A Reduced Rank Approach for Covariance Matrix Estimation in EEG Signal Classification

Naoki Tomida†, Masao Yamagishi†, Isao Yamada†, and Toshihisa Tanaka‡

***Abstract*— *Common Spatial Pattern* (CSP) methods are widely used to extract the brain activity for brain machine interfacing (BMI) based on electroencephalogram (EEG). For each mental task, CSP methods estimate a covariance matrix of EEG signals and adopt the uniform average of the sample covariance matrices over trials. However, the uniform average is sensitive to outliers caused by e.g. unrelated brain activity. In this paper, we propose an improvement of the estimated covariance matrix utilized in CSP methods by reducing the influence of the outliers as well as guaranteeing positive definiteness. More precisely, our estimation is the projection of the uniform average onto the intersection of two convex sets: the first set is a special reduced dimensional subspace which alleviates the influence of the outliers; the second is the positive definite cone. A numerical experiment supports the effectiveness of the proposed technique.**

I. INTRODUCTION

Brain machine interfacing (BMI) is a challenging application of signal processing, machine learning, and neuroscience [1]. BMIs capture brain activities associated with mental tasks and external stimuli and realize non-muscular communication and a control channel for conveying messages and commands to the external world [1]–[3], by using recordings of brain activities such as noninvasive electroencephalogram (EEG), which is the recordings of the electrical activity of neurons on the scalp level [4]. In particular, a noninvasive BMI associated with motor-imagery (MI-BMI) [5], [6] is a crucial field because it realizes to assist that people who have severe motor disabilities control complex movements and to offer a new useful application in medical rehabilitation for paralyzed stroke patients [7]. It is known that motor-imagery tasks evoke the so-called *mu rhythm* [2], [3]. Hence, in the presence of measurement noise and unrelated brain activities, accurately catching such a brain activity has of great importance for realization of MI-BMI.

A well known approach to capture the brain activity for MI-BMI is so-called Common Spatial Pattern (CSP) methods (see also Fig. 1 for the entire flowchart) [8]–[15]. The CSP is a set of spatial weight coefficients corresponding to each electrode in a multichannel EEG. These coefficients are determined in such a way that the variances of the signal extracted by the spatial weights differ between two tasks as much as possible. For maximizing the differences of the exact variances of extracted signals, CSP methods estimate a pair of the true covariance matrices of two different tasks by uniformly averaging the sample covariance matrices of EEG signals observed several times (called trial). However, the uniform average is relatively sensitive to outliers caused by e.g. measurement noise and the unrelated brain activity, which may deteriorate classification performance [16].[1]

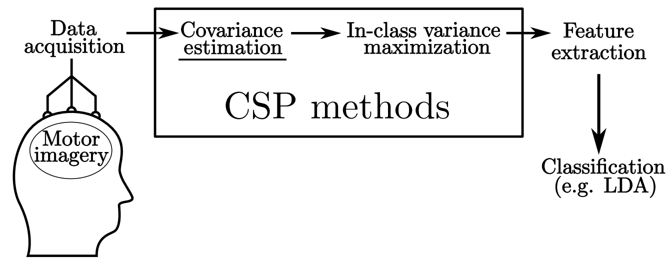

Fig. 1. Flowchart of motor-imagery BMI from data acquisition to classification.

In this paper, we propose a novel covariance matrix estimation to alleviate the sensitivity, for betterment on classification performance of CSP methods.[2] The key idea of our estimation is to reduce the influence of the outliers as well as to guarantee an inherent property of covariance matrices, i.e., positive definiteness of the estimate. More precisely, to alleviate the influence, we focus on a natural observation: for the same mental task, the sample covariance matrices obtained at every trial have similarity except for the outliers; that is, most sample covariance matrices can be approximated well by a linear combination of a few "main" matrices, or there exists a reduced dimensional subspace that can approximate most sample covariance matrices except for the outliers; Hence adopting the subspace as search region of our estimation reduces the influence of the outliers. To specify the subspace, we utilize the singular value decomposition (SVD) of a matrix consisting of the (column) vectorizations of the sample covariance matrices. In addition, to guarantee the positive definiteness, we also utilize the convex cone of positive definite matrices [18], [19]. Consequently, following the minimal disturbance principle, we adopt as our estimation the projection onto the intersection of the reduced dimensional subspace and the cone. To compute the projection, the Dykstra's algorithm (e.g. [20], [21]) is utilized. A numerical experiment demonstrates the effectiveness of the proposed technique.

