



# Common Spatial Pattern with L21-Norm

Jingyu Gu<sup>1</sup> · Mengting Wei<sup>2</sup> · Yiyun Guo<sup>3</sup> · Haixian Wang<sup>1</sup>

Accepted: 17 June 2021 / Published online: 29 June 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

As one of the most commonly used algorithms in the field of feature extraction, common spatial pattern (CSP) has a good effect on multichannel electroencephalogram (EEG) signal classification, especially for motor imagery-based signals. However, the formulation of the conventional CSP based on the L2-norm is sensitive to outliers. Whereas the L1-norm-based common spatial pattern (CSP-L1) proposed in recent years can seek robust spatial filters to effectively alleviate the impact of outliers, the L1-norm is unable to characterize the geometric structure of the data well. To further improve the robustness of CSP, in this paper, we propose a new extension to CSP called the L21-norm-based common spatial pattern (CSP-L21), which is formulated by using the L21-norm rather than the L2-norm. Moreover, CSP-L21 has the advantages of rotational invariance and geometric structure characterization. We provide a non-greedy iterative algorithm to maximize the objective function of CSP-L21. Experiments on a toy example and three popular data sets of BCI competitions illustrate that the proposed method can efficiently extract discriminative features.

**Keywords** Brain-computer interfaces (BCI) · Common spatial pattern (CSP) · L21-norm · Robust feature extraction

## 1 Introduction

The brain-computer interface (BCI) is an external information communication and control technology established between the human brain and a computer or other electronic devices that is independent of the conventional brain information output pathway (peripheral nerve and pathway tissue) [1, 2]. One of the main objectives of BCI studies

---

✉ Haixian Wang  
hxwang@seu.edu.cn

<sup>1</sup> Key Laboratory of Child Development and Learning Science of Ministry of Education, School of Biological Science & Medical Engineering, Southeast University, Nanjing, Jiangsu 210096, People's Republic of China

<sup>2</sup> Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, People's Republic of China

<sup>3</sup> Qingdao Port International Company, Ltd, Qingdao, Shandong 266011, People's Republic of China

is to provide communication for patients with neuromuscular dysfunction or paralysis. The typical research goal is to facilitate the daily lives of patients. According to different kinds of signals and scenarios, BCI systems can be divided into three types: invasive, partially invasive and noninvasive [3, 4]. The former two types can cause trauma and thus are commonly used in animals or in patients with severe brain diseases. The electroencephalogram-based brain-computer interface system is widely researched and used for its noninvasive, high temporal resolution and simple equipment. However, its low spatial resolution makes it vulnerable to all kinds of noise and outliers. Therefore, feature extraction and EEG signal classification are key technologies to improve the performance of brain-computer interface system.

The BCI system generally consists of five modules, including signal acquisition, signal processing, feature extraction, classification and control signal output. The electroencephalogram (EEG) signal is very weak and has the characteristics of nonlinearity, nonstationary and noise sensitivity, which makes the effect of direct EEG signal classification poor. Therefore, the feature extraction module is very important. In the literature, feature extraction is usually considered from two aspects: temporal filtering and spatial filtering. The autoregressive model (AR) [30], power spectrum density estimation [31] and wavelet transform [32] are commonly used temporal filtering methods. Additionally, CSP and independent component analysis (ICA) [33] are the classical methods for spatial filtering. Among them, CSP is a popular approach that is widely used in data analysis of multichannel EEG signals. It aims to find the optimal spatial filters that project the EEG signals onto a subspace such that the variance ratio of the two EEG classes is maximized. It is generally known that the conventional CSP employs the covariance in terms of the L2-norm. This makes the method sensitive to noise and outliers. Given this situation, there are many extensions to CSP [22, 23] that have been put forward in recent years, such as local temporal common spatial patterns (LTCSP) [7], local temporal correlation common spatial patterns (LTCCSP), L1-norm-based CSP (CSP-L1) [29], local temporal joint recurrence common spatial patterns (LTRCSP) [21] and Lp-norm-based local temporal correlation common spatial patterns (LTCCSP-Lp) [6]. There is still room to enhance the robust modeling of the CSP method to extract more discriminative features.

