mathbf{W}_k^{(c)T} \right); (M)(M) \right] \right] \\ &\qquad = \mathcal{D}, \\ &\sum_{c=1}^{C} \sum_{i=1}^{n_c} \frac{1}{n_c} \left[\left[\left(\mathcal{X}_{c;i} \prod_{k=1}^{M-1} \times_k \mathbf{W}_k^{(c)T} \right) \circ \left(\mathcal{X}_{c;i} \prod_{k=1}^{M-1} \times_k \mathbf{W}_k^{(c)T} \right); (M)(M) \right] \right] \\ &\qquad = \mathcal{I}, \end{aligned} \tag{9}$$

where $\mathbf{W}_k^{(c)}|_{k=1}^{M-1}$ denote k-th projection matrix on each of $M-1$ modes respectively, and only retain H_k projection directions corresponding to $H_k/2$ largest eigenvalues and $H_k/2$ smallest eigenvalues.

This objective functions can be further interpreted as that by projecting the training samples $\mathcal{X}_{c;i}$ to $\mathbf{W}_k^{(c)}|_{k=1}^{M-1}$ on each of $M-1$ modes, the averaged covariance tensor, defined in Eq.(5), of c-th class data is superdiagonal tensor and the averaged covariance tensor of all classes data is unit superdiagonal tensor. Therefore, CTDA obtains projection directions on multi-dimensions, which can maximize variance for one class and at the same time minimize variance for the other class.

The problem defined in Eq.(9) does not have a closed form solution, so we choose to use the alternating projection method, which is an iterative procedure, to obtain a numerical solution. Therefore, Eq.(9) is decomposed into $M-1$ different optimization subproblems, as follows,

$$\frac{1}{n_c}\sum_{i=1}^{n_c}\left[\left[\left(\mathcal{X}_{c;i}\prod_{k=1}^{M-1}\times_k \mathbf{W}_k^{(c)^T}\right)\circ\left(\mathcal{X}_{c;i}\prod_{k=1}^{M-1}\times_k \mathbf{W}_k^{(c)^T}\right);(\bar{l})(\bar{l})\right]\right] = \mathbf{D}_l,$$
$$\sum_{c=1}^{C}\sum_{i=1}^{n_c}\frac{1}{n_c}\left[\left[\left(\mathcal{X}_{c;i}\prod_{k=1}^{M-1}\times_k \mathbf{W}_k^{(c)^T}\right)\circ\left(\mathcal{X}_{c;i}\prod_{k=1}^{M-1}\times_k \mathbf{W}_k^{(c)^T}\right);(\bar{l})(\bar{l})\right]\right] = \mathbf{I}_l, \quad (10)$$

where $l = 1,\ldots,M-1$ and $\mathcal{D} = \mathbf{D}_1 \circ \mathbf{D}_2 \circ \ldots \circ \mathbf{D}_{M-1}, \mathcal{I} = \mathbf{I}_1 \circ \mathbf{I}_2 \circ \ldots \circ \mathbf{I}_{M-1}$. This can be interpreted as that the tensor data $\mathcal{X}_{c;i}$ filtered by matrices $\mathbf{W}_k^{(c)}|_{k=1}^{M-1}$ on $M-1$ modes respectively and matricized on each l-th mode would be diagonal matrices of $\mathbf{D}_l, \mathbf{I}_l \in \mathbb{R}^{H_l \times H_l}$.

To simplify Eq.(10), we define $\mathbf{U}_l^{(c)}$ as

$$\frac{1}{n_c}\sum_{i=1}^{n_c}\left[\mathrm{mat}_l\left(\mathcal{X}_{c;i}\prod_{k=1;k\neq l}^{M-1}\times_k \mathbf{W}_k^{(c)^T}\right)\mathrm{mat}_l^T\left(\mathcal{X}_{c;i}\prod_{k=1;k\neq l}^{M-1}\times_k \mathbf{W}_k^{(c)^T}\right)\right], \quad (11)$$

and define $\mathbf{T}_l$ as

$$\sum_{c=1}^{C}\sum_{i=1}^{n_c}\frac{1}{n_c}\left[\mathrm{mat}_l\left(\mathcal{X}_{c;i}\prod_{k=1;k\neq l}^{M-1}\times_k \mathbf{W}_k^{(c)^T}\right)\mathrm{mat}_l^T\left(\mathcal{X}_{c;i}\prod_{k=1;k\neq l}^{M-1}\times_k \mathbf{W}_k^{(c)^T}\right)\right]. \quad (12)$$

