# Subspace-Constrained Multilinear Discriminant Analysis for ERP-based Brain Computer Interface Classification

Hiroshi Higashi*‡, Tomasz M. Rutkowski†‡, Toshihisa Tanaka‡§, and Yuichi Tanaka‡
* Toyohashi University of Technology, Aichi, Japan.
E-mail: higashi@tut.jp Tel: +81-532-44-6779
† University of Tsukuba, Ibaraki, Japan.
E-mail: tomek@tara.tsukuba.ac.jp Tel: +81-29-853-6261
‡ Tokyo University of Agriculture and Technology, Tokyo, Japan.
E-mail: tanakat@cc.tuat.ac.jp, ytnk@cc.tuat.ac.jp Tel: +81-42-388-7123, +81-42-388-7150
§ RIKEN Brain Science Institute, Saitama, Japan.

***Abstract*—Classification in brain computer interfaces (BCIs) frequently suffers from small sample problem which leads an ill-posed problem and overfitting/overtraining. To avoid and reduce the problems, we propose a multilinear formed linear discriminant analysis with constraints which are derived from other datasets. The proposed method prevents the ill-posed problem by reducing variables and improves robustness by transferring information from other datasets. The experimental results show that the proposed method improves classification accuracy in event-related potential-based BCIs.**

## I. Introduction

Decoding brain activity from brain signals is an important and challenging technology [1]. One of the applications of the brain decoding is a brain computer/machine interfaces (BCIs/BMIs). The BCI is an interface which uses brain signals and connects a human brain and an external device. A brain activity evoked by a certain task is assigned to the command of the device. A user of the interface performs the assigned tasks to generate the commands. The tasks inducing the brain activities are not limited to muscular movements. Certain mental tasks such as turning attention to external stimuli and imaging of something can drive BCIs [2], [3]. Therefore, the BCIs provide non-muscular communication and control channel for conveying messages and commands to the external world [3]–[5].

For acquisition of brain signals for the BCI, invasive and noninvasive methods are used [6]. The invasive methods need surgery installing electrodes on a cortex or a cerebral ventricle and measure electrical activities of brain neurons. The invasive methods can measure the brain activities with noise much less than the noninvasive methods [1]. On the other hand, the noninvasive methods do not require such medical surgeries. The noninvasive methods are considered that they impose a less loads on participants [7]. Because of this reason, noninvasively measured data such as EEG [8], magnetoencephalogram (MEG) [9], [10], functional magnetic resonance imaging (fMRI) [11], and near-infrared spectroscopy [12] are widely used to the BCI research. Among them, because of its highest temporal resolution [13], EEG is considered a suitable measurement device for the BCIs that require a quick response and a realtime processing [14], [15]. Moreover, the devices for recording EEG are low cost and relatively smaller than the other devices. Although EEG is considered to be practical for the BCI because of the above reasons, decreases of desired components, high noise, and low spatial resolution are the significant problems to be solved for the BCIs. By some techniques in signal processing, pattern recognition, and machine learning, several BCI systems with EEG (EEG-based BCI) such as inputing of letters [16], [17] and controlling an object in a monitor [18], [19], a wheel chair [20], and a robot [14], [21] have been developed.

For the EEG-based BCIs, multi-channel recording systems are widely used for capturing spatial features of brain activities related to the assigned task. As many channels/electrodes are used and spatial dense recoding is demanded, the size of the signals is increasing [22]. On the other hand, because the brain pattern depends on individuals and measurement environments, the BCI system needs to tune for individuals (user) and environments (calibration). However, it is difficult to obtain enough number of the samples (trials) for calibration of the BCI systems in some case. The recording for the calibration is time-consuming and the user can be tired. In the case that the size (dimension) of the data is large and the number of the samples is small, ill-pose problems and overfitting/overtraining can happen [23]. To avoid such problems, unsupervised dimensional reduction techniques such as principal component analysis (PCA) [23], regularizations [24]–[28], and calibrations with data from other users [22], [29]

This work was supported by JSPS KAKENHI (Grant Number 15K21079 and 26240043, and 15H04002), and Tateishi Science and Technology Foundation.
Y. Tanaka is supported in part by MEXT Tenure Track Promotion Program.

are useful. Moreover, multilinear/multiway signal processing techniques have been proposed [30], [31]. The multilinear signal processing can process a tensor having more than three modes (e.g. a vector is a one-mode tensor and a matrix is a two-mode tensor). The structures of the recorded data and/or converted data can be kept by the multilinear approach. Because the structures are given as an additional information and work as regularization, the multilinear approaches can perform well for EEG classification [32], [33]. Additionally, if an alternating optimization for each mode is applied [31], [34]–[37], the number of the variables to be optimized can be small for each mode. Therefore, it can avoid the ill-posed problems.

This paper proposes a robust multilinear discriminant analysis (MLDA) [38] by modifying the direct general tensor discriminant analysis (DGTDA) algorithm [34] for single trial event-related potential (ERP)-based BCI classification. The classification problems for the ERP-based BCIs, in which a linear discriminant analysis (LDA) classifier is frequently used [39], [40], is easy to be ill-posed because the dimension of an extracted feature vector can be large. For address this problem, we employ a multilinear form of the LDA classifier [34]. Moreover, we develop constraints for the parameter space to improve the robustness [24], [41] for the classifier. The constraints are derived from other datasets (observed in other experiment participants). We call the proposed method subspace-constrained DGTDA (SC-DGTDA). The experimental results of offline classification show that SC-DGTDA improves the classification accuracy in the ERP-based BCIs.