In this paper, a new robust form of the CSP algorithm, which replaces the L2-norm with the L21-norm, is considered. We term this method as the L21-norm-based common spatial pattern (CSP-L21). In fact, the L21-norm has been widely used in other methods of feature extraction in the machine learning field, such as R1-PCA [8, 9], e-LDA-L21 [10], 2DLDA-L21 [12] and LPP-L21 [11]. Experimental results show that all the methods based on the L21-norm have enhanced robustness and achieved better classification performance. In addition, compared with the L1-norm, the L21-norm not only can effectively suppress outliers and noises, but also has rotation invariance and can well characterize the geometric structure of data. It is worthwhile to mention the following three innovations of CSP-L21. 1) In this robust method, the L2-norm is plugged to measure the distance along the space dimension and the L1-norm is used to sum all data points, which effectively reduces the influence of the square term on the noises and outliers. 2) A non-greedy iterative algorithm [13] is introduced to solve the optimization of the objective function of CSP-L21. It turns out to be theoretically feasible. 3) We use the L21 dispersion as features that are fed into the linear discriminant analysis (LDA) for classification.

The remainder of this paper is organized as follows. The conventional CSP is briefly reviewed in Sect. 2. In Sect. 3, we propose the CSP-L21 method. In addition, a non-greedy iterative algorithm is introduced and theoretically justified. Section 4 presents the experimental

results and discussions on a toy example and three EEG data sets. Finally, we provide concluding remarks in Sect. 5.

## 2 Brief Review of CSP

In the field of EEG processing, CSP is a spatial filtering algorithm, which is used to extract the spatial features of multichannel EEG signals [24, 25]. It is generally used for two categories. Let  $X^1, X^2, \dots, X^{t_x} \in R^{C \times N}$  be the motor imagery (MI)-based EEG signals that belong to one mental task and  $Y^1, Y^2, \dots, Y^{t_y} \in R^{C \times N}$  be the condition of the other class. Here,  $C$  is the number of channels,  $N$  is the number of samples in a single EEG trial, and  $t_x$  and  $t_y$  represent the numbers of trials from the two kinds of EEG signals. We relabel the columns of  $X$  and  $Y$  as  $X = (x_1, x_2, \dots, x_m) \in R^{C \times m}$  and  $Y = (y_1, y_2, \dots, y_n) \in R^{C \times n}$ , respectively. Here,  $m = N \times t_x$  and  $n = N \times t_y$ . They represent the numbers of samples in the total EEG trials from the two classes.

The EEG signals then go through the filtering of a specific frequency band, the decentralization of the mean value and the preprocessing of normalization [26]. For simplicity of expression, we assume that the above symbols represent the preprocessed EEG data matrix rather than the original EEG signals. Then, the formulation of the objective function is given as follows:

$$J_{CSP}(w) = \frac{w^T C^x w}{w^T C^y w} \tag{1}$$

where  $w \in R^C$  is a spatial filter that projects multichannel EEG signals into a new space such that the variance of one class is maximized while that of the other class is minimized. Here,  $C^x \in R^{C \times C}$  and  $C^y \in R^{C \times C}$  are the covariance matrices of the two classes, given by Eqs. (2) and (3), respectively:

$$C^x = \frac{1}{t_x} X X^T \tag{2}$$

$$C^y = \frac{1}{t_y} Y Y^T \tag{3}$$

Our aim is to determine the spatial filter  $w$  by solving the generalized eigenvalue equation:

$$C^x w = \lambda C^y w \tag{4}$$

For classification, we select the smallest number of leading eigenvectors associated with the largest and the smallest eigenvalues as spatial filters. The variances of the filtered signals from the two classes (possibly after a log-transformation) are used as features sets, which are fed into the classifier of the linear discriminant analysis (LDA).

### 3 L21-Norm-Based Common Spatial Pattern (CSP-L21)

With the widely increasing application of CSP, many problems have come to light. The random non-stationarity and noninvasive acquisition way of EEG signals lead to noise caused by electromyograms (EMGs), electrooculograms (EOGs) and spikes mixed in the signals [14]. To solve this problem effectively has always been the focus of research.

It is generally known that many effective methods have done well in alleviating the impact of outliers in recent years, such as CSP-L1 [5], CSP-L1 with the form of waveform length [19], improving generalization CSP-L1 [28], even regularized CSP-L1 [18, 20] and so on. However, the L1-norm cannot well characterize the geometric structure of data and does not obey the rotational invariance. To obtain more discriminative features, we consider it from the perspective of group sparsity. A new extension to CSP, called the L21-norm-based common spatial pattern (CSP-L21), is proposed in this paper.