Therefore, Eq.(10) are simplified as

$$\mathbf{W}_l^{(c)^T}\mathbf{U}_l^{(c)}\mathbf{W}_l^{(c)} = \mathbf{D}_l, \quad \mathbf{W}_l^{(c)^T}\mathbf{T}_l\mathbf{W}_l^{(c)} = \mathbf{I}_l, \quad (13)$$

where $l \in [1:M-1], c \in [1:C]$, $\mathbf{T}_l$ is equal to $\sum_{c=1}^{C}\mathbf{U}_l^{(c)}$, and $\mathbf{W}_l^{(c)}$ denotes projection matrix on l-th mode for c-th class. Thus, the CDTA problem is equivalent to $M-1$ sub-problems which can be solved by two step PCA method. Similarly, we can combine the $\mathbf{W}_l^{(c)}$ corresponding to each c class as:

$$\mathbf{W}_l = [\mathbf{W}_l^{(1)},\ldots,\mathbf{W}_l^{(C)}], \quad l = 1,\ldots,M-1. \quad (14)$$

Therefore, for the M order training tensors $\mathcal{X}$, CTDA would obtain $M-1$ optimal projection matrices $\mathbf{W}_l|_{l=1}^{M-1}$ by solving the $M-1$ alternative sub-problems defined in Eq.(13). Unfortunately, the matrices $\mathbf{U}_l^{(c)}|_{c=1}^{C}and\mathbf{T}_l$ are not fixed while depend on $\mathbf{W}_k|_{k=1;k\neq l}^{M-1}$. Therefore, the projection matrices $\mathbf{W}_l|_{l=1}^{M-1}$ can not be computed independently. A solution for that is to use an iterative procedure for simultaneous optimization of $\mathbf{W}_l|_{l=1}^{M-1}$.

Once we obtain the projection matrices based on CTDA, the M order tensor data $\mathcal{X}$ could be projected on them defined as

$$\mathcal{Z} = \mathcal{X}\prod_{l=1}^{M-1}\times_l \mathbf{W}_l{}^T. \quad (15)$$

Hence, the $M-1$ projection matrices can maximize mode-M variance for one class and keep the sum of mode-M variance constant, i.e., minimize mode-M variance for the others. The projected tensors $\mathcal{Z}$ has maximum differentiation with the variances on M-th mode.

The feature vector of tensor data $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times \ldots \times N_M}$ used for classification is composed of the $H_1 \times \ldots \times H_{M-1}$ variances normalized by the total variance of the projections retained, and log-transformed,

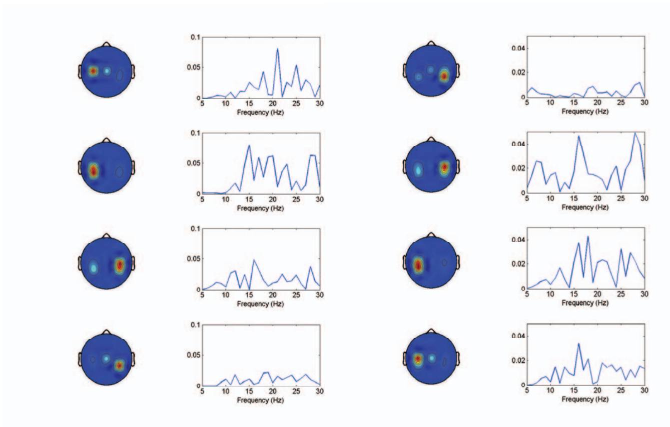
$$\mathbf{f} = \log\left\{\frac{\mathrm{diag}\left[\mathrm{mat}_M^T(\mathcal{Z})\,\mathrm{mat}_M(\mathcal{Z})\right]}{\mathrm{tr}\left[\mathrm{mat}_M^T(\mathcal{Z})\,\mathrm{mat}_M(\mathcal{Z})\right]}\right\}. \quad (16)$$

The transformation to logarithmic values is done in order to make the distribution of the elements in $\mathbf{f}$ normal. The feature vectors from the training data are used to estimate the parameters of a classifier which are used to classify new data using the projection matrices obtained from the training data.

