Robust EEG-based Motor Imagery Brain-Computer Interface using Lp-norm-based Common Spatial Patterns

Ji-Hack Lee
Department of Electronics and Communications Engineering
Kwangwoon University
Seoul, Republic of Korea
dlwlgkr159@ kw.ac.kr

Hyun-Kyu Lee
Department of Electronics and Communications Engineering
Kwangwoon University
Seoul, Republic of Korea
skgusrb12@kw.ac.kr

Young-Seok Choi*
Department of Electronics and Communications Engineering
Kwangwoon University
Seoul, Republic of Korea
yschoi@kw.ac.kr

***Abstract*—Electroencephalogram (EEG) signals are inevitably contaminated by outliers, artifacts, and noise since EEGs are non-invasively recorded on the scalp. Motor imagery (MI) based brain-computer interface (BCI) makes use of multivariate EEG signals, which aims at classifying MI features. The existence of outliers leads to a degradation of the classification performance of MI-based BCI (MI-BCI). Previous popular common spatial patterns (CSPs) based on L2-norm and L1-norm dispersion are sensitive to outliers in EEG signals. This paper presents a generalized Lp-norm based CSP (p<1) to yield robustness to outliers for MI-BCI. By utilizing the Lp-norm dispersion of the filtered EEG samples instead of L2-norm or L1-norm ones, the robust MI-BCI to outliers is expected. Through simulations using a toy dataset and public EEG BCI dataset, we validate the capacity of the proposed Lp-norm BCI (CSP-Lp) and confirm its increased robustness to outliers.**

Keywords—brain-computer interfaces, electroencephalogram, motor imagery, common spatial patterns, Lp-norm

I. INTRODUCTION

Brain-computer interface (BCI) provides an alternative way of transmitting human intentions or responses to external devices by interpreting the brain functional activities under specific environments [1]–[3]. Among the monitoring modalities of brain functional activities, electroencephalogram (EEG), which is recorded on the scalp, is widely used in the field of noninvasive BCI. Depending on the type of EEGs, there are several categories of noninvasive BCI models [4], [5]. Typically, motor imagery (MI), event-related potentials (ERP) and steady-state visual evoked potential (SSVEP) are the most widespread noninvasive BCI models. Among those, the MI-based BCI (MI-BCI) utilizes the brain patterns in cases of the imagination of motor movements, which is most intuitive than others.

The main notable patterns of EEGs while imaging motor movements are a decrease and an increase of alpha and beta band powers, which is known as event-related desynchronization (ERD) and event-related synchronization (ERS), respectively [6]. In addition, ERD and ERS are observed in EEG recorded from the motor cortex area of the scalp. Recent advances in machine learning approaches enable researchers to detect MI patterns with the improved performance [7], [8].

To identify MI patterns from multivariate EEG signals, spatial filtering is usually carried out. Common spatial patterns (CSP) in [9], [10] has been a standard among various spatial filtering methods for a two-classes scenario. CSP seeks few spatial filters which not only maximize a variance of the filtered EEG signals of one class, but also minimize that of the other class based on the L2-norm space. Due to its promising performance for single-trial classification of BCI model, a variety of variants of CSP has been presented [11]–[14]. In [11], the authors have presented a regularized CSP (RCSP) by utilizing the Tikhonov regularization and Laplacian penalty. Further, other extended versions of CSP such as common spatio-spectral patterns (CSSP) [12], sparse CSP (SCSP) [13], and filter bank common spatial pattern (FBCSP) [14] have been introduced and shown promising results. However, when EEG signals are contaminated by outliers caused by eye blinks, head movement, and electrical noise, the L2-norm variance computation exaggerates outliers, which results in a deterioration of the performance of L2-norm based CSP (hereafter CSP-L2) or extended CSP- L2.