#### 3.1 Objective Function

According to Eq. (1), Eqs. (2) and (3), the objective function of classical CSP can be further reformulated as:

$$J_{CSP}(w) = \frac{w^T C^x w}{w^T C^y w} = \frac{\frac{1}{I_x} \|w^T X\|_2^2}{\frac{1}{I_y} \|w^T Y\|_2^2} = \frac{\frac{1}{I_x} \sum_{i=1}^m (w^T x_i)^2}{\frac{1}{I_y} \sum_{j=1}^n (w^T y_j)^2} \tag{5}$$

where  $\|\cdot\|$  denotes the L2-norm. Clearly, the square term potentially magnifies the effect of outliers. Motivated by the advantages of the L21-norm, we propose the objective function given by:

$$J_{CSP-l_{2,1}}(W) = \frac{\|W^T X\|_{2,1}}{\|W^T Y\|_{2,1}} = \frac{\sum_{i=1}^m \|W^T x_i\|_2}{\sum_{j=1}^n \|W^T y_j\|_2} \tag{6}$$

where  $W \in R^{C \times d}$  ( $d < C$ ) is an optimal projection matrix which projects the samples into the lower  $d$ -dimension subspace. Here,  $d$  denotes the number of projection vectors, i.e., the number of extracted features that are set as the input to the linear discriminant analysis (LDA) for classification, and  $\|\cdot\|_{2,1}$  represents the L21-norm. For an arbitrarily matrix  $U$  ( $U \in R^{a \times b}$ ),  $\|U\|_{2,1}$  is defined as follows:

$$\|U\|_{2,1} = \sum_{j=1}^b \sqrt{\sum_{i=1}^a U_{ij}^2} = \sum_{j=1}^b \|u_j\|_2 \tag{7}$$

where  $U_j$  is the  $j$  column of  $U$ .

In order to solve the objective function (6) more conveniently, the following corollary is introduced.

Corollary 1: Objective function (6) is equal to the following formulation:

$$J_{CSP-l_{2,1}}(W) = \frac{tr(W^T X D_x X^T W)}{tr(W^T Y D_y Y^T W)} \tag{8}$$

$$D_x = \text{diag}\left(\frac{1}{\|x_1\|_2}, \frac{1}{\|x_2\|_2}, \dots, \frac{1}{\|x_d\|_2}\right) \tag{9}$$

$$D_y = \text{diag}\left(\frac{1}{\|y_1\|_2}, \frac{1}{\|y_2\|_2}, \dots, \frac{1}{\|y_d\|_2}\right) \tag{10}$$

where  $tr(\cdot)$  is the trace operator.

Proof: According to Eq. (6), the following equation is obtained by using simple algebraic theory:

$$\|U\|_{2,1} = \sum_{j=1}^b \|u_j\|_2 = \sum_{j=1}^b \frac{u_j^T u_j}{\|u_j\|_2} = \text{tr}(UDU^T) \tag{11}$$

$$D = \text{diag}\left(\frac{1}{\|u_1\|_2}, \frac{1}{\|u_2\|_2}, \dots, \frac{1}{\|u_b\|_2}\right) \tag{12}$$

According to Eq. (11), we have:

$$\|W^T X\|_{2,1} = \text{tr}(W^T X D_x X^T W) \tag{13}$$

$$\|W^T Y\|_{2,1} = \text{tr}(W^T Y D_y Y^T W) \tag{14}$$

Substituting Eqs. (13) and (14) into Eq. (6), objective function (6) can finally be rewritten as:

$$J_{CSP-l_{2,1}}(W) = \frac{\|W^T X\|_{2,1}}{\|W^T Y\|_{2,1}} = \frac{\text{tr}(W^T X D_x X^T W)}{\text{tr}(W^T Y D_y Y^T W)} \tag{15}$$

### 3.2 Iterative algorithm

Computationally, finding the derivative of objective function (15) about  $W$  and letting its reciprocal be zero is extremely complex and difficult. In addition, different from the form of the traditional trace ratio, the matrices  $D_x$  and  $D_y$  in Eq. (15) are related to the projection matrix  $W$ . Thus, we decided to use a non-greedy iterative algorithm which constructs an auxiliary function, with the help of the sub-gradient algorithm and Armijo line search method to obtain optimal projection matrix  $W$ .

Before formally settling the problem, we introduce a theorem as follows [10, 11]:

Theorem 1: Suppose that the matrix functions  $M(U)$  and  $N(U)$  are positive definite, we have:

$$\lambda_{\max} = \frac{M(U^*)}{N(U^*)} = \max_{U^T U = I_p} \frac{M(U)}{N(U)} \tag{16}$$

if and only if:

$$M(U^*) - \lambda_{\max}N(U^*) = \max(M(U) - \lambda_{\max}N(U)) = 0 \tag{17}$$

Consequently, this theorem provides an auxiliary function for the objective optimization.