## II. Multilinear Algebra and Multilinear Discriminant Analysis – Review

### A. Basics for Multilinear Algebra

In this paper, lower-case bold-face characters represent vectors (e.g. $\boldsymbol{x}$ and $\boldsymbol{y}$), upper-case bold-face characters represent matrices (e.g. $\boldsymbol{X}$ and $\boldsymbol{Y}$), upper-case calligraphic capitals represent tensors (e.g. $\mathcal{X}$ and $\mathcal{Y}$). A tensor is a multidimensional array represented as $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ which is a tensor with $N$ modes and the dimension $I_n$ for its $n$th mode.

The $i$th element of a vector $\boldsymbol{x}$ is given by $[\boldsymbol{x}]_i$. The entry in the $i$th row and $j$th column of a matrix $\boldsymbol{X}$ is given by $[\boldsymbol{X}]_{i,j}$. The entry in the $i_n$th index in the $n$-modes for $n = 1, \ldots, N$ of a tensor $\mathcal{X}$ is given by $[\mathcal{X}]_{i_1,i_2,\ldots,i_N}$.

The $n$-mode vectors of $\mathcal{X}$ are defined as the $I_n$-dimensional vectors obtained from $\boldsymbol{X}$ by varying the index $i_n$ while keeping the other indices fixed. The unfolding for a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ along the $n$th mode is an operator denoted by $(\cdot)_{(n)}$ to transform the tensor to the matrix. The matrix transformed by $(\mathcal{X})_{(n)}$ is denoted by $\boldsymbol{X}_{(n)} \in \mathbb{R}^{I_n \times I_{\bar{n}}}$ where $I_{\bar{n}} = \prod_{i=1, i\neq n}^{N} I_i$. The column vectors of $\boldsymbol{X}_{(n)}$ are the $n$-mode vectors of $\mathcal{X}$. The $n$-mode product of a tensor $\mathcal{X}$ by a matrix $\boldsymbol{A} \in \mathbb{R}^{J_n \times I_n}$ is denoted by $\mathcal{X} \times_n \boldsymbol{A}$. The elements of the $n$-mode product of $\mathcal{X}$ and $\boldsymbol{A}$ are defined as

$$[\mathcal{X} \times_n \boldsymbol{A}]_{i_1,\ldots,i_{n-1},j_n,i_{n+1},\ldots,i_N} = \sum_{i_n=1}^{I_n} [\mathcal{X}]_{i_1,\ldots,i_{n-1},i_n,i_{n+1},\ldots,i_N} [\boldsymbol{A}]_{j_n,i_n} \quad (1)$$

The $n$-mode unfolding of the $n$-mode product can be obtained by

$$(\mathcal{X} \times_n \boldsymbol{A})_{(n)} = \boldsymbol{A}^\top \boldsymbol{X}_{(n)}. \quad (2)$$

For convenience, we denote

$$\mathcal{X} \prod_{n=1}^{N} \times_n \boldsymbol{A}_n = \mathcal{X} \times_1 \boldsymbol{A}_1 \times_2 \boldsymbol{A}_2 \times_3 \cdots \times_N \boldsymbol{A}_N. \quad (3)$$

The Frobenius norm of a tensor $\mathcal{X}$ is defined as

$$\|\mathcal{X}\|_F^2 = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} |[\mathcal{X}]_{i_1,i_2,\ldots,i_N}|^2. \quad (4)$$

### B. Direct General Tensor Discriminant Analysis

There are some algorithms for multilinear form of LDA [38], [42]. Most of the algorithms employ an alternating optimization for optimizing the LDA cost function that does not have a closed-form solution. In this paper, we modify DGTDA which provides a closed-form solution and propose a new method for classifying for ERP-based BCIs. We review DGTDA in this section.

Consider a problem in which an observed sample is classified to a class $\omega$ out of $N_c$ classes ($\omega \in \{\omega_1, \ldots, \omega_{N_c}\}$). The observed sample is given as an $N$-mode tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ in the problem. DGTDA [34] provides a supervised dimensional reduction by a similar idea of LDA. DGTDA finds projection matrices for each mode and reduces the dimension by

$$\mathcal{Y} = \mathcal{X} \prod_{n=1}^{N} \times_n \boldsymbol{U}_n^\top, \quad (5)$$

where $\mathcal{Y}$ is the dimensionally-reduced tensor the size of which is $I'_1 \times \cdots \times I'_N$ and $\boldsymbol{U}_n \in \mathbb{R}^{I_n \times I'_n}$ is the projection matrix for the $n$th mode. Because DGTDA aims to reduce the dimension for each mode, the dimension of the mode for the projected tensor $\mathcal{Y}$ is less than it in the original tensor $\mathcal{X}$ such that $\{I'_n \leq I_n\}_{n=1}^{N}$.

Let $\{\mathcal{X}^{(m)}, \beta^{(m)}\}_{m=1}^{N_m}$ is $N_m$ pairs of the observed sample and its class label in a given learning data, where $\mathcal{X}^{(m)} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ and $\beta^{(m)} \in \{\omega_1, \ldots, \omega_{N_c}\}$. DGTDA solves an optimization problem [34];

$$\max_{\{\boldsymbol{U}_n\}_{n=1}^{N}} \frac{\sum_{c=1}^{N_c} |\Omega_c| \|\tilde{\mathcal{M}}_{\omega_c} - \tilde{\mathcal{M}}\|_F^2}{\sum_{m=1}^{N_m} \|\mathcal{Y}^{(m)} - \tilde{\mathcal{M}}_{\beta_m}\|_F^2} \quad \text{subject to} \quad \boldsymbol{U}_n^\top \boldsymbol{U}_n = \boldsymbol{I}_{I'_n}, \quad \forall n, \quad (6)$$

where $\tilde{\mathcal{M}}$ is the mean tensor of the projected tensors defined as

$$\tilde{\mathcal{M}} = \frac{1}{N_m} \sum_{m=1}^{N_m} \mathcal{Y}^{(m)}, \quad (7)$$