To deal with vulnerability to outliers of CSP- L2 families, the robust versions of the CSP- L2 by applying L1-norm computation, named CSP-L1 [15] and regularized CSP-L1 [16], have been developed. These works have exploited the L1-norm based dispersion instead of L2-norm based variance, which has yielded robustness to outliers. However, CSP-L1 still suffer from performance degradation due to inevitable outliers or artifacts. Comparing to L1-norm and L2-norm, Lp-norm ($0<p<1$) has rendered improved robustness against outliers in various signal processing applications [17], [18]. In [17] and [18], Lp-norm based principal component analysis (PCA-Lp) and linear discriminant analysis (LDA-Lp) demonstrated enhanced performance compared to L1-norm-based methodologies, respectively.

In this work, we generalized a CSP framework in order to restrict the outlier influence for MI-BCI by using Lp-norm ($p<1$) dispersion in cases where computing L1-norm and L2-norm previously. The proposed Lp-norm based CSP (CSP-Lp) calculates the Lp-norm dispersions of filtered EEG signals rather than L1 and L2-norms. Thus, the proposed CSP-Lp is the generalization of the conventional CSP-L2 and CSP- L1. CSP-Lp based MI-BCI is analyzed and evaluated using a toy dataset and public EEG BCI dataset provided by BCI Competition. The experimental results demonstrate the capability of the proposed CSP-Lp to decode MI-BCI in cases where severe outliers exist.

II. Common Spatial Patterns (CSPs)

A. L2-norm based CSP (CSP-L2)

Let $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{\text{channel} \times \text{sample}}$ be pre-processed EEG samples, the objective function of CSP-L2 is given by

$$J_{L2}(\boldsymbol{w}) = \frac{\boldsymbol{w}^T \boldsymbol{X}\boldsymbol{X}^T \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{Y}\boldsymbol{Y}^T \boldsymbol{w}} = \frac{\boldsymbol{w}^T \boldsymbol{C}_{\boldsymbol{X}} \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{C}_{\boldsymbol{Y}} \boldsymbol{w}} = \frac{\|\boldsymbol{w}^T \boldsymbol{X}\|_2^2}{\|\boldsymbol{w}^T \boldsymbol{Y}\|_2^2} \quad (1)$$

where $\boldsymbol{w} \in \mathbb{R}^{\text{channel}}$ is a spatial filter of CSP, T denotes transpose operator, and $\boldsymbol{C}_{\boldsymbol{X}}$, $\boldsymbol{C}_{\boldsymbol{Y}}$ represent the covariances of X and Y, respectively. Since CSP-L2 aims at obtaining few spatial filters that maximize variance for one class MI-BCI while minimizing variance for other class MI-BCI, the goal is to find an optimum spatial filter that maximizes the objective function $J_{L2}(w)$ in (1). The optimal spatial filter $\boldsymbol{w}$ is obtained by solving the following generalized eigenvalue equation as follows as:

$$\boldsymbol{C}_{\boldsymbol{X}} \boldsymbol{w} = \lambda \boldsymbol{C}_{\boldsymbol{Y}} \boldsymbol{w} \quad (2)$$

In (2), we can see that the spatial filter $\boldsymbol{w}$ is the eigenvector of covariance matrices $\boldsymbol{C}_{\boldsymbol{X}}$, $\boldsymbol{C}_{\boldsymbol{Y}}$. In addition, the eigenvalue λ represents the ratio of two covariances, $\boldsymbol{C}_{\boldsymbol{X}}$ and $\boldsymbol{C}_{\boldsymbol{Y}}$. Thus, among the eigenvectors of the number of channels, the eigenvectors corresponding to the largest and the smallest eigenvalues are chosen as the spatial filters. For the classification, the variance of the logarithm of the spatial filtered EEG signals is used as the feature in the CSP-L2 based MI-BCI.