According to Eqs. (16) and (17), we invert objective function (15) into the following corresponding trace difference objective function:

$$W_{opt} = \arg \max_{W^T W = I_d} \frac{\|W^T X\|_{2,1}}{\|W^T Y\|_{2,1}} = \arg \max_{W^T W = I_d, \lambda} (M(W) - \lambda N(W)) \tag{18}$$

$$M(W) = \|W^T X\|_{2,1} \tag{19}$$

$$N(W) = \|W^T Y\|_{2,1} \tag{20}$$

We see that two unknown variables  $D_x$  and  $D_y$  are included in objective function (18). We therefore resort to alternatively updating  $W$  (while fixing  $\lambda$ ) and  $\lambda$  (while fixing  $W$ ).

Specifically, we assume that  $W_{t-1}$  is the solution at the  $(t - 1)$  th iteration. Meanwhile,  $\lambda_t$  is calculated as follows:

$$\lambda_t = \frac{M(W_{t-1})}{N(W_{t-1})} \tag{21}$$

According to Theorem 1 and Eq. (18),  $\lambda_t$  can be fixed by optimizing the following formula:

$$\arg \max_{W^T W = I_d} F(W) \tag{22}$$

where

$$F(W) = M(W) - \lambda_t N(W) \tag{23}$$

Obviously,  $\lambda_t$  ( $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{t-1} \leq \lambda_t$ ) is a sequence. It is concluded that the problem can be transformed into the inequality (24), which is settled by the Armijo line search method:

$$F(W) = M(W) - \lambda_t N(W) \geq M(W_{t-1}) - \lambda_t N(W_{t-1}) = 0 \tag{24}$$

We consider applying a sub-gradient algorithm to solve Eq. (22). The sub-gradient of  $F(W)$  is achieved by calculating the derivative of Eq. (23) with respect to  $W$ :

$$\nabla F(W) = \nabla M(W) - \lambda_t \nabla N(W) \tag{25}$$

$$\nabla M(W) = \frac{\partial M}{\partial W} = \frac{\partial(\sum_j \|W^T X(:,j)\|_2)}{\partial W} = \sum_j \frac{1}{\|W^T X(:,j)\|_2} X(:,j)X(:,j)^T W = X D_x X^T W \tag{26}$$

where

$$D_x = \text{diag}\left(\frac{1}{\|W^T X(:,1)\|_2}, \frac{1}{\|W^T X(:,2)\|_2}, \dots, \frac{1}{\|W^T X(:,d)\|_2}\right) \tag{27}$$

$$\nabla N(W) = \frac{\partial N}{\partial W} = \frac{\partial(\sum_j \|W^T Y(:,j)\|_2)}{\partial W} = \sum_j \frac{1}{\|W^T Y(:,j)\|_2} Y(:,j) Y(:,j)^T W = Y D_y Y^T W \tag{28}$$

where

$$D_y = \text{diag}\left(\frac{1}{\|W^T Y(:,1)\|_2}, \frac{1}{\|W^T Y(:,2)\|_2}, \dots, \frac{1}{\|W^T Y(:,d)\|_2}\right) \tag{29}$$

Substituting Eqs. (26) and (28) into Eq. (25), the sub-gradient of  $F(W)$  finally becomes:

$$\nabla F(W) = \nabla M(W) - \lambda_t \nabla N(W) = X D_x X^T W - \lambda_t Y D_y Y^T W \tag{30}$$

where  $D_x$  and  $D_y$  are represented as Eqs. (27) and (29), respectively.

Furthermore, to obtain more discriminative features, the orthogonalization constraint of  $W$  must be satisfied. In each iteration,  $W$  must be projected into an orthogonal cone by using Eq. (31) for the new projection matrix  $W_{new}$ :

$$W_{new} = H(W) = W(W^T W)^{-\frac{1}{2}} \tag{31}$$

The pseudocode of the entire solution process is shown in Table 1.

### 3.3 Algorithm validation

It is necessary to prove that the previous iterative algorithm for optimizing the objective function of CSP-L21 is convergent. For this purpose, we verify that the objective function is monotonically increasing in each step and has an upper bound as follows.

**Table 1** Iterative algorithm procedure of CSP-L21

---

Input:  
 Two classes of EEG data matrices  $X$  and  $Y$ , Armijo line search parameter  $\beta$  ( $0 < \beta < 1$ ), and iteration number  $t$ . Initialize  $W_0 \in R^{C \times d}$  randomly, satisfying  $W_0^T W_0 = I_d$