4. EXPERIMENTAL RESULTS

In our application, EEG signals with only 5 electrodes (i.e., C3, Cp3, Cz, Cp4, C4) over the motor cortex were recorded from the scalp at a sampling rate of 250Hz for 2 and 3 classes MI-based BCI experiments. The EEG are transformed using a Morlet continuous wavelet transform (CWT) with center frequency $\omega_c = 1$ and bandwidth parameter $\omega_b = 2$. Thus, we obtain EEG tensor representation $\mathcal{X} \in \mathbb{R}^{N_d \times N_f \times N_t}$ which is the three-way time-varying EEG wavelet coefficients array, where N_d, N_f, N_t are the number of channels, steps of frequency, and time points, respectively. We apply the original CSP algorithm and the proposed CTDA algorithm for feature extraction and linear support vector machines (SVM) for classification. The frequency band of 5-30Hz is thus adopted for both the band filter in CSP algorithm and establishing EEG tensor by CWT time-frequency transform in CTDA algorithm.

. For further illustrations of the proposed method, we will pick one specific dataset of subject S1 to visualize the projection patterns on each mode which are obtained by CTDA method (see Fig.2). There are two figures which denote the projection tensors for two classes, i.e., left and right hand MI respectively. In each figure, the two upper rows show largest projections on each mode, i.e., spatial and frequency domains while the two lower rows show smallest projections. Each row represents one projection tensor and the spatial and frequency projections are shown in left column and right column respectively. By simultaneous optimization on multi-way tensor, we obtain optimal spatial filters and frequency combinations which contain the most discriminative information. In Fig.2(a), compared with classes of right hand, the largest variance of left class mainly focuses on left scalp map and a large peak around 21Hz and 15Hz, which demonstrates the ERS phenomena with specific spatial and frequency domains. Meanwhile, the smallest variance of left hand class, i.e., ERD phenomena, mainly focuses on right area and frequency band of 10-12Hz and 15-20Hz when compared with right hand, which demonstrates the ERD and ERS have not only different spatial distributions but also have a slight variability across frequency bins. Similarly, Fig.2(b) also illustrates the variability across frequency bins of ERS and ERD for right hand class. Furthermore, it is clearly shown that the ERD/ERS of different classes have distinct spatial and frequency distributions simultaneously. This indicates that the frequency information are not only subject-dependent but also class-dependent. Hence, these frequency patterns can provide further discriminative information which can not be obtained by only spatial filters. As expected, the CTDA can find the most discrimination spatial filters and frequency bands simultaneously. So instead of having a spatial projection onto a broad band (5-30Hz) signal as a solution given by the CSP, the CTDA can split information furthermore by projecting onto multi-frequency signals of the same local origin, stemming from different sub-bands, such that each projection fulfills the optimization criterion of maximizing the variance for one class, while having minimal variance for the sum of all classes. Summarizing, this yields an improved spatio-frequency resolution of the discriminative signals. Therefore, by combining the tensor representations of EEG and CTDA, more discriminative information hidden in raw signals can be obtained automatically by learning optimal projections on multi-dimensions simultaneously.



(a) Left hand (b) Right hand

Fig. 2. Optimized projection matrices on each mode of EEG tensor for each class respectively, i.e., (a) left hand class; (b) right hand class. Left column represents projection matrices on spatial modality by scalp map and right column represents projection matrices on frequency modality.

The 10×10-fold cross validation results for three subjects with two classes MI (i.e., left vs. right) and three classes case (i.e., left, right and foot) are presented in Table.1. Compared with matrix-based CSP algorithm, the tensor-based CTDA demonstrates improved classification performance.