B. L1-norm based CSP (CSP-L1)

L2-norm is generally known to be sensitive to outliers since it exaggerates the effect of outliers by using the variance on Euclidean distance. As a solution for this issue, the L1-norm based CSP was developed with the following objective function:

$$J_{L1}(\boldsymbol{w}) = \frac{\|\boldsymbol{w}^T \boldsymbol{X}\|_1}{\|\boldsymbol{w}^T \boldsymbol{Y}\|_1} = \frac{\sum_{i=1}^{m} (\boldsymbol{w}^T \boldsymbol{x}_i)}{\sum_{j=1}^{n} (\boldsymbol{w}^T \boldsymbol{y}_j)} \quad (3)$$

where $\boldsymbol{x}_i, \boldsymbol{y}_j \in \mathbb{R}^{\text{channel}}$ are the vectors of one sample of each class MI EEG data, and m and n are samples of each class MI EEG data. Unlike CSP-L2, it is not available to compute differential calculation in CSP-L1, so iterative algorithm is used. The optimized spatial filter is obtained by combining a spatial filter to maximize the objective function $J_{\mathrm{L1}}(\boldsymbol{w})$ and a spatial filter to maximize $1/J_{\mathrm{L1}}(\boldsymbol{w})$.

III. Proposed Method

A. Lp-norm based CSP (CSP-Lp)

Although CSP-L1 has been proposed to improve sensitivity to outliers, it still appears to be sensitive to large outliers. To tackle this obstacle, we develop a novel objective function by using Lp-norm (0<p<1) instead of L2-norm or L1-norm. The goal of CSP-Lp is to find an optimal spatial filter $\boldsymbol{w}$ that maximizes the following objective function:

$$J_{L_P}(\boldsymbol{w}) = \frac{\|\boldsymbol{w}^T \boldsymbol{X}\|_P^P}{\|\boldsymbol{w}^T \boldsymbol{Y}\|_P^P} = \frac{\sum_{i=1}^{m} |\boldsymbol{w}^T \boldsymbol{x}_i|^P}{\sum_{j=1}^{n} |\boldsymbol{w}^T \boldsymbol{y}_j|^P} \quad (4)$$

where $0 < p < 1$.

We use a gradient ascent approach to get the maximum value of the objective function $J_{\mathrm{L1}}(\boldsymbol{w})$. First, we differentiate the objective function to obtain the gradient of $J_{\mathrm{L1}}(\boldsymbol{w})$ with respect to $\boldsymbol{w}$. However, owing of the absolute value formula of $J_{L_P}(\boldsymbol{w})$, the gradient may not be clearly defined. To avoid this problem, we rewrite the objective function $J_{\mathrm{L1}}(\boldsymbol{w})$ with a sign function for absolute value operations.

$$J_{L_P}(\boldsymbol{w}) = \frac{\sum_{i=1}^{m} [\mathrm{sgn}(\boldsymbol{w}^T \boldsymbol{x}_i) \boldsymbol{w}^T \boldsymbol{x}_i]^P}{\sum_{j=1}^{n} [\mathrm{sgn}(\boldsymbol{w}^T \boldsymbol{y}_j) \boldsymbol{w}^T \boldsymbol{y}_j]^P} \quad (5)$$

The sgn(a) function yields a value of -1 or 1 depending on the sign of a. The gradient of $J_{L_P}(\boldsymbol{w})$ with respect to $\boldsymbol{w}$ is given by

$$\nabla_{\boldsymbol{w}} = \frac{\partial J_{L_P}(\boldsymbol{w})}{\partial \boldsymbol{w}} = \frac{\mathrm{A} \times \mathrm{B} - \mathrm{C} \times \mathrm{D}}{\mathrm{E}} \quad (6)$$

$$\mathrm{A} = P \sum_{i=1}^{m} \mathrm{sgn}(\boldsymbol{w}^T \boldsymbol{x}_i) |\boldsymbol{w}^T \boldsymbol{x}_i|^{P-1} \boldsymbol{x}_i,$$

$$\mathrm{B} = \sum_{j=1}^{n} [\mathrm{sgn}(\boldsymbol{w}^T \boldsymbol{y}_j) \boldsymbol{w}^T \boldsymbol{y}_j]^P,$$

$$\mathrm{C} = \sum_{i=1}^{m} [\mathrm{sgn}(\boldsymbol{w}^T \boldsymbol{x}_i) \boldsymbol{w}^T \boldsymbol{x}_i]^P,$$