$\tilde{\mathcal{M}}_{\omega_c}$ is the mean tensor of the projected tensors belonging to class $\omega_c$ defined as

$$\tilde{\mathcal{M}}_{\omega_c} = \frac{1}{|\Omega_c|} \sum_{m \in \Omega_c} \mathcal{Y}^{(m)}, \tag{8}$$

$\Omega_c$ is a set of the sample index belonging to class $\omega_c$ given by $\Omega_c = \{m' \mid \beta^{(m')} = \omega_c, m' = 1, \ldots, N_m\}$, the operator $|\cdot|$ for a set gives the number of the elements of the set, and $\boldsymbol{I}_D$ is an identity matrix of the size $D \times D$. The projections $\{\boldsymbol{U}_n\}_{n=1}^N$ are given by solving

$$\begin{aligned} \max_{\boldsymbol{U}_n} \quad & \operatorname{tr}\left(\boldsymbol{U}_n^\top (\boldsymbol{B}_n - \zeta \boldsymbol{W}_n) \boldsymbol{U}_n\right) \\ \text{subject to} \quad & \boldsymbol{U}_n^\top \boldsymbol{U}_n = \boldsymbol{I}_{I'_n}, \end{aligned} \tag{9}$$

for each mode, where the between-class scatter matrix is defined as

$$\boldsymbol{B}_n = \sum_{c=1}^{N_c} |\Omega_c| (\mathcal{M}_{\omega_c} - \mathcal{M})_{(n)} (\mathcal{M}_{\omega_c} - \mathcal{M})_{(n)}^\top, \tag{10}$$

the within-class scatter matrix is defined as

$$\boldsymbol{W}_n = \sum_{m=1}^{N_m} (\mathcal{X}^{(m)} - \mathcal{M}_{\beta^{(m)}})_{(n)} (\mathcal{X}^{(m)} - \mathcal{M}_{\beta^{(m)}})_{(n)}^\top, \tag{11}$$

$\zeta$ is the largest eigenvalue of $\boldsymbol{W}_n^{-1}\boldsymbol{B}_n$, and the operator $\operatorname{tr}(\cdot)$ gives the sum of the diagonal elements of a square matrix. The problem (9) is obtained as the singular value decomposition (SVD);

$$\bar{\boldsymbol{U}} \bar{\boldsymbol{S}} \bar{\boldsymbol{V}}^\top = \boldsymbol{B}_n - \zeta \boldsymbol{W}_n, \tag{12}$$

where $\bar{\boldsymbol{U}} \in \mathbb{R}^{I_n \times I_n}$ and $\bar{\boldsymbol{V}} \in \mathbb{R}^{I_n \times I_n}$ are unitary matrices, and $\bar{\boldsymbol{S}}$ is a diagonal matrix containing the singular values in its diagonal entries. The projection $\boldsymbol{U}_n$ is given as

$$\boldsymbol{U}_n = [\bar{\boldsymbol{u}}_{f_1}, \bar{\boldsymbol{u}}_{f_2}, \ldots, \bar{\boldsymbol{u}}_{f_{I'_n}}], \tag{13}$$

where $\bar{\boldsymbol{u}}_i$ is the $i$th column vector of $\bar{\boldsymbol{U}}$ and the $\{f_1, f_2, \ldots, f_{I'_n}\}$ are the index of the $I'_n$ largest singular values as $[\bar{\boldsymbol{S}}]_{f_1,f_1} \geq [\bar{\boldsymbol{S}}]_{f_2,f_2} \geq \cdots \geq [\bar{\boldsymbol{S}}]_{f_{I_n},f_{I_n}}$.

## III. Subspace-Constrained Direct General Tensor Discriminant Analysis for ERP classification

This section describes the proposed SC-DGTDA algorithm. It is well known as regularization that adding some information performs the preventing of overfitting and the solving of an ill-posed problem [23]. The proposed method performs adding some information by giving constraints of parameter spaces as well as regularization.

### *A. Subspace Constrained Optimization*

The proposed method constrains the parameter spaces for the column vectors in each mode projection. This idea is formulated as

$$\begin{aligned} \max_{\{\boldsymbol{U}_n\}_{n=1}^N} \quad & \frac{\sum_{c=1}^{N_c} |\Omega_c| \|\tilde{\mathcal{M}}_{\omega_c} - \tilde{\mathcal{M}}\|_F^2}{\sum_{m=1}^{N_m} \|\mathcal{Y}^{(m)} - \tilde{\mathcal{M}}_{\beta_m}\|_F^2} \\ \text{subject to} \quad & \boldsymbol{U}_n^\top \boldsymbol{U}_n = \boldsymbol{I}_{I'_n}, \quad \forall n, \\ & \boldsymbol{u}_n^{(i_n)} \in S_n, \quad \forall i_n, \forall n, \end{aligned} \tag{14}$$

where, $\boldsymbol{u}_n^{(i_n)}$ is a column vector of $\boldsymbol{U}_n$ defined as $\boldsymbol{U}_n = [\boldsymbol{u}_n^{(i_1)}, \boldsymbol{u}_n^{(i_2)}, \ldots \boldsymbol{u}_n^{(i_n)}]$ and $S_n$ is a subspace in the parameter space for the column vector of $\boldsymbol{U}$. The subspaces work to constraint the projection vector in the subspace for each mode. The design of $\{S_n\}_{n=1}^N$ is discussed in Sec. III-B.