II. PRELIMINARIES

Let $\mathbb{R}$ denotes the set of all real numbers. For a given positive integer M, We define a *real Hilbert Space* $\mathcal{H}$ as the space of all $M \times M$ real symmetric matrices, i.e. $\mathcal{H} := \{\boldsymbol{X} \in \mathbb{R}^{M\times M} \mid \boldsymbol{X} = \boldsymbol{X}^\top\}$. Here, the *inner product* is defined by $\langle \boldsymbol{X}, \boldsymbol{Y}\rangle = \mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{Y})$, ($\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{H}$), and $\mathrm{tr}(\boldsymbol{A}) := \sum_{i=1}^{M} a_{i,i}$ stands for the trace of the matrix $\boldsymbol{A} := [a_{i,j}] \in \mathbb{R}^{M\times M}$. The induced *norm* $\|\boldsymbol{X}\|_F := \sqrt{\langle \boldsymbol{X}, \boldsymbol{X}\rangle}$, ($\boldsymbol{X} \in \mathcal{H}$), becomes the *Frobenius norm*. A matrix $\boldsymbol{A} \in \mathbb{R}^{M\times M}$ is called *positive (semi) definite* if $\boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v} (\geqslant) > 0$ for all

†N. Tomida, M. Yamagishi, and I. Yamada are with Department of Communications and Computer Engineering, Tokyo Institute of Technology, 2-12-1-S3-60, Ookayama, Meguro-ku, Tokyo, 152-8550, Japan. E-mail: {tomida, myamagi, isao}@sp.ce.titech.ac.jp

‡T. Tanaka is with Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology, 2-24-16, Naka-cho, Koganei-shi, Tokyo, 184-8588, Japan. He is also affiliated with RIKEN Brain Science Institute, Wako-shi, Saitama, Japan. E-mail: tanakat@cc.tuat.ac.jp

This work is supported in part by JSPS Grants-in-Aid (24800022, B-21300091, 24360146).

[1] Instead of the uniform average, a weighted average of sample covariance matrices at every trial was examined in [17].

[2] The proposed estimation technique can be applied to the variants of the CSP method and has potential to improve their performance. However, we show for simplicity the effectiveness of the proposed technique by utilizing the standard CSP method [8], [9] in this paper.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

nonzero $\boldsymbol{v} \in \mathbb{R}^M$, where $\cdot^\top$ denotes *transposition*. Additionally, given the matrices $\boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{M\times M}$, the notation $\boldsymbol{B}(\succcurlyeq) \succ \boldsymbol{C}$ is equivalent to that $\boldsymbol{B}-\boldsymbol{C}$ is positive (semi) definite.

A set $\mathcal{C} \subset \mathcal{H}$ is called *convex* if $\forall \boldsymbol{X}, \boldsymbol{Y} \in \mathcal{C}$ and $\lambda \in (0,1)$, $\lambda\boldsymbol{X} + (1-\lambda)\boldsymbol{Y} \in \mathcal{C}$. Given a *closed convex* set $\mathcal{C} \subset \mathcal{H}$, the *metric projection* onto $\mathcal{C}$ is the mapping $P_\mathcal{C} : \mathcal{H} \to \mathcal{C} : \boldsymbol{X} \mapsto P_\mathcal{C}(\boldsymbol{X})$ s.t. $d(\boldsymbol{X}, \mathcal{C}) := \min\{\|\boldsymbol{X}-\boldsymbol{Y}\|_F \mid \boldsymbol{Y} \in \mathcal{C}\} = \|\boldsymbol{X} - P_\mathcal{C}(\boldsymbol{X})\|_F$.

A. Common Spatial Pattern (CSP)

We briefly review the standard CSP method [8], [9]. The CSP is given as a spatial weight vector, $\boldsymbol{w} \in \mathbb{R}^M$, which attempts to maximize the in-class variance for a self task (class). Consider a desired case that we have exact knowledge on the covariance matrices $\boldsymbol{S}_d \in \mathcal{H}$ $(d \in \{+,-\})$ in the two classes (e.g., imagination of left and right hand movements), where $+$ and $-$ indicate the class labels. Then for a class d, the CSP is obtained as a minimizer of the following problem:

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top \boldsymbol{S}_d \boldsymbol{w}, \\ \text{subject to} \quad & \boldsymbol{w}^\top(\boldsymbol{S}_+ + \boldsymbol{S}_-)\boldsymbol{w} = 1. \end{aligned} \tag{1}$$

The solution of (1) is given by the generalized eigenvector corresponding to the largest generalized eigenvalue:

$$\boldsymbol{S}_d\boldsymbol{w} = \lambda(\boldsymbol{S}_+ + \boldsymbol{S}_-)\boldsymbol{w}. \tag{2}$$

In practice, to estimate the covariance matrices of two classes, the uniform average of multiple measurements of EEG signals is adopted, i.e.,

$$\overline{\boldsymbol{S}}_d = \frac{1}{K_d}\sum_{k\in\mathfrak{C}_d} \boldsymbol{S}^{(k)}, \; d \in \{+,-\}, \tag{3}$$

where $\mathfrak{C}_d$ is the indices of training data containing the signals observed at all trials belonging to class d, K_d is the cardinality of $\mathfrak{C}_d$, and $\boldsymbol{S}^{(k)}$ is the sample covariance matrix at the kth trial defined by

$$\boldsymbol{S}^{(k)} := \frac{1}{N}\boldsymbol{X}^{(k)}\boldsymbol{X}^{(k)\top} \in \mathbb{R}^{M\times M}, \tag{4}$$

where $\boldsymbol{X}^{(k)} \in \mathbb{R}^{M\times N}$ is a matrix consisting of zero mean M channel signals with N samples at the kth trial.[3] In general, each channel signal in $\boldsymbol{X}^{(k)}$ passed through a bandpass filter which passes the frequency components related to the target brain activity.

B. Dykstra's Algorithm

The Dykstra's algorithm computes the projection onto the intersection of two convex sets.[4] Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be closed convex sets of $\mathcal{H}$ such that $\mathcal{C}_1 \cap \mathcal{C}_2 \neq \phi$, and for an arbitrary initial $\boldsymbol{X}_0 \in \mathcal{H}$, generate sequences $\boldsymbol{X}_n, \boldsymbol{Y}_n \in \mathcal{H}$ as follows:

$$\begin{cases} \boldsymbol{p}_0 = \boldsymbol{q}_0 = 0 \\ \boldsymbol{Y}_n = P_{\mathcal{C}_1}(\boldsymbol{X}_n + \boldsymbol{p}_n) \\ \boldsymbol{p}_{n+1} = \boldsymbol{X}_n + \boldsymbol{p}_n - \boldsymbol{Y}_n \\ \boldsymbol{X}_{n+1} = P_{\mathcal{C}_2}(\boldsymbol{Y}_n + \boldsymbol{q}_n) \\ \boldsymbol{q}_{n+1} = \boldsymbol{Y}_n + \boldsymbol{q}_n - \boldsymbol{X}_{n+1}. \end{cases} \tag{5}$$

Then, $\{\boldsymbol{X}_n\}_{n=0}^\infty$ and $\{\boldsymbol{Y}_n\}_{n=0}^\infty$ converge to $P_{\mathcal{C}_1\cap\mathcal{C}_2}(\boldsymbol{X}_0)$.

III. Reduced Rank Estimation of Covariance Matrices

Consider an estimation of a covariance matrix $\boldsymbol{S}_d$ by using sample covariance matrices $\boldsymbol{S}^{(k)}$, $k \in \mathfrak{C}_d$, for one of the classes (the same procedure will be applied for each class). We improve the covariance matrix estimation utilized in the CSP method, i.e., $\overline{\boldsymbol{S}}_d$ in (3), by projecting the intersection of the two closed convex sets: The first set $\mathcal{C}_R$ is a subspace utilized to alleviate the influence of the outliers, the second $\mathcal{C}_P$ is a cone of positive definite matrices. The resulting matrix $\boldsymbol{S}_d^\star$ is adopted as a new estimate (see Fig. 2).