$$\mathrm{D} = P \sum_{j=1}^{n} \mathrm{sgn}(\boldsymbol{w}^T \boldsymbol{y}_j) \left|\boldsymbol{w}^T \boldsymbol{y}_j\right|^{P-1} \boldsymbol{y}_j,$$

$$\mathrm{E} = \left(\sum_{j=1}^{n} [\mathrm{sgn}(\boldsymbol{w}^T \boldsymbol{y}_j) \boldsymbol{w}^T \boldsymbol{y}_j]^P \right)^2$$

For the gradient in (7) to be defined, the objective function $J_{L_P}(\boldsymbol{w})$ must satisfy two conditions, i.e., $\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i \neq 0$ and $\boldsymbol{w}^{\mathrm{T}} \boldsymbol{y}_i \neq 0$. These two conditions are validated by singularity checks at the beginning of every iteration.

The overall gradient ascent algorithm to obtain the optimal spatial filter of the CSP-Lp consists of 6 steps as follows as:

1) Initialize $\boldsymbol{w}_{t=0} = \boldsymbol{w}(0)$ such that $\|\boldsymbol{w}\|_2 = 1$

2) Singularity check

If $\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i = 0$ or $\boldsymbol{w}^{\mathrm{T}} \boldsymbol{y}_i = 0$, $\boldsymbol{w}(t) = \frac{\boldsymbol{w}(t) + \boldsymbol{\delta}}{\|\boldsymbol{w}(t) + \boldsymbol{\delta}\|_2}$

where $\boldsymbol{\delta}$ is a small valued random vector.

3) Compute $\nabla_{\boldsymbol{\omega}}$ in (6)

4) Learn $\boldsymbol{w}(t+1)$ from $\boldsymbol{w}(t)$

$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) + \mu \nabla_{\boldsymbol{w}}$, where μ is a learning rate.

5) Normalize $\boldsymbol{w}(t)$

$$\boldsymbol{w}(t) = \frac{\boldsymbol{w}(t)}{\|\boldsymbol{w}(t)\|_2}$$

6) Convergence check

If $\|\boldsymbol{w}(t) - \boldsymbol{w}(t-1)\|_2 \geq \epsilon$, go to 2).

Else, $\boldsymbol{w}^* = \boldsymbol{w}(t)$. Iteration Stops.

Like the CSP-L2, the features are extracted by combining two spatial filters that maximize and minimize the objective function $J_{L_P}(\boldsymbol{w})$. A spatial filter that minimizes the objective function is like getting a spatial filter that maximizes

B. Extension to Multiple Filters

For CSP-L2, the number of spatial filters equals the number of channels of MI EEG signals and each spatial filter is orthogonal to each other. We utilize orthogonality to extend CSP-Lp into multiple spatial filters. The first spatial filter $\boldsymbol{w}_1$ is obtained through an iterative algorithm described in the previous section. The second spatial filter $\boldsymbol{w}_2$ must maximize the objective function while satisfying the condition $\boldsymbol{w}_1^{\mathbf{T}}\boldsymbol{w}_2 = 0$. We can rewrite the second spatial filter $\boldsymbol{w}_2$ as following.

$$\boldsymbol{w}_2 = (\boldsymbol{I}_C - \boldsymbol{w}_1\boldsymbol{w}_1^{\boldsymbol{T}})\boldsymbol{\alpha}_1 \tag{7}$$

where c means the number of the channel of EEG signals, $\boldsymbol{I}_C$ and $\boldsymbol{\alpha}_1$ are c-dimensional identity matrix and c-dimensional coefficient vectors, respectively. Thus, the objective function for the second spatial filter $\boldsymbol{w}_2$ is represented as

$$\frac{\sum_{i=1}^{m}|\boldsymbol{w}_2{}^T\boldsymbol{x}_i|^P}{\sum_{j=1}^{n}|\boldsymbol{w}_2{}^T\boldsymbol{y}_j|^P} = \frac{\sum_{i=1}^{m}|\boldsymbol{\alpha}_1{}^T(\boldsymbol{I}_C - \boldsymbol{w}_1\boldsymbol{w}_1^{\boldsymbol{T}})\boldsymbol{x}_i|^P}{\sum_{j=1}^{n}|\boldsymbol{\alpha}_1{}^T(\boldsymbol{I}_C - \boldsymbol{w}_1\boldsymbol{w}_1^{\boldsymbol{T}})\boldsymbol{y}_j|^P} \tag{8}$$