As well as DGTDA, the optimization problem for SC-DGTDA (14) can be reduced to subproblems for each mode as

$$\begin{aligned} \max_{\boldsymbol{U}_n} \quad & \operatorname{tr}\left(\boldsymbol{U}_n^\top (\boldsymbol{B}_n - \zeta \boldsymbol{W}_n) \boldsymbol{U}_n\right) \\ \text{subject to} \quad & \boldsymbol{U}_n^\top \boldsymbol{U}_n = \boldsymbol{I}_{I'_n}, \\ & \boldsymbol{u}_n^{(i_n)} \in S_n, \quad \forall i_n. \end{aligned} \tag{15}$$

Let $\{\boldsymbol{s}_n^{(1)}, \boldsymbol{s}_n^{(2)}, \ldots, \boldsymbol{s}_n^{(D_n)}\}$ be an orthonormal basis of $S_n$ and $\boldsymbol{S}_n \in \mathbb{R}^{I_n \times D_n}$ be a matrix defined as $\boldsymbol{S}_n = [\boldsymbol{s}_n^{(1)}, \boldsymbol{s}_n^{(2)}, \ldots, \boldsymbol{s}_n^{(D_n)}]$, where $D_n$ should be more than $I'_n$ ($I_n \geq D_n \geq I'_n$) and ${\boldsymbol{s}_n^{(i)}}^\top \boldsymbol{s}_n^{(j)} = \delta_{i,j}$ for $i, j = 1, \ldots, D_n$. By using the basis, the column vectors for the projection matrix are formulated as a linear combination;

$$\boldsymbol{u}_n^{(i_n)} = a_1 \boldsymbol{s}_n^{(1)} + a_2 \boldsymbol{s}_n^{(2)} + \cdots + a_{D_n} \boldsymbol{s}_n^{(D_n)} = \boldsymbol{S}_n \boldsymbol{a} \tag{16}$$

where $\{a_d\}_{d=1}^{D_n}$ are the coefficients for the linear combination and $\boldsymbol{a}$ is the vector defined as $\boldsymbol{a} = [a_1, \ldots, a_{D_n}]^\top$. By this formulation, the projections can be represented by

$$\boldsymbol{U}_n = \boldsymbol{S}_n \boldsymbol{A}_n, \tag{17}$$

where the $i$th row vector of $\boldsymbol{A}_n \in \mathbb{R}^{D_n \times I'_n}$ have the coefficients of the linear combination with the $i$th column vector of $\boldsymbol{U}_n$. Then, the optimization problem (15) can be translated to a problem of finding $\boldsymbol{A}$ by change of variables. The problem of finding $\boldsymbol{A}$ is formulated as

$$\begin{aligned} \max_{\boldsymbol{A}_n} \quad & \operatorname{tr}\left(\boldsymbol{A}_n^\top \boldsymbol{S}_n^\top (\boldsymbol{B}_n - \zeta \boldsymbol{W}_n) \boldsymbol{S}_n \boldsymbol{A}_n\right) \\ \text{subject to} \quad & \boldsymbol{A}_n^\top \boldsymbol{A}_n = \boldsymbol{I}_{I'_n}. \end{aligned} \tag{18}$$

The orthonormal constraints $\boldsymbol{U}_n^\top \boldsymbol{U}_n = \boldsymbol{I}_{I'_n}$ can be reduced to $\boldsymbol{A}_n^\top \boldsymbol{A}_n = \boldsymbol{I}$ because $\boldsymbol{U}_n^\top \boldsymbol{U}_n = \boldsymbol{A}_n^\top \boldsymbol{S}_n^\top \boldsymbol{S}_n \boldsymbol{A} = \boldsymbol{A}_n^\top \boldsymbol{I}_{D_n} \boldsymbol{A}_n = \boldsymbol{A}_n^\top \boldsymbol{A}_n$. As well as DGTDA, the solution of the optimization problem (18) is given by SVD;

$$\bar{\boldsymbol{A}} \bar{\boldsymbol{S}} \bar{\boldsymbol{V}}^\top = \boldsymbol{S}_n^\top (\boldsymbol{B}_n - \zeta \boldsymbol{W}_n) \boldsymbol{S}_n, \tag{19}$$

where $\bar{\boldsymbol{A}} \in \mathbb{R}^{D_n \times D_n}$ and $\bar{\boldsymbol{V}} \in \mathbb{R}^{D_n \times D_n}$ are unitary matrices. The coefficient matrix $\boldsymbol{A}_n$ is given as

$$\boldsymbol{A}_n = [\bar{\boldsymbol{a}}_{f_1}, \bar{\boldsymbol{a}}_{f_2}, \ldots, \bar{\boldsymbol{a}}_{f_{I'_n}}], \tag{20}$$

where $\bar{\boldsymbol{a}}_i$ is the $i$th column vector of $\bar{\boldsymbol{A}}$ and the $\{f_1, f_2, \ldots, f_{I'_n}\}$ are the index of the $I'_D$ largest singular values as $[\bar{\boldsymbol{S}}]_{f_1,f_1} \geq [\bar{\boldsymbol{S}}]_{f_2,f_2} \geq \cdots \geq [\bar{\boldsymbol{S}}]_{f_{D_n},f_{D_n}}$.

### *B. Design of Subspaces for ERP Classification*

As the subspaces $\{S_n\}_{n=1}^N$ for SC-DGTDA, we propose the subspace design method using datasets from other participants [22], [25]. Let $\{\mathcal{X}^{(m)} \in \mathbb{R}^{I_1 \times \cdots \times I_N}\}_{m=1}^{N_o}$ be feature tensors from other dataset. For the subspace of the $n$th mode, we utilize the mean correlation of the $I_n$ variables in the $n$th mode. The correlation matrix along the $n$th mode was defined

$$[\boldsymbol{C}_n]_{i,j} = \sum_{m=1}^{N_o} \frac{\sum_{k=1}^{\bar{I}_n} [\tilde{\boldsymbol{X}}_{(n)}^{(m)}]_{i,k} [\tilde{\boldsymbol{X}}_{(n)}^{(m)}]_{j,k}}{\sqrt{\sum_{k=1}^{\bar{I}_n} \left([\tilde{\boldsymbol{X}}_{(n)}^{(m)}]_{i,k}\right)^2} \sqrt{\sum_{k=1}^{\bar{I}_n} \left([\tilde{\boldsymbol{X}}_{(n)}^{(m)}]_{j,k}\right)^2}}, \quad (21)$$