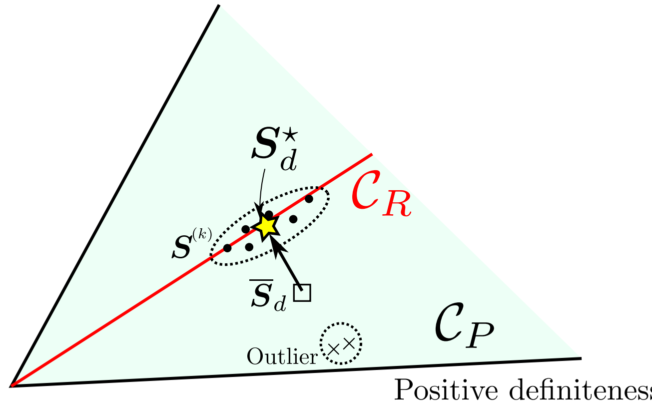

Fig. 2. Illustration of the proposed technique. Since the subspace $\mathcal{C}_R$ well approximates a group of sample covariance matrices $\boldsymbol{S}^{(k)}$, the metric projection onto the intersection of $\mathcal{C}_R$ and $\mathcal{C}_P$ results in a reasonable covariance estimate $\boldsymbol{S}_d^\star$ from the unstable covariance estimate $\overline{\boldsymbol{S}}_d$ utilized in the standard CSP method.

The designs of the convex sets are the following: First, to alleviate the influence of the outliers, we adopt a specially designed reduced dimensional subspace of $\mathbb{R}^{M^2}$ $(\equiv \mathbb{R}^{M\times M})$. Assume that the sample covariance matrices $\boldsymbol{S}^{(k)}$ are similar except for the outliers, and the outliers are not dominant. Then we can expect that there exists a subspace that approximates well many similar sample covariance matrices and does poorly the outliers. Hence finding the estimate of the covariance matrix $\boldsymbol{S}_d$ from such a subspace would reduce the influence of outliers. We embody this idea by the Singular Value Decomposition (SVD) of the following matrix[5];

$$\boldsymbol{Z} = \left[\mathrm{vec}(\boldsymbol{S}^{(1)}), \mathrm{vec}(\boldsymbol{S}^{(2)}), \ldots, \mathrm{vec}(\boldsymbol{S}^{(K_d)})\right] \in \mathbb{R}^{M^2\times K_d}, \tag{6}$$

where $\mathrm{vec}(\cdot)$ denotes the vectorization operator. We denote SVD of $\boldsymbol{Z}$ as

$$\boldsymbol{Z} = \boldsymbol{U}_{\boldsymbol{Z}}\boldsymbol{\Sigma}_{\boldsymbol{Z}}\boldsymbol{V}_{\boldsymbol{Z}}^\top, \tag{7}$$

where the orthonormal matrices $\boldsymbol{U}_{\boldsymbol{Z}} := [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{M^2}] \in \mathbb{R}^{M^2\times M^2}$ and $\boldsymbol{V}_{\boldsymbol{Z}} := [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{K_d}] \in \mathbb{R}^{K_d\times K_d}$ consist of the left and right singular vectors, and the diagonal matrix $\boldsymbol{\Sigma}_{\boldsymbol{Z}} \in \mathbb{R}^{M^2\times K_d}$ contains the singular values $\{\sigma_i(\boldsymbol{Z})\}_{i=1}^{r_{\max}}$ on its main diagonal in decreasing order and 0's elsewhere $(r_{\max} := \min\{M^2, K_d\})$. Then, except for outliers, sample covariance matrices can be approximated well by the subspace spanned by the left singular vectors associated with large singular values. Hence the true covariance matrix $\boldsymbol{S}_d$ is expected to be near

$$\mathcal{C}_R := \{\boldsymbol{S} \in \mathcal{H} \mid \mathrm{vec}(\boldsymbol{S}) \in \mathcal{R}(\boldsymbol{U}_{\boldsymbol{Z}}^r)\}, \tag{8}$$

where r $(\leq r_{\max})$ is the number of remaining singular values,

$$\boldsymbol{U}_{\boldsymbol{Z}}^r = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r] \in \mathbb{R}^{M^2\times r},$$
$$\mathcal{R}(\boldsymbol{U}_{\boldsymbol{Z}}^r) := \{\boldsymbol{U}_{\boldsymbol{Z}}^r\boldsymbol{k}_r \in \mathbb{R}^{M^2} \mid \forall \boldsymbol{k}_r \in \mathbb{R}^r\}.$$

[3] The sample mean is subtracted from the observed data if necessary (the M channel signals have nonzero mean).

[4] Although the Dykstra's algorithm can compute the projection onto the intersection of finitely many convex sets, its special form is shown here for simplicity.

[5] If $\boldsymbol{Z}$ has column vectors with extremely large norm, they should be normalized before applying SVD.