Output:  
 Optimal spatial filter matrix  $W_t$

Procedure:  
 1) Set  $t = 1$   
 2) Suppose the solution of the  $(t-1)$  th iteration has been obtained, which is  $W_{t-1}$ . Calculate  $\lambda_t$  by Eqs. (19)–(21) as  $\lambda_t = \frac{M(W_{t-1})}{N(W_{t-1})} = \frac{\|(W_{t-1})^T X\|_{2,1}}{\|(W_{t-1})^T Y\|_{2,1}}$   
 3) Calculate  $\nabla F(W_{t-1})$  by Eq. (30), where  $D_x$  and  $D_y$  are represented as Eqs. (27) and (29), respectively  
 4) Update  $W_{new} = H(W_{t-1} + \beta \cdot \nabla F(W_{t-1}))$ , which satisfies Eq. (31)  
 5) Calculate  $F(W_{new})$  by Eq. (23). If  $F(W_{new}) \geq F(W_{t-1}) = 0$ , let  $W_t = W_{new}$  and go to step 6; otherwise update  $\beta = \frac{\beta}{1+\alpha}$  ( $\alpha \in [0, 1]$ )  $[0, 1]$  and go to step 4  
 6) Set  $t \leftarrow t + 1$   
 7) Stop the iteration until the objective function denoted by Eq. (15) converges  
 8) Return the final spatial filter matrix  $W_t$

---

Supposing  $W_t$  is the solution at the  $t$ -th iteration, the formula  $F(W_t) \geq F(W_{t-1}) = 0$  is valid. Then, we have:

$$F(W) = M(W) - \lambda_t N(W) = \|W^T X\|_{2,1} - \lambda_t \|W^T Y\|_{2,1} \geq 0 \tag{32}$$

Let  $J(W_t) = \frac{M(W_t)}{N(W_t)}$ . Because  $N(W_t)$  is positive definite, namely,  $N(W_t) > 0$ , according to Eq. (32), we have:

$$J(W_t) = \frac{M(W_t)}{N(W_t)} \geq \lambda_t \tag{33}$$

Substituting Eq. (21) into Eq. (33), Eq. (34) is given as:

$$J(W_t) = \frac{M(W_t)}{N(W_t)} \geq \frac{M(W_{t-1})}{N(W_{t-1})} = J(W_{t-1}) \tag{34}$$

Therefore, the conclusion that the objective function of CSP-L21 is monotonically increasing in each interaction step is shown. Next, the fact that the objective function has an upper bound is justified.

According to  $\sqrt{a^2 + b^2 + c^2} \leq |a| + |b| + |c|$ , in the numerator of the objective function, we have:

$$\|W^T X\|_{2,1} \leq \sum_{j=1}^N \|W^T X(:,j)\|_1 = \sum_{j=1}^N \sum_{k=1}^d |W(:,k)^T X(:,j)| \tag{35}$$

Through the Cauchy inequality  $|\langle x, y \rangle| \leq \|x\|_2 \|y\|_2$ , it is shown that:

$$\sum_{j=1}^N \sum_{k=1}^d |W(:,k)^T X(:,j)| \leq \sum_{j=1}^N d \|X(:,j)\|_2 \tag{36}$$

$$\|W^T X\|_{2,1} \leq \sum_{j=1}^N d \|X(:,j)\|_2 \tag{37}$$

On the basis of  $\sqrt{a^2 + b^2 + c^2} \leq |a| + |b| + |c|$ , we have:

$$\|W^T X\|_{2,1} = \sum_{j=1}^N \|W^T X(:,j)\|_2 \geq \sqrt{\sum_{j=1}^N \|W^T X(:,j)\|_2^2} = \sqrt{\text{tr}(W^T X X^T W)} \tag{38}$$

Due to the matrix theory:

$$\text{tr}(W^T X X^T W) \geq \sum_{i=1}^d \eta_i (\eta_1 \leq \eta_2 \leq \dots \eta_d) \tag{39}$$

We have:

$$\|W^T X\|_{2,1} \geq \sum_{i=1}^d \eta_i \tag{40}$$

where  $\eta_i$  is the eigenvalue of symmetric matrix  $X X^T$ . Combining Eqs. (37) and (40), we have:



$$\sum_{i=1}^d \eta_i \leq \|W^T X\|_{2,1} \leq \sum_{j=1}^N d \|X(:,j)\|_2 \tag{41}$$

Similarly, for the denominator of the objective function, we still have:

$$\sum_{i=1}^d \gamma_i \leq \|W^T Y\|_{2,1} \leq \sum_{j=1}^N d \|Y(:,j)\|_2 \tag{42}$$

According to Eqs. (41) and (42), it is justified that the objective function has an upper bound. Thus, the objective function is monotonically increasing in each step and has an upper bound. Namely, the proposed iterative procedure for the objective function of CSP-L21 is convergent.