where $\bar{I}_n$ is defined as $\bar{I}_n = \left(\prod_{l=1}^{N} I_l\right)/I_n$, $\tilde{\boldsymbol{X}}_{(n)}^{(m)}$ is the centered unfolded matrix defined as

$$[\tilde{\boldsymbol{X}}_{(n)}^{(m)}]_{i,j} = [\boldsymbol{X}_{(n)}^{(m)}]_{i,j} - \sum_{k=1}^{\bar{I}_n} [\boldsymbol{X}_{(n)}^{(m)}]_{i,k}. \quad (22)$$

We consider that the correlation matrix $\boldsymbol{C}_n$ is an adjacent matrix on the graph which connects the $I_n$ variables in the $n$th mode [43]. We utilize this graph as the additional information for the parameter space. Employing the method proposed in [28], we obtain a basis of the parameter space that is smooth across the graph by the following way. The graph Laplacian is defined as

$$\boldsymbol{L}_n = \boldsymbol{D}_n - \boldsymbol{C}_n, \quad (23)$$

where $\boldsymbol{D}_n$ is a diagonal matrix whose diagonal elements are given by $[\boldsymbol{D}_n]_{i,i} = \sum_{k=1}^{I_n} [\boldsymbol{C}_n]_{i,k}$. The orthonormal basis of $S_n$ satisfies

$$\boldsymbol{L}_n \boldsymbol{s}_n^{(i)} = \lambda_i \boldsymbol{s}_n^{(i)}. \quad (24)$$

The indexes of the eigenvalues and the corresponding eigenvectors are decided in such a way that $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{I_n}$.

## IV. Experiment with BCI Datasets

We conducted classification experiment of ERP-based BCIs to evaluate the classification performance of the proposed SC-DGTDA classifier. The SC-DGTDA classifier was compared with the LDA classifier with PCA dimensional reduction.

### A. Data Description

We used some types of datasets of P300-based BCIs. Although the P300-based BCIs provides multi-class commands and need to solve multi-class classification problem, we reduce the classification problem into two-class single-trial classification problem (a trial is belonging to target or non-target class) for simple discussion.

*1) Tactile-force Stimulus:* In the dataset, a P300-based BCI were served with tactile-force stimulus delivered to the hand holding a force-feedback joystick. For the detailed information about this dataset, readers can refer to [44]. The EEG signals were recorded with 16 electrodes (Cz, CPz, P3, P4, C3, C4, CP5, CP6, P1, P2, POz, C1, C2, FC1, FC2, and FCz) and 512 Hz of the sampling frequency. In this paper, the dataset is called *DATASET-T*.

*2) Audio/Visual/Audio-visual Stimulus:* In the dataset, a P300-based BCI were served with auditory/visual/audio-visual stimulus. For the detailed information about this dataset, readers can refer to [45]. The EEG signals were recorded with 16 electrodes (Cz, CPz, POz, Pz, P1, P2, C3, C4, O1, O2, T7, T8, P3, P4, F3, and F4) and 512 Hz of the sampling frequency. In this paper, the datasets of the audio, visual, audio-visual stimuli are called *DATASET-A*, *DATASET-V*, and *DATASET-AV*, respectively. Figure 1 depicts the grandaveraged ERP waveforms of *DATASET-V*. We can observe the difference in the amplitudes of the P300 components between the target class and the non-target class.

### B. Classification Procedure

For removing trials contaminated by components related to eye-blink or other muscles, the trials the signals in which had over 80 $\mu$V amplitude were removed. The signal for a single trial was drawn from 0–800 ms interval after the stimulus presented. The Butterworth bandpass filter whose passband was 0.1–31 Hz were applied. We furthermore downsampled them to 64 Hz.

For the LDA classifier, we vectorized the signals that contained in a matrix. Because the dimension of the vectors exceeded the number of the learning samples and it led an ill-posed problem, we applied PCA before obtaining the LDA projection. The dimension of the feature vector reduced by the PCA projection is a parameter for the LDA classifier. For the SC-DGTDA classifier, we did not vectorized the feature matrix. The subspaces for SC-DGTDA were given by the datasets of the other participants. The dimensions of the subspaces are parameters for the SC-DGTDA classifier.

The LDA or SC-DGTDA classifier projects the feature vector/matrix of a trial into a one-dimensional space. This means that the parameters controlling the projected tensor dimension of SC-DGTDA were set to $I'_1 = I'_2 = 1$. And then, the projected samples is classified by the Bayes' rule [46] using a Gaussian distribution;

$$\hat{c} = \arg \max_{c \in \{\text{target,non-target}\}} \mathcal{N}(y \mid m_c, \sigma_c^2) p(c), \quad (25)$$

where $y$ is a projected sample, $m_c$ and $\sigma_c^2$ are estimated as the sample mean and the sample variance of the projected features in the learning samples belonging to class $c$, respectively, and $p(c)$ is estimated as $p(\text{target}) = p(\text{non-target}) = 0.5$.