Algorithm 1 solver for optimization problem (10)

Given $\boldsymbol{S}^{(k)}(k=1,\ldots,K_d)$.
Choose number r of remaining singular values.
Set $\boldsymbol{S}_0=\overline{\boldsymbol{S}}_d=\frac{1}{K_d}\sum_{k=1}^{K_d}\boldsymbol{S}^{(k)}$ (see also (3)), $\boldsymbol{p}_0=\boldsymbol{0}$, $\boldsymbol{q}_0=\boldsymbol{0}$, and $\boldsymbol{Q}=\epsilon\boldsymbol{I}_M$.
$\boldsymbol{Z}=[\mathrm{vec}(\boldsymbol{S}^{(1)}),\mathrm{vec}(\boldsymbol{S}^{(2)}),\ldots,\mathrm{vec}(\boldsymbol{S}^{(K_d)})]$.
$\boldsymbol{U}_{\boldsymbol{Z}}^r$ = r leading left singular vectors of $\boldsymbol{Z}$.
repeat
1. $\widetilde{\boldsymbol{S}}_n=\mathrm{vec}^{-1}\left[\boldsymbol{U}_{\boldsymbol{Z}}^r\boldsymbol{U}_{\boldsymbol{Z}}^{r\top}\mathrm{vec}(\boldsymbol{S}_n+\boldsymbol{p}_n)\right]$.
2. $\boldsymbol{p}_{n+1}=\boldsymbol{S}_n+\boldsymbol{p}_n-\widetilde{\boldsymbol{S}}_n$.
3. Compute an eigendecomposition $\boldsymbol{U}_{\widetilde{\boldsymbol{S}}_n+\boldsymbol{q}_n-\boldsymbol{Q}}\boldsymbol{\Lambda}\boldsymbol{U}_{\widetilde{\boldsymbol{S}}_n+\boldsymbol{q}_n-\boldsymbol{Q}}^{\top}$ of $\widetilde{\boldsymbol{S}}_n+\boldsymbol{q}_n-\boldsymbol{Q}$.
4. Construct $\underline{\boldsymbol{\Lambda}}$ by replacing the negative entries in $\boldsymbol{\Lambda}$ by zeros.
5. $\boldsymbol{S}_{n+1}=\boldsymbol{U}_{\widetilde{\boldsymbol{S}}_n+\boldsymbol{q}_n-\boldsymbol{Q}}\underline{\boldsymbol{\Lambda}}\boldsymbol{U}_{\widetilde{\boldsymbol{S}}_n+\boldsymbol{q}_n-\boldsymbol{Q}}^{\top}+\boldsymbol{Q}$.
6. $\boldsymbol{q}_{n+1}=\widetilde{\boldsymbol{S}}_n+\boldsymbol{q}_n-\boldsymbol{S}_{n+1}$.

until converged.

Next, to guarantee the positive definiteness of the resulting matrix, we adopt a positive definite cone

$$\mathcal{C}_P=\{\boldsymbol{S}\in\mathcal{H}\mid\boldsymbol{S}\succcurlyeq\boldsymbol{Q}\},\tag{9}$$

where $\boldsymbol{Q}\in\mathcal{H}$ is a positive definite matrix. A typical choice is $\boldsymbol{Q}:=\epsilon\boldsymbol{I}_M$, with a predefined parameter $\epsilon>0$ and the identity matrix $\boldsymbol{I}_M\in\mathcal{H}$.

Consequently, our estimation of the covariance matrix is

$$\boldsymbol{S}_d^{\star}=P_{\mathcal{C}_R\cap\mathcal{C}_P}\left(\overline{\boldsymbol{S}}_d\right)=\underset{\boldsymbol{S}\in\mathcal{C}_R\cap\mathcal{C}_P}{\arg\min}\ \frac{1}{2}\left\|\boldsymbol{S}-\overline{\boldsymbol{S}}_d\right\|_F^2,\tag{10}$$

where $\overline{\boldsymbol{S}}_d=\frac{1}{K_d}\sum_{k=1}^{K_d}\boldsymbol{S}^{(k)}$. To solve the optimization problem in (10), we adopt the Dykstra's algorithm (5) with $\mathcal{C}_1:=\mathcal{C}_R$, $\mathcal{C}_2:=\mathcal{C}_P$, and $\boldsymbol{X}_0:=\overline{\boldsymbol{S}}_d$. The detailed algorithm is shown in Algorithm 1.