Using the first spatial filter $\boldsymbol{w}_1$, and $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$ are MI EEG data, the followings are calculated

$$\begin{aligned}\boldsymbol{x}_i^{(2)} &= (\boldsymbol{I}_C - \boldsymbol{w}_1\boldsymbol{w}_1^{\boldsymbol{T}})\boldsymbol{x}_i \\ \boldsymbol{y}_j^{(2)} &= (\boldsymbol{I}_C - \boldsymbol{w}_1\boldsymbol{w}_1^{\boldsymbol{T}})\boldsymbol{y}_j\end{aligned} \tag{9}$$

Substituting (9) into (8), we can rewrite eq. (8) as follows:

$$\frac{\sum_{i=1}^{m}|\boldsymbol{\alpha}_1{}^T\boldsymbol{x}_i^{(2)}|^P}{\sum_{j=1}^{n}|\boldsymbol{\alpha}_1{}^T\boldsymbol{y}_j^{(2)}|^P} \tag{10}$$

Finally, through the above iterative algorithm, we obtain $\boldsymbol{\alpha}_1$ that maximizes the objective function in (10). The second spatial filter $\boldsymbol{w}_2$ is obtained through $\boldsymbol{\alpha}_1$ as (7). In general, if the k spatial filters $\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_k$ have been obtained, the (k+1)th spatial filters $\boldsymbol{w}_{k+1}$ satisfying condition $\boldsymbol{w}_1^{\mathbf{T}}\boldsymbol{w}_{k+1} = \cdots = \boldsymbol{w}_k^{\mathbf{T}}\boldsymbol{w}_{k+1} = 0$ is obtained as follows:

$$\boldsymbol{w}_{k+1} = (\boldsymbol{I}_C - \boldsymbol{W}_k\boldsymbol{W}_k^{\boldsymbol{T}})\boldsymbol{\alpha}_k \tag{11}$$

where $\boldsymbol{W}_k$ is a c-dimension square matrix and consists of k spatial filters, ($\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_k$). The obtained $\boldsymbol{w}_{k+1}$ is normalized as $\frac{\boldsymbol{w}_{k+1}}{\|\boldsymbol{w}_{k+1}\|_2}$ and then padded into the $\boldsymbol{W}_k$.

C. Feature Extraction

When the spatial filters obtained through maximizing and minimizing the objective function are represented as $\boldsymbol{w}_1, \cdots, \boldsymbol{w}_{nx}$ and $\boldsymbol{w}_1', \cdots, \boldsymbol{w}_{ny}'$, respectively, the feature vector is given by

$$\boldsymbol{f} = \begin{pmatrix} \|\boldsymbol{w}_1^{\boldsymbol{T}}\boldsymbol{Z}\|_1 \\ \vdots \\ \|\boldsymbol{w}_{nx}^{\boldsymbol{T}}\boldsymbol{Z}\|_1 \\ \|(\boldsymbol{w}_1')^{\boldsymbol{T}}\boldsymbol{Z}\|_1 \\ \vdots \\ \left\|(\boldsymbol{w}_{ny}')^{\boldsymbol{T}}\boldsymbol{Z}\right\|_1 \end{pmatrix} \tag{12}$$