### 3.4 Rotational invariance

Commonly, there are two main aspects of rotational invariance. One is that the projection matrix  $W$  will rotate to  $\Gamma W$  after the rotation of the feature space  $\Gamma$ . The other is that the result of the data projection in a high-dimensional space will remain unchanged when the sample space rotates. Then the rotational transformations can be defined by the rotation matrix  $\Gamma \in R^{C \times C}$  as follows:

$$X \leftarrow \Gamma X Y \leftarrow \Gamma Y x_i^j \leftarrow \Gamma x_i^j y_i^j \leftarrow \Gamma y_i^j W \leftarrow \Gamma W \Gamma^T \Gamma = I \tag{43}$$

According to objective function (6), Eq. (6) can be rewritten as:

$$J_{CSP-L21}(W) = \frac{\|W^T X\|_{2,1}}{\|W^T Y\|_{2,1}} = \frac{\sum_{i=1}^m \|W^T x_i\|_2}{\sum_{j=1}^n \|W^T y_j\|_2} = \frac{\sum_{i=1}^m \|W^T \Gamma^T x_i\|_2}{\sum_{j=1}^n \|W^T \Gamma^T y_j\|_2} = \frac{\sum_{i=1}^m \|W^{\wedge T} \Gamma x_i\|_2}{\sum_{j=1}^n \|\hat{W}^T \Gamma y_j\|_2} \tag{44}$$

where  $\hat{W} = \Gamma W$ .

If  $W$  is the solution of objective function (6), then  $\hat{W}$  is the solution of objective function (6) after a rotational transformation by rotation matrix  $\Gamma \in R^{C \times C}$ . That is, we have:

$$[\hat{W}^T \Gamma x_i = W^T x_i, \hat{W}^T \Gamma y_j = W^T y_j] \tag{45}$$

In summary, the low-dimensional feature extracted by the rotation transformation objective function remains unchanged under the rotational transformation.

### 3.5 Geometric structure

The geometric structure of the classical CSP is represented by the covariance matrices, as shown in Eq. (1). By contrast, the objective function of CSP-L21 is described by Eq. (15) while the solution is mainly obtained by Eq. (30). The objective function of CSP-L21 is defined by  $X D_x X^T$  and  $Y D_y Y^T$  while the solution is also related to  $X D_x X^T$  and  $Y D_y Y^T$ .

These two matrices are in fact the weighted data covariance matrices. That is, CSP-L21 preserves the geometry of the data well.

### 3.6 Feature extraction

Through the non-greedy iterative algorithm above, we obtain the optimal projection matrix  $W$ . For feature extraction, we relabel the columns of  $W$  as a set of orthonormal spatial filters  $w_1, w_2, \dots, w_d$ . Thus, for any EEG trial  $Z$ , the feature is extracted as:

$$\left[ \left[ f = (\|w_1 Z\|_2, \|w_2 Z\|_2, \dots, \|w_d Z\|_2)^T \right] \right] \quad (46)$$

where  $d$  is the number of spatial filters and  $f$  is a  $d$ -dimensional feature vector for training a classifier.

## 4 Experiments

In this section, we build a toy data set and introduce the outliers first for preliminary verification. Afterwards, three public BCI competition EEG data sets, data sets IIIa and IVa of BCI competition III and data set IIa of BCI competition IV, are used to compare the effectiveness of the proposed CSP-L21 approach and the other extensions to the CSP methods. To verify the robustness, which is the main contribution of this paper, outliers of different frequencies are introduced.

### 4.1 Description of Data Sets

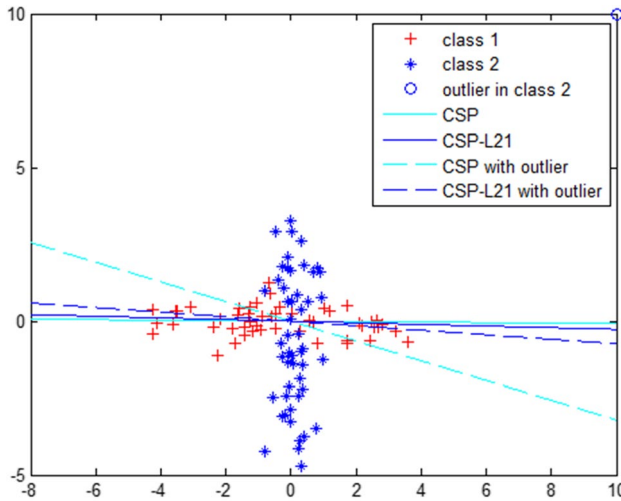
#### 4.1.1 Toy Example

The 2-D artificial data with 50 points per class are generated from two Gaussian classes that are set up with zero means and covariance matrices of  $[5, 0; 0, 0.2]$  and  $[0.2, 0; 0, 5]$ , respectively. We aim to examine the difference in the projection directions when the data set is with and without outliers. The two classes of samples are specified by “+” in red and “\*” in blue. For testing the performance of CSP and CSP-L21 under the influence of outliers, an outlier, 10 is introduced to class “\*” by using the blue “o”. The spatial filters of CSP and CSP-L21 are optimal when maximizing the filtered scatter of class “+” while minimizing the other class “\*”. The directions of the filters are shown in Fig. 1.