Remark 1 (Computation of Projections): The metric projection $P_{\mathcal{C}_R}:\mathcal{H}\to\mathcal{C}_R$ is given by

$$P_{\mathcal{C}_R}(\boldsymbol{S})=\mathrm{vec}^{-1}\left[\boldsymbol{U}_{\boldsymbol{Z}}^r\boldsymbol{U}_{\boldsymbol{Z}}^{r\top}\mathrm{vec}(\boldsymbol{S})\right].\tag{11}$$

The metric projection $P_{\mathcal{C}_P}:\mathcal{H}\to\mathcal{C}_P$ is given by

$$P_{\mathcal{C}_P}(\boldsymbol{S})=\boldsymbol{U}_{\boldsymbol{S}-\boldsymbol{Q}}\underline{\boldsymbol{\Lambda}}\boldsymbol{U}_{\boldsymbol{S}-\boldsymbol{Q}}^{\top}+\boldsymbol{Q},\tag{12}$$

where $\boldsymbol{U}_{\boldsymbol{S}-\boldsymbol{Q}}\boldsymbol{\Lambda}\boldsymbol{U}_{\boldsymbol{S}-\boldsymbol{Q}}^{\top}$ is the eigendecomposition of $\boldsymbol{S}-\boldsymbol{Q}$, i.e. $\boldsymbol{U}_{\boldsymbol{S}-\boldsymbol{Q}}$ is the $M\times M$ orthonormal matrix whose columns are the eigenvectors of the matrix $\boldsymbol{S}-\boldsymbol{Q}$ as well as $\boldsymbol{\Lambda}$ is the diagonal matrix consisting of the eigenvalues of $\boldsymbol{S}-\boldsymbol{Q}$, and the matrix $\underline{\boldsymbol{\Lambda}}$ is formed by replacing the negative entries in $\boldsymbol{\Lambda}$ by zeros.

Remark 2 (Selection of Criterion): Although we adopt for simplicity the Frobenius norm to measure dissimilarity of two covariance matrices in (10), employing other convex criteria has potential to improve further. For example, in consideration of symmetricity of covariance matrices, one of natural choice is a weighted Frobenius norm such as

$$\frac{1}{2}\left\|\boldsymbol{\Theta}\odot(\boldsymbol{S}-\overline{\boldsymbol{S}}_d)\right\|_F^2,$$

where $\boldsymbol{\Theta}:=[\theta_{i,j}]\in\mathbb{R}^{M\times M}$ is defined by

$$\theta_{i,j}=\begin{cases}1 & (j\geq i)\\ 0 & (j<i)\end{cases},$$

and $\odot$ implies the Hadamard product (or entry-wise product). In addition, a convex relaxation of the geodesic distance between two symmetric positive definite matrices (e.g. [22]) is also a strong candidate.

TABLE I
DATA DESCRIPTION OF DATA SET IVA AND DATA SET 1.

	data set IVa	data set 1
subject labels	*aa,al,av,aw,ay*	*a,b,f,g*
the number of channels	118	59
signal length (sec)	3.5	4.0
sampling rate (Hz)	1000	100
the number of trials per class	140	100

IV. EXPERIMENT OF TWO EEG CLASSIFICATION

We conduct an experiment of binary classification of EEG signal during motor imagery, to confirm effectiveness of the proposed covariance matrix estimation. Although the proposed technique can be utilized in most of CSP methods, we apply it to the standard CSP method [8], [9] for showing its fundamental property.

In this experiment, we used two datasets. The first is dataset IVa from BCI competition III, which was provided by Fraunhofer FIRST (Intelligent Data Analysis Group) and Campus Benjamin Franklin of the Charité - University Medicine Berlin (Department of Neurology, Neurophysics Group) [23]. This dataset consists of EEG signals during right hand and right foot motor-imageries. In this experiment, we furthermore applied to those data a bandpass filter whose passband is 7–30 Hz and downsampled to 100 Hz. The second is dataset 1 from BCI competition IV, which was provided by Berlin Institute of Technology (Machine Learning Laboratory), Fraunhofer FIRST (Intelligent Data Analysis Group) and Campus Benjamin Franklin of the Charité - University Medicine Berlin (Department of Neurology, Neurophysics Group) [24]. This dataset consists of two motor-imageries, which were selected from the three classes left hand, right hand, and foot (side chosen by the subject; optionally also both feet). In this experiment, we furthermore applied to those data a bandpass filter whose passband is 7–30 Hz. A detailed description of two datasets is shown in Table I.