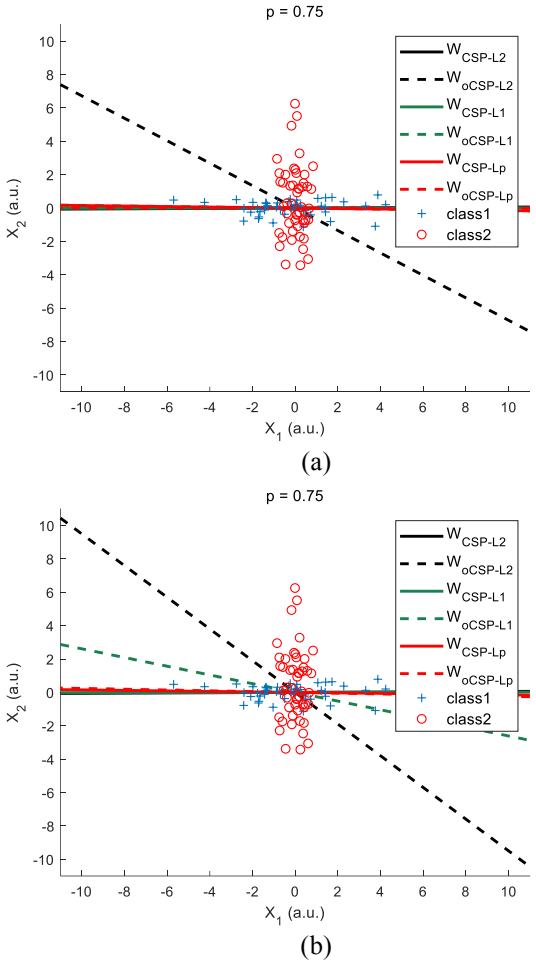

Fig. 1: A 2-D toy dataset and the spatial filters using CSP-L2, CSP-L1 and CSP-Lp. For CSP-Lp, p = 0.75 is chosen. The solid and dashed lines denote the spatial filters without and with outliers, respectively. (a) an outlier is on (20,20) position and (b) an outlier is on (60,60) position.

IV. Experiments

A. Toy Dataset

The toy dataset used for verification is two-dimensional data of two classes and is obtained through Gaussian random variance. The zero mean and covariance matrices [5, 0; 0, 0.2], [0.2, 0; 0, 5] are used. In addition, each class consists of 50 samples. The outliers were added to the class 2 data. For CSP-Lp, p = 0.75 is chosen. Fig. 1 illustrates the spatial filters in cases where distinct outliers exist. The dataset of class 1 and class 2 are represented by '+' and 'o', respectively. The solid lines demonstrate spatial filters when no outlier exists in class 2 data while the dotted lines show spatial filters when outliers exist in class 2 data. Fig. 1. (a) illustrates the spatial filters of different CSPs in the case where the outlier is on (20,20) position. We can see that CSP-Lp is comparable to CSP-Lp regardless of the outlier, while the CSP-L2 filter shows a high sensitivity to the outlier. However, when the outlier is increased on (60, 60) position as shown in Fig 1. (b), both CSP-L2 and CSP-L1 are sensitive to the outlier, leading to performance degradation. On the other hand, CSP-Lp with the outlier shows almost a similar performance in the case where

Table 1. Classification accuracy (%) of 7 subjects using CSP-L2, CSP-L1, and CSP-Lp with increasing occurrence frequency of outliers (from 0% to 50% of training dataset)