As compared with CSP, the CTDA helps to reduce the number of parameters needed to model the data. For example, when a tensor $\mathcal{X}$ has the size $N_1 \times \ldots \times N_M$, we need to estimate the projection matrix $\mathbf{W}$ with the size $N_1 \ldots N_{M-1} \times H$ by vectorization operation and CSP, but we only need to estimate the projection matrices $\mathbf{W}_k|_{k=1}^{M-1}$ with the corresponding size $N_k \times H_k, k = 1, \ldots, M-1$ in CTDA. The advantage is the number of the parameters in CTDA is much less than that of CSP. Furthermore, when the number of the training samples is limited, the vectorization operation always leads to the under sample problem. That is, for a small training set, CSP will over-fit the data. Furthermore, the vectorization of a tensor makes it hard to keep track of the information in spatial constraints. Hence, when the number of the training samples is limited, CTDA performs better than CSP.

In our experiments, only three-way tensors are used to analyze EEG signals, thus, CTDA would learn several two-way tensor projections. To further explore the advantages of CTDA, the EEG should be represented as tensors more than three order and hence more detailed discriminative information will be extracted from multiaspect.

Table 1. Classification Accuracies ± Standard Deviation (%) of 2 classes (left Vs. Right) and 3 classes (left, right and foot) MI EEG

Subjects	2 classes		3 classes	
	CSP	CTDA	CSP	CTDA
S1	94.66±0.84	95.78±0.75	88.95±1.60	91.80±0.57
S2	88.50±0.70	92.50±1.18	83.11±1.39	87.56±0.75
S3	88.00±1.00	94.50±0.67	84.93±0.95	88.33±0.35

5. CONCLUSIONS

This paper focuses on the tensor representation and tensor-based feature extraction method for EEG signals. According to the idea of CSP algorithm, we propose a new generalization of CSP algorithm in the tensor case, called CTDA, by using tensor based subspace learning. Different from traditional CSP, the proposed algorithm is performed in the tensor space rather than the vector space. Experimental analysis for classification of MI based EEG demonstrates the better performance of proposed algorithm and the advantages of tensor analysis methods when applied in BCI.

6. ACKNOWLEDGMENT

The work was supported by the Science and Technology Commission of Shanghai Municipality (Grant No. 08511501700), the National High-Tech Research Program of China (Grant No. 2006AA01Z125) and the National Natural Science Foundation of China (Grant No. 60775007).

7. REFERENCES

[1] A.K. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis with Applications in the Chemical Sciences*, Wiley, 2004.

[2] M. Heiler and C. Schnorr, "Controlling Sparseness in Non-negative Tensor Factorization," *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006: Proceedings*, 2006.

[3] A. Cichocki, R. Zdunek, S. Choi, R. Plemmons, and S. Amari, "Non-Negative Tensor Factorization using Alpha and Beta Divergences," *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 3, 2007.

[4] M. Mørup, L.K. Hansen, C.S. Herrmann, J. Parnas, and S.M. Arnfred, "Parallel Factor Analysis as an exploratory tool for wavelet transformed event-related EEG," *Neuroimage*, vol. 29, no. 3, pp. 938–947, 2006.

[5] F. Miwakeichi, E. Martínez-Montes, P.A. Valdés-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchi, "Decomposing EEG data into space–time–frequency components using Parallel Factor Analysis," *Neuroimage*, vol. 22, no. 3, pp. 1035–1045, 2004.

[6] H.Ramoser, J.Müller-Gerking, and G.Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement," vol. 8, no. 40, pp. 441–446, Dec 2000.

[7] Y. Wang, Z. Zhang, Y. Li, X. Gao, S. Gao, and F. Yang, "BCI competition 2003-data set IV: An algorithm based on CSSD and FDA for classifying single-trial EEG," *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 6, pp. 1081–1086, 2004.

[8] D. Tao, X. Li, X. Wu, and S.J. Maybank, "General Tensor Discriminant Analysis and Gabor Features for Gait Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1700–1715, 2007.

[9] X. He, D. Cai, and P. Niyogi, "Tensor subspace analysis," *Advances in Neural Information Processing Systems*, vol. 18, 2005.