Because the number of the trials was different for each participant, we randomly selected 20 or 200 samples (10 or 100 samples per class) for the learning samples and remaining samples were used to evaluate the classification accuracy. We repeated this procedure 100 times and obtained the classification accuracy averaged over the evaluations. The parameters for each method were tuned with the learning samples for each evaluation. For the tuning, a leave-one-out cross validation was adopted. The reduced dimension in the LDA classifier was chosen out of $\{1, 2, \ldots, 100\}$. The dimension of the subspace for the spatial domain in the SC-DGTDA classifier was chosen out of $\{1, 2, \ldots, 10\}$. The dimension of the subspace for the

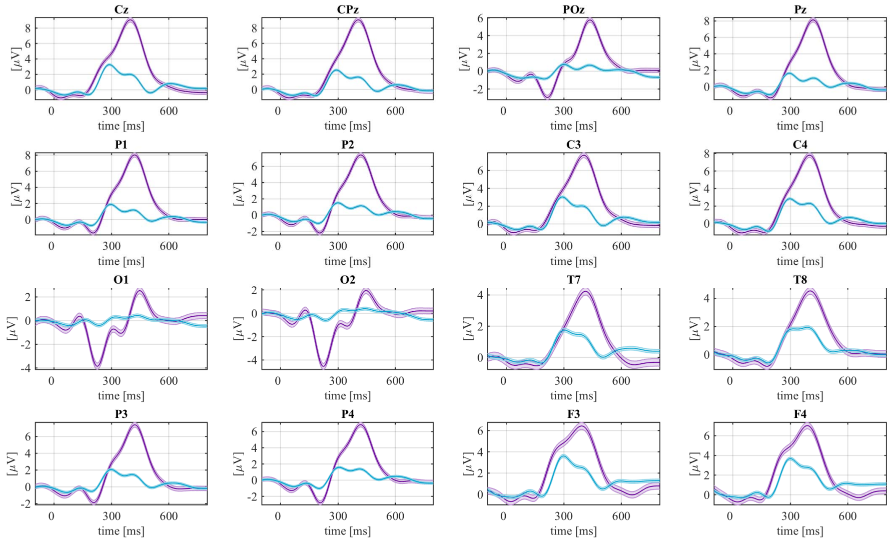

Fig. 1. The grandaveraged waveforms of *DATASET-V*. The purple lines are the grandaveraged waveforms belonging to the class target. The blue lines are the grandaveraged waveforms belonging to the class non-target. The regions the colors of which are the same of the plot lines show the standard error at each time point over the participants.

TABLE I
CLASSIFICATION ACCURACY FOR *DATASET-T*.

| | Method | | | |
|---|---|---|---|---|
| | $N_L = 20$ | | $N_L = 200$ | |
| Participant | LDA | SC-DGTDA | LDA | SC-DGTDA |
| 1 | 52.22 | 53.47 | 54.00 | 55.56 |
| 2 | 50.62 | 54.16 | 53.03 | 56.47 |
| 3 | 55.81 | 56.26 | 58.85 | 63.24 |
| 4 | 52.05 | 54.06 | 52.59 | 56.76 |
| 5 | 53.64 | 54.62 | 60.48 | 58.19 |
| Ave. | 52.87 | 54.51 | 55.79 | 58.04 |

time domain in the SC-DGTDA classifier was chosen out of $\{1, 2, \ldots, 20\}$.

### C. Result

The classification accuracies are shown in Tables I, II, III, and IV. $N_L$ denotes the number of the learning samples. The proposed SC-DGTDA can improve classification accuracy.

TABLE II
CLASSIFICATION ACCURACY FOR *DATASET-A*.

| | Method | | | |
|---|---|---|---|---|
| | $N_L = 20$ | | $N_L = 200$ | |
| Participant | LDA | SC-DGTDA | LDA | SC-DGTDA |
| 1 | 70.08 | 74.05 | 73.46 | 78.15 |
| 2 | 62.44 | 66.86 | 67.50 | 72.22 |
| 3 | 71.49 | 72.16 | 76.19 | 75.60 |
| 4 | 56.90 | 58.41 | 60.16 | 60.16 |
| 5 | 56.53 | 60.28 | 57.69 | 64.19 |
| 6 | 54.48 | 56.97 | 56.00 | 57.79 |
| 7 | 69.73 | 68.90 | 74.35 | 76.96 |
| 8 | 61.06 | 58.41 | 63.33 | 61.43 |
| 9 | 63.91 | 65.17 | 66.22 | 68.65 |
| 10 | 58.20 | 61.24 | 59.10 | 64.10 |
| 11 | 55.27 | 57.26 | 56.04 | 58.44 |
| 12 | 53.19 | 55.60 | 54.40 | 56.04 |
| 13 | 63.78 | 65.40 | 69.17 | 68.33 |
| 14 | 54.06 | 54.71 | 54.77 | 56.48 |
| 15 | 58.25 | 57.52 | 60.35 | 62.30 |
| 16 | 55.18 | 56.52 | 57.10 | 59.03 |
| Ave. | 60.28 | 61.84 | 62.86 | 64.93 |

## V. CONCLUSIONS

We develop an supervised projection algorithm with DGTDA and the subspace constraints for the parameter spaces. Our experimental results show that the proposed algorithm can improve classification accuracy in ERP-based BCIs by deriving the constraints from other participants datasets.

## REFERENCES


[1] G. Muller-Putz, R. Leeb, M. Tangermann, J. Hohne, A. Kubler, F. Cincotti, D. Mattia, R. Rupp, K.-R. Muller, and J. d. R. Millan, "Towards noninvasive hybrid brain–computer interfaces: Framework, practice, clinical application, and beyond," *Proceedings of the IEEE*, vol. 103, no. 6, pp. 926–943, 2015.

[2] B. He, S. Gao, H. Yuan, and J. Wolpaw, "BrainComputer Interfaces," in *Neural Engineering*, B. He, Ed. Springer US, 2013, pp. 87–151.

[3] B. He, B. Baxter, B. J. Edelman, C. C. Cline, and W. W. Ye, "Noninvasive brain-computer interfaces based on sensorimotor rhythms," *Proceedings of the IEEE*, vol. 103, no. 6, pp. 1–19, 2015.