	Subject A01E							Subject A02E							Subject A04E						
Frequency	0	10	20	30	40	50	Mean	0	10	20	30	40	50	Mean	0	10	20	30	40	50	Mean
CSP-L2	85.4	83.9	**74.3**	**74.3**	**74.3**	**74.3**	**74.3**	**74.3**	**74.3**	54.7	53.6	53.1	53.7	53.5	**74.3**	66.1	59.5	56.9	54.4	53.8	60.8
CSP-L1	83.3	84.9	69.4	69.4	69.4	69.4	69.4	69.4	69.4	53.7	55.4	51.7	50.6	53.9	69.4	65.6	65.2	65.3	62.1	58.2	64.3
CSP-Lp	**91.0**	**87.9**	68.8	68.8	68.8	68.8	68.8	68.8	68.8	**56.5**	**56.8**	**58.1**	**56.5**	**57.1**	68.8	**67.6**	**65.9**	**66.7**	**66.2**	**62.0**	**66.2**
	Subject A06E							Subject A08							Subject A09						
Frequency	0	10	20	30	40	50	Mean	0	10	20	30	40	50	Mean	0	10	20	30	40	50	Mean
CSP-L2	**65.3**	**62.6**	61.8	58.3	59.7	55.1	60.5	**93.8**	93.1	93.1	93.1	**93.1**	93.1	93.2	91.7	87.6	81.2	76.1	77.0	75.6	81.5
CSP-L1	62.5	61.9	59.6	58.1	59.2	57.8	59.8	93.1	**94.4**	**94.4**	**93.8**	90.3	**94.4**	**93.4**	**92.4**	89.7	87.6	89.1	89.9	88.6	89.5
CSP-Lp	63.9	62.1	**62.6**	**62.1**	**60.4**	**62.4**	**62.2**	**93.8**	93.1	**94.4**	91.0	**93.1**	93.1	93.1	91.7	**92.4**	**91.9**	**91.2**	**91.3**	**91.0**	**91.6**

the outlier is absent. In other words, CSP-Lp achieves improved robustness to outliers compared to CSP-L2 and CSP-L1 when severe outliers occur.

B. EEG Dataset

To verify the proposed CSP-Lp using the real EEG signals, we used the BCI Competition dataset which is widely used in the BCI study. We used a dataset 2a of BCI Competition IV. The dataset consisted of nine subjects and was recorded four classes MI task, left hand, right hand, foot, and tongue MI, through 22-channel electrodes. The training dataset and test dataset are equally composed of 72 trials for each class. For a two-class classification scenario, we utilize two MI tasks, i.e., the left hand and right hand MIs. EEG signals were bandpass filtered using a fifth-order Butterworth filter from 8Hz to 30Hz to reflect the ERD/ERS pattern of MI. Based on the winner BCI Competition IV and [8, 12], we used the EEG data from 0.5 to 2.5 seconds after the cue. We used 7 subject's EEG data excluding A03E and A07E.

To validate sensitivity to outliers, we applied multivariate outliers to the training dataset. The outliers were generated with a Gaussian distribution $N_C(m + 3\sigma, 3\Sigma)$ of 22 dimensions, where m, σ, and Σ are mean, a standard deviation of each channel of training dataset, and covariance of the training dataset. The frequency of outliers was increased from 0% to 50% of the number of training samples with step 10%.

Classification performance was computed by a linear discriminant analysis (LDA) based on the feature vectors obtained in the case where $nx = ny = 3$. The classification performance obtained from 7 subjects can be seen in Table 1. As the occurrence of outliers increases, the classification accuracies of all three CSPs decreases. The proposed CSP-Lp indicates enhanced accuracies in most cases of severe outliers compared to CSP-L2 and CSP-L1 and is comparable to CSP-L1 in few cases. These results suggest that the proposed CSP-Lp is most effective for suppressing the effect of outliers than other two CSP methods.

V. CONCLUSION

We have presented a robust CSP for MI-BCI by utilizing the Lp-norm based dispersion of filtered EEG signals. The use of Lp-norm in CSP-Lp results in insensitivity to outliers. By validating the capability of the proposed CSP-Lp on synthetic toy dataset and real EEG BCI dataset, we confirm that the proposed CSP-Lp achieves improved robustness to outliers in cases where severe outliers occur.

ACKNOWLEDGMENT

This work was partly supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2022R1F1A1075043) and by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.229800, Development of quantitative emotion-sensibility evaluation model of non-face-to-face environment users and commercialization of interactive digital content application)

REFERENCES

[1] 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, Jun. 2002, doi: 10.1016/S1388-2457(02)00057-3.

[2] J. Wang, L. Bi, and W. Fei, "EEG-Based Motor BCIs for Upper Limb Movement: Current Techniques and Future Insights," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 4413–4427, 2023, doi: 10.1109/TNSRE.2023.3330500.