A. Classification Algorithm

We defined the feature vector as the output of feature extraction using the CSP (and with the proposed covariance matrix estimation). Although the solution of (1) is given by the eigenvector corresponding to the smallest eigenvalue in (2), other eigenvectors can be utilized for improving classification accuracy [25]. Following this strategy, we defined the feature vector $\boldsymbol{y}$ by

$$\begin{aligned}&\boldsymbol{y}=[y_1,\ldots,y_l,y_{M-l+1},\ldots,y_M]^{\top}\in\mathbb{R}^{2l},\\ &y_i=\frac{1}{N}\left\|\hat{\boldsymbol{w}}_i^{\top}\boldsymbol{X}\right\|_2^2,i\in\{1,\ldots,l,M-l+1,\ldots,M\}\end{aligned}\tag{13}$$

with the first l and the last l eigenvectors, for classification of unlabeled data $\boldsymbol{X}$, where $\hat{\boldsymbol{w}}_i$ is the eigenvector corresponding to the ith smallest eigenvalue of (2) ($i=1,2,\ldots,M$).

In this experiment, we adopted the linear discriminant analysis (LDA) [26] with the feature vector $\boldsymbol{y}$ as its input. In all cases, for the sake of simplicity of comparison, the number l of the associated spatial weights in (13) is fixed to 3.

B. Result

Table II shows results of the experiment of the standard CSP method with two covariance estimations: the uniform average (3) or the proposed technique (10). The results were obtained by conducting 5-fold cross validation. For all the subjects, our technique improves the resulting classification accuracy, which clearly

TABLE II

CLASSIFICATION ACCURACY [%] GIVEN BY 5-FOLD CROSS VALIDATION. IN PROPOSED 1 (THE STANDARD CSP WITH THE PROPOSED TECHNIQUE), WE SHOW THE HIGHEST CLASSIFICATION ACCURACY AT EACH SUBJECT AMONG ALL POSSIBLE PAIR (r_+, r_-), WHICH IS A PAIR OF THE NUMBER OF REMAINING SINGULAR VALUES IN CLASS $+$ AND IN CLASS $-$, IN PROPOSED 2, WE ALSO SHOW THE HIGHEST ACCURACY AMONG EVERY (r_+, r_-) SUCH THAT $r_+ = r_-$. DUE TO 5 FOLD CROSS VALIDATION, IN DATA SET IVA, r_+ AND r_- ARE BOUNDED ABOVE BY $r_{\max}(=\min\{M^2, K_d\}) = 112$ BECAUSE OF $M = 118$ CHANNEL SIGNALS AND $K_d = 112$ TRIALS IN EACH CLASS. SIMILARLY, FOR DATA SET 1, WE HAVE $r_{\max} = 80$ BY $(M, K_d) = (59, 80)$. NOTE THAT IN ALL THE METHODS THE NUMBER OF THE ASSOCIATED SPATIAL WEIGHTS, $2l$ IN (13), IS FIXED TO 6.

Method	subjects (data set IVa)						subjects (data set 1)				
	aa	*al*	*av*	*aw*	*ay*	Ave.	*a*	*b*	*f*	*g*	Ave.
CSP	75.71	93.57	63.21	97.86	92.86	84.64	66.00	71.50	88.50	89.00	78.75
Proposed 1	80.71	95.00	71.07	98.57	94.29	87.93	81.00	78.50	90.00	91.00	85.13
(r_+, r_-)	(5, 47)	(55, 5)	(2, 101)	(103,26)	(112, 20)		(25, 3)	(5, 43)	(73, 39)	(30, 80)	
Proposed 2	79.29	93.57	69.29	98.21	93.21	86.71	78.50	76.50	89.00	90.00	83.50
$(r_+ = r_-)$	(5)	(23)	(9)	(58)	(25)		(5)	(3)	(71)	(32)	

demonstrates the effectiveness of the proposed technique. Finally, note that computational cost for Algorithm 1 is significantly low in this experiment because it takes only one iteration to reach the solution $\boldsymbol{S}_d^\star$ of (10) for every number of remaining singular values.