[4] G. Dornhege, J. d. R. Millan, T. Hinterberger, D. McFarland, and K.-R. Muller, Eds., *Toward Brain-Computer Interfacing*. The MIT Press, 2007.

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

[6] M. A. Lebedev and M. A. L. Nicolelis, "Brain machine interfaces: past, present and future," *Trends in Neurosciences*, vol. 29, no. 9, pp. 536–546, 2006.

[7] D. J. McFarland and J. R. Wolpaw, "Brain-computer interfaces for communication and control," *Communications of the ACM*, vol. 54, no. 5, pp. 60–66, 2011.

[8] E. Niedermeyer and F. H. L. Da Silva, *Electroencephalography: Basic*

TABLE III
CLASSIFICATION ACCURACY FOR *DATASET-V*.

| | Method | | | |
|---|---|---|---|---|
| | $N_L = 20$ | | $N_L = 200$ | |
| Participant | LDA | SC-DGTDA | LDA | SC-DGTDA |
| 1 | 71.26 | 74.49 | 74.68 | 78.44 |
| 2 | 65.24 | 65.61 | 70.31 | 70.94 |
| 3 | 73.77 | 74.23 | 75.50 | 76.63 |
| 4 | 56.88 | 57.62 | 60.34 | 61.70 |
| 5 | 56.93 | 58.61 | 59.20 | 62.87 |
| 6 | 54.04 | 56.10 | 53.74 | 59.01 |
| 7 | 70.29 | 70.87 | 74.21 | 78.95 |
| 8 | 59.35 | 58.26 | 61.86 | 61.02 |
| 9 | 64.10 | 64.34 | 69.70 | 73.03 |
| 10 | 57.63 | 60.84 | 61.21 | 64.00 |
| 11 | 55.21 | 56.76 | 55.87 | 57.17 |
| 12 | 52.70 | 54.16 | 53.68 | 57.82 |
| 13 | 64.57 | 64.14 | 75.50 | 71.00 |
| 14 | 53.88 | 54.78 | 54.05 | 56.19 |
| 15 | 56.62 | 58.87 | 61.33 | 61.08 |
| 16 | 54.35 | 56.23 | 56.55 | 58.79 |
| Ave. | 60.43 | 61.62 | 63.61 | 65.54 |

TABLE IV
CLASSIFICATION ACCURACY FOR *DATASET-AV*.

| | Method | | | |
|---|---|---|---|---|
| | $N_L = 20$ | | $N_L = 200$ | |
| Participant | LDA | SC-DGTDA | LDA | SC-DGTDA |
| 1 | 71.71 | 76.50 | 75.62 | 79.63 |
| 2 | 63.08 | 66.92 | 71.43 | 72.50 |
| 3 | 71.83 | 74.41 | 75.13 | 77.11 |
| 4 | 56.29 | 58.38 | 59.09 | 60.91 |
| 5 | 55.41 | 59.10 | 61.21 | 62.89 |
| 6 | 53.14 | 56.50 | 55.29 | 58.51 |
| 7 | 68.31 | 71.54 | 72.67 | 79.33 |
| 8 | 60.95 | 58.67 | 63.82 | 62.91 |
| 9 | 66.20 | 66.71 | 68.62 | 74.14 |
| 10 | 57.85 | 61.13 | 60.74 | 63.68 |
| 11 | 55.36 | 57.83 | 55.01 | 59.67 |
| 12 | 51.73 | 55.26 | 55.78 | 56.27 |
| 13 | 65.91 | 63.33 | 73.12 | 70.00 |
| 14 | 53.46 | 56.08 | 55.13 | 57.00 |
| 15 | 60.47 | 59.30 | 61.01 | 61.90 |
| 16 | 54.71 | 56.15 | 56.67 | 60.56 |
| Ave. | 60.40 | 62.30 | 63.83 | 66.00 |


*Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins, 2005.

[9] J. Mellinger, G. Schalk, C. Braun, H. Preissl, W. Rosenstiel, N. Birbaumer, and A. Kbler, "An MEG-based brain computer interface (BCI)," *NeuroImage*, vol. 36, no. 3, pp. 581–593, 2007.

[10] K.-i. Morishige, T. Yoshioka, D. Kawawaki, N. Hiroe, M.-a. Sato, and M. Kawato, "Estimation of hyper-parameters for a hierarchical model of combined cortical and extra-brain current sources in the MEG inverse problem," *NeuroImage*, vol. 101, pp. 320–36, Nov. 2014.

[11] R. Sitaram, A. Caria, R. Veit, T. Gaber, G. Rota, A. Kuebler, and N. Birbaumer, "fMRI brain-computer interface: A tool for neuroscientific research and treatment," *Computational Intelligence and Neuroscience*, vol. 2007, pp. 1:1–1:10, 2007.

[12] S. M. Coyle, T. E. Ward, and C. M. Markham, "Brain-computer interface using a simplified functional near-infrared spectroscopy system," *Journal of Neural Engineering*, vol. 4, no. 3, p. 219, 2007.

[13] S. Sanei and J. Chambers, *EEG Signal Processing*. Wiley-Interscience, 2007.

[14] D. J. McFarland and J. R. Wolpaw, "Brain-computer interface operation of robotic and prosthetic devices," *Computer*, vol. 41, no. 10, pp. 52–56, 2008.

[15] C. Zhang, Y. Kimura, H. Higashi, and T. Tanaka, "A simple platform of brain-controlled mobile robot and its implementation by SSVEP," in *Proceedings of The 2012 International Joint Conference on Neural Networks (IJCNN)*, 2012, pp. 1–7.

[16] R. Scherer, G. R. Muller, C. Neuper, B. Graimann, and G. Pfurtscheller, "An asynchronously controlled EEG-based virtual keyboard: Improvement of the spelling rate," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 979–984, 2004.