[3] P. Arpaia, A. Esposito, A. Natalizio, and M. Parvis, "How to successfully classify EEG in motor imagery BCI: a metrological analysis of the state of the art," *J. Neural Eng.*, vol. 19, no. 3, p. 031002, Jun. 2022, doi: 10.1088/1741-2552/ac74e0.

[4] S. Gao, Y. Wang, X. Gao, and B. Hong, "Visual and auditory brain-computer interfaces," *IEEE Trans Biomed Eng*, vol. 61, no. 5, pp. 1436–1447, May 2014, doi: 10.1109/TBME.2014.2300164.

[5] J. Mladenović, "Standardization of protocol design for user training in EEG-based brain–computer interface," *J. Neural Eng.*, vol. 18, no. 1, p. 011003, Feb. 2021, doi: 10.1088/1741-2552/abcc7d.

[6] J. Wolpaw and E. W. Wolpaw, *Brain-Computer Interfaces: Principles and Practice*. Oxford University Press, USA, 2012.

[7] S. K. Mandal and M. N. B. J. Naskar, "MI brain-computer interfaces: A concise overview," *Biomedical Signal Processing and Control*, vol. 86, p. 105293, Sep. 2023, doi: 10.1016/j.bspc.2023.105293.

[8] A. Al-Saegh, S. A. Dawwd, and J. M. Abdul-Jabbar, "Deep learning for motor imagery EEG-based classification: A review," *Biomedical Signal Processing and Control*, vol. 63, p. 102172, Jan. 2021, doi: 10.1016/j.bspc.2020.102172.

[9] X. Zhang, L. Yao, X. Wang, J. J. M. Monaghan, D. Mcalpine, and Y. Zhang, "A survey on deep learning-based non-invasive brain signals: recent advances and new frontiers," *J. Neural Eng.*, Nov. 2020, doi: 10.1088/1741-2552/abc902.

[10] H. K. Lee and Y.-S. Choi, "Application of Continuous Wavelet Transform and Convolutional Neural Network in Decoding Motor Imagery Brain-Computer Interface," *Entropy*, vol. 21, no. 12, Art. no. 12, Dec. 2019, doi: 10.3390/e21121199.

[11] F. Lotte and C. Guan, "Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2011, doi: 10.1109/TBME.2010.2082539.

[12] S. Lemm, B. Blankertz, G. Curio, and K.- Muller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 9, pp. 1541–1548, Sep. 2005, doi: 10.1109/TBME.2005.851521.

[13] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing Spatial Filters by Minimizing Within-Class Dissimilarities in Electroencephalogram-Based Brain #x2013;Computer Interface," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 4, pp. 610–619, 2013, doi: 10.1109/TNNLS.2013.2239310.

[14] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface," in *IEEE International Joint Conference on Neural Networks, 2008. IJCNN*

2008. (IEEE World Congress on Computational Intelligence), 2008, pp. 2390–2397. doi: 10.1109/IJCNN.2008.4634130.

[15] H. Wang, Q. Tang, and W. Zheng, "L1-Norm-Based Common Spatial Patterns," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 3, pp. 653–662, Mar. 2012, doi: 10.1109/TBME.2011.2177523.

[16] H. Wang and X. Li, "Regularized Filters for L1-Norm-Based Common Spatial Patterns," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 2, pp. 201–211, Feb. 2016, doi: 10.1109/TNSRE.2015.2474141.

[17] N. Kwak, "Principal component analysis by Lp-norm maximization," *IEEE Trans Cybern*, vol. 44, no. 5, pp. 594–609, May 2014, doi: 10.1109/TCYB.2013.2262936.

[18] J. H. Oh and N. Kwak, "Generalization of linear discriminant analysis using Lp-norm," *Pattern Recognition Letters*, vol. 34, no. 6, pp. 679–685, Apr. 2013, doi: 10.1016/j.patrec.2013.01.016.