After introducing the outliers, the deviation angle of the filter direction of the CSP algorithm is larger than that of the CSP-L21 algorithm, which effectively proves the robustness of CSP-L21. The above is merely a preliminary experiment, and thus, we will further prove the result on the real data sets.

#### 4.1.2 Real EEG Data Sets

1. Data set IIIa of BCI competition III: This data set contains EEG signals recorded from three individuals, s1, s2 and s3, by using 60 channels. Specifically, 90, 60 and 60 trials



**Fig. 1** A toy example of spatial filtering by using CSP and CSP-L21 on a 2-D data set. Data points and projection vectors obtained by CSP and CSP-L21 in the cases without and with outliers are shown

are used for training and testing for s1, s2 and s3, respectively. We set the sampling frequency as 250 Hz and focus on the classification of the left and right hands MI.

2. Data set IVa of BCI competition III: The EEG signals contain five subjects, aa, al, av, aw and ay, are down-sampled at 100 Hz for analysis with 118 channels. The numbers of trials used for training for each subject are 168, 224, 84, 56, and 28, respectively, while the numbers of trials used for testing are 112, 56, 196, 224 and 252, respectively. Our aim is to classify the EEG signals of right hand and foot MI.
3. Data set IIa of BCI competition IV: The EEG data gathered from 22 electrodes are constituted by collecting EEG signals from nine people: A01E–A09E. Two sessions recorded on different days are provided. Each session consists of 288 trials with 72 trials per class. The signals are sampled at 250 Hz and the motor imageries of the left and right hand are considered for classification in this paper.

As shown in Table 2, we summarize the information of the three real data sets.

**Table 2** Statistical information of three real EEG data sets for experiment

Data set	BCI competition III								BCI competition IV
	Data set IIIa			Data set IVa					Data set IIa
Subject	s1	s2	s3	aa	al	av	aw	ay	A01E-A09E
No. of training trials	90	60	60	168	224	84	56	28	144
No. of testing trials	90	60	60	112	56	196	224	252	144
No. of total trials	180	120	120	280	280	280	280	280	288
No. of channels	60			118					22
Sample frequency	250 Hz			100 Hz					250 Hz
MI classified	Left vs. right hand			Right hand vs. foot					Left vs. right hand

## 4.2 Preprocessing of EEG signals

Before the experiment, the original EEG signals from the three real data sets are preprocessed first. The raw signals are filtered with a cutoff frequency of 8–35 Hz composing both the  $\alpha$ -band and the  $\beta$ -band by a fifth order Butterworth filter. It should be noted that, when the order increases, the slope of the filter decodes. Thus the speed of cutoff will be faster as the order becomes larger. In particular, when the order becomes infinite, the gain becomes a rectangular function. Therefore, according to the suggestion of Lotte and Guan [15], the order of Butterworth filter is set to five. For the first and third data sets, time segments from 0.5 to 2.5 s after the visual cues are chosen. Additionally, following the winner of BCI competition IV, the EEG segments recorded from 0.5 to 3.75 s after the visual cue are selected for the second data set.

## 4.3 Experimental Settings

There are two parameters in the line search during the calculation. Empirically, the set  $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$  for  $\beta$  is selected on an approximate scale while  $\alpha$  randomly takes a value between 0 and 1 in each interaction. Because the random value increases the uncertainty, we decide to run the program ten times in succession to ensure the stability and superiority of the result. As a method of comparison, the regularization parameter of TRCSP determined by ten-fold cross validation is searched in the set  $\{1e-6, 1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1e1, 1e2\}$ .

In addition, unlike previous research on extensions to the CSP algorithm, the pairs of filters termed as  $m$  are not fixed here. We observe the changes in accuracies as  $m$  varies from 1 to  $0.5 \times C$ , where  $C$  represents the number of electrodes (i.e., the channels), rather than using one signal value. Finally, we obtain a  $d$ -D feature vector that is set as an input to the linear discriminant analysis (LDA) for classification, where  $d$  is the number of filters. The highest accuracy is achieved during the process. Besides, in order to save iteration time and ensure effective search, this paper sets the initial value of  $W$  as the basic space filter obtained by the traditional CSP algorithm.