[17] G. Townsend, B. K. LaPallo, C. B. Boulay, D. J. Krusienski, G. E. Frye, C. K. Hauser, N. E. Schwartz, T. M. Vaughan, J. R. Wolpaw, and E. W. Sellers, "A novel P300-based brain-computer interface stimulus presentation paradigm: Moving beyond rows and columns," *Clinical Neurophysiology*, vol. 121, no. 7, pp. 1109–1120, 2010.

[18] F. Galán, M. Nuttin, E. Lew, P. W. Ferrez, G. Vanacker, J. Philips, and J. d. R. Millán, "A brain-actuated wheelchair: Asynchronous and non-invasive Brain-computer interfaces for continuous control of robots," *Clinical Neurophysiology*, vol. 119, no. 9, pp. 2159–2169, 2008.

[19] R. Leeb, D. Friedman, G. R. Müller-Putz, R. Scherer, M. Slater, and G. Pfurtscheller, "Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with a tetraplegic," *Computational Intelligence and Neuroscience*, vol. 2007, pp. 7:1—-7:12, 2007.

[20] K. Choi and A. Cichocki, "Control of a wheelchair by motor imagery in real time," in *Intelligent Data Engineering and Automated Learning 2008*, 2008, pp. 330–337.

[21] C. J. B. Rao, P. Shenoy, R. Chalodhorn, and R. P. N, "Control of a humanoid robot by a noninvasive brain-computer interface in humans," *Journal of Neural Engineering*, vol. 5, no. 2, p. 214, 2008.

[22] S. Fazli, S. Dahne, W. Samek, F. Bieszmann, and K.-R. Muller, "Learning from more than one data source: data fusion techniques for sensorimotor rhythm-based brain–computer interfaces," *Proceedings of the IEEE*, vol. 103, no. 6, pp. 891–906, 2015.

[23] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[24] B. Blankertz, M. Kawanabe, R. Tomioka, F. Hohlefeld, V. Nikulin, and K. R. Müller, "Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing," *Advances in Neural Information Processing Systems*, vol. 20, pp. 113–120, 2008.

[25] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces," *Proceedings of the IEEE*, vol. 103, no. 6, pp. 871–890, 2015.

[26] H. Lu, H.-L. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial pattern with aggregation for EEG classification in small-sample setting," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 12, pp. 2936–2946, 2010.

[27] H. Higashi and T. Tanaka, "Regularization using similarities of signals observed in nearby sensors for feature extraction of brain signals," in *Proceedings of 2013 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 7420–7423.

[28] H. Higashi, T. Tanaka, and Y. Tanaka, "Smoothing of spatial filter by graph Fourier transform for EEG signals," in *Proceedings of 2014 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, 2014, pp. 1–8.

[29] W. Samek, M. Kawanabe, and K.-R. Müller, "Divergence-based framework for common spatial patterns algorithms," *IEEE Reviews in Biomedical Engineering*, vol. 7, pp. 50–72, 2013.

[30] A. Cichocki, Y. Washizawa, T. Rutkowski, H. Bakardjian, A.-H. Phan, S. Choi, H. Lee, Q. Zhao, L. Zhang, and Y. Li, "Noninvasive BCIs: Multiway signal-processing array decompositions," *Computer*, vol. 41, pp. 34–42, 2008.

[31] A. Cichocki, D. Mandic, A.-H. Phan, C. Caiafa, G. Zhou, Q. Zhao, and L. De Lathauwer, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145–163, 2014.

[32] F. Cong, Q.-H. Lin, L.-D. Kuang, X.-F. Gong, P. Astikainen, and T. Ristaniemi, "Tensor decomposition of EEG signals: A brief review." *Journal of Neuroscience Methods*, vol. 248, pp. 59–69, Jun. 2015.

[33] Y. Liu, Q. Zhao, and L. Zhang, "Uncorrelated Multiway Discriminant Analysis for Motor Imagery EEG Classification," *International Journal of Neural Systems*, vol. 25, no. 04, p. 1550013, Feb. 2015.

[34] Q. Li and D. Schonfeld, "Multilinear Discriminant Analysis for Higher-Order Tensor Data Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 12, pp. 2524–2537, 2014.

[35] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Transactions on Neural Networks*, vol. 19, no. 1, pp. 18–39, 2008.

[36] H. Higashi and T. Tanaka, "Simultaneous design of FIR filter banks and spatial patterns for EEG signal classification," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 4, pp. 1100–1110, 2013.
[37] ——, "Common spatio-time-frequency patterns for motor imagery-based brain machine interfaces." *Computational Intelligence and Neuroscience*, vol. 2013, p. 537218, 2013.
[38] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A survey of multilinear subspace learning for tensor data," *Pattern Recognition*, vol. 44, no. 7, pp. 1540–1551, Jul. 2011.
[39] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.
[40] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, "Single-trial analysis and classification of ERP components–A tutorial." *NeuroImage*, vol. 56, no. 2, pp. 814–25, May 2011.
[41] W. Samek, V. Carmen, K.-R. Müller, and M. Kawanabe, "Stationary common spatial patterns for braincomputer interfacing," *Journal of Neural Engineering*, vol. 9, no. 2, p. 26013, 2012.
[42] 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*, vol. 29, no. 10, pp. 1700–1715, 2007.
[43] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.
[44] S. Kono and T. Rutkowski, "Tactile-force brain-computer interface paradigm," *Multimedia Tools and Applications*, pp. 1–13, 2014.
[45] M. Chang, N. Nishikawa, Z. R. Struzik, K. Mori, S. Makino, D. Mandic, and T. M. Rutkowski, "Comparison of P300 responses in auditory, visual and audiovisual spatial speller BCI paradigms," *Proceedings of the Fifth International Brain-Computer Interface Meeting 2013*, p. 156, 2013.
[46] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. Taylor & Francis, 2014.