V. CONCLUDING REMARKS

In this paper, we have proposed an improvement of the covariance matrix estimation utilized in CSP methods. The underlying idea is to alleviate the influence of the outliers as well as to guarantee the inherent property of covariance matrices, enjoying the virtue of the conventional methods. This idea has been embodied as the projection onto the intersection of two convex sets: A special reduced dimensional subspace that can reduce the influence of the outliers; the positive definite matrix cone. Numerical experiments have shown that the proposed method achieves improved classification accuracy with a low additional computational cost. Future work includes (i) extensions of the proposed technique for variants of the CSP method (e.g. [10]–[15]), (ii) systematic choices of the number of remaining singular values, e.g., by extending a technique utilized in context of image processing [27], (iii) sophisticated designs of dissimilarity criterion of two covariance matrices, and (iv) further performance improvements of the proposed technique by combining the result in [17], i.e., the use of a weighted average, instead of the uniform average utilized in the CSP method, of sample covariance matrices.

REFERENCES

[1] G. Dornhege, J. d. R. Millan, T. Hinterberger, D. McFarland, and K.-R. Müller, *Toward Brain-Computer Interfacing*. Cambridge, MA: MIT Press, 2007.

[2] S. Sanei, *Adaptive Processing of Brain Signals*. Hoboken, NJ: John Wiley & Sons, April 2013.

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

[4] D. J. McFarland and J. R. Wolpaw, "Brain-computer interface operation of robotic and prosthetic devices," *Computer*, vol. 41, no. 6, pp. 52–56, 2008.

[5] ——, "Brain-computer interfaces for communication and control," *Commun. ACM*, vol. 54, no. 5, pp. 60–66, 2011.

[6] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans," *Nat. Acad. Sci.*, vol. 101, no. 51, pp. 17 849–17 854, 2004.

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

[8] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clin. Neurophysiol.*, vol. 110, no. 5, pp. 787–798, 1999.

[9] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, 2000.

[10] S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 9, pp. 1541–1548, 2005.

[11] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Müller, "Combined optimization of spatial and temporal filters for improving brain-computer interfacing," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 11, pp. 2274–2281, 2006.

[12] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. 2008 Int. Joint Conf. Neural Netw. (IJCNN)*, 2008, pp. 2390–2397.

[13] H. Higashi and T. Tanaka, "Simultaneous design of FIR filter banks and spatial patterns for EEG signal classification," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 4, pp. 1100–1110, 2013.

[14] ——, "Common spatio-time-frequency patterns for motor-imagery based brain machine interfaces," *Computational Intelligence and Neuroscience*, vol. 2013, 2013.

[15] 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 Trans. Biomed. Engineering*, vol. 57, no. 12, pp. 2936–2946, 2010.

[16] X. Yong, R. K. Ward, and G. E. Birch, "Robust common spatial patterns for EEG signal preprocessing," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*. IEEE, 2008, pp. 2087–2090.

[17] N. Tomida, H. Higashi, and T. Tanaka, "A joint tensor diagonalization approach to active data selection for EEG classification," in *Proc. IEEE ICASSP*, 2013, pp. 983–987.

[18] K. Slavakis, I. Yamada, and K. Sakaniwa, "Computation of symmetric positive definite Toeplitz matrices by the hybrid steepest descent method," *Signal Processing*, vol. 83, no. 5, pp. 1135–1140, 2003.

[19] N. J. Higham, "Computing a nearest symmetric positive semidefinite matrix," *Linear algebra and its applications*, vol. 103, pp. 103–118, 1988.

[20] J. P. Boyle and R. L. Dykstra, "A method for finding projections onto the intersection of convex sets in Hilbert spaces," in *Advances in Order Restricted Statistical Inference*, R. Dykstra, T. Robertson, and F. T. Wright, Eds. Berlin, Germany: Springer, 1986, pp. 28–47.

[21] H. H. Bauschke and P. L. Combettes, "A Dykstra-like algorithm for two monotone operators," *Pacific Journal of Optimization*, vol. 4, no. 3, pp. 383–391, 2008.

[22] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain–computer interface classification by Riemannian geometry," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 4, pp. 920–928, 2012.

[23] B. Blankertz, K.-R. Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlögl, G. Pfurtscheller, J. del R. Millan, M. Schröder, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 153–159, 2006.

[24] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio, "The non-invasive Berlin brain–computer interface: fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539–550, 2007.

[25] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, 2008.

[26] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006.

[27] M. Yamagishi and I. Yamada, "A rank selection of MV-PURE with an unbiased predicted-MSE criterion and its efficient implementation in image restoration," in *Proc. IEEE ICASSP*, 2013, pp. 1573–1577.