## 4.4 Outlier Simulation

Outliers with different frequencies generated from a  $C$ -dimensional Gaussian distribution  $N(m + 3\sigma, \Sigma)$ , where  $m$  is the mean vector of EEG training data,  $\sigma$  is the standard deviation vector of the EEG training data and  $\Sigma$  is the covariance matrix of the EEG training data, are introduced to the data sets for verifying the robustness of the algorithm we propose in this paper. Here, we compare the recognition rates with frequencies from the list  $\{0.1, 0.2, 0.3, 0.4, 0.5\}$ .

## 4.5 Results and Discussion

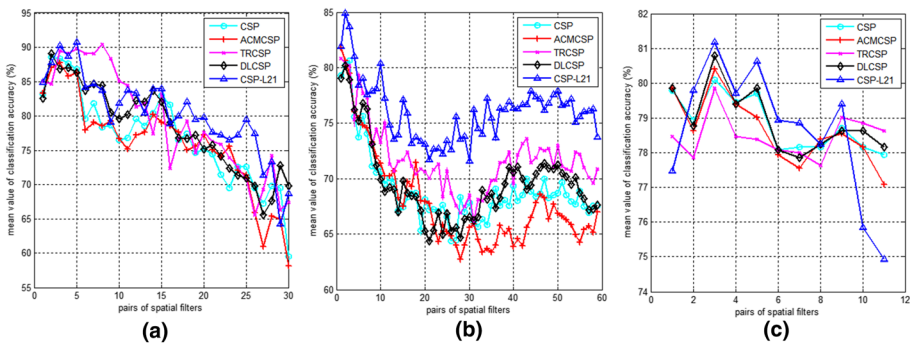
Other than the classical CSP method, other extensions used for comparison include TRCSP (the regularized CSP with Tikhonov regularization) [15], ACMCSP, which uses the weighted average covariance matrix to replace the conventional covariance matrix [16], and DLCSP, which learns the regularized CSP filters based on diagonal loading (DL) to discriminate the two mental states in the EEG signals [17]. Figure 2 shows the changes in

**Table 3** Classification accuracies of CSP, ACMCSP, TRCSP, DLCSP and CSP-L21 on the subjects of the data sets IIIa and IVa of BCI competition III without outliers added

Data set	BCI competition III									
	Data set IIIa				Data set IVa					
Subject	s1	s2	s3	Mean	aa	al	av	aw	ay	Mean
CSP	<b>98.89</b>	68.33	<b>98.33</b>	88.52	69.64	<b>100.00</b>	65.31	<b>93.75</b>	74.60	80.66
ACMCSP	96.67	70.00	96.67	87.78	65.18	<b>100.00</b>	71.94	92.86	78.97	81.79
TRCSP	94.44	<b>80.00</b>	96.67	90.37	65.18	98.21	<b>72.45</b>	89.73	78.57	80.83
DLCSP	<b>98.89</b>	70.00	<b>98.33</b>	89.07	71.43	91.07	65.31	92.86	80.16	80.16
CSP-L21	<b>98.89</b>	75.00	<b>98.33</b>	<b>90.74</b>	<b>78.57</b>	<b>100.00</b>	66.33	91.96	<b>87.70</b>	<b>84.91</b>
BCI winner	<u>0.82</u>	<u>0.76</u>	<u>0.80</u>	<u>0.79</u>	95.50	100.00	80.60	100.00	97.60	94.74

The BCI winner values on the data set IIIa of BCI competition III are the Kappa scores

The bold values represent the highest recognition rates for each subject



**Fig. 2** The changes of average classification accuracies with pairs of spatial filters on **a** data set IIIa of BCI competition III, **b** data set IVa of BCI competition III and **c** data set IVa of BCI competition IV by five different extensions to CSP: classical CSP, ACMCSP, TRCSP, DLCSP and CSP-L21

the average classification accuracies with the pairs of spatial filters in three real data sets by using the above five methods. It can be seen that the mean values of the accuracy are more than 55%, 60% and 74%, respectively, and the recognition rates of CSP-L21 not only are higher than that of the other methods in most cases but also obtain the best accuracy.

Moreover, we can achieve the corresponding filter pairs of the five methods when the optimal recognition rates are reached in the three real data sets. For data set IIIa of the BCI competition III, which has three subjects, the optimal spatial filter pairs used to classify the motor imagery-based signals for the five methods CSP, ACMCSP, TRCSP, DLCSP and CSP-L21 are 2, 3, 8, 2 and 5, respectively. On data set IVa of the BCI competition III, which has five subjects, 3, 1, 1, 2, and 2 filter pairs are selected for the five methods mentioned above. Finally, applying three filter pairs results in the best accuracy for all the above methods on data set IVa of BCI competition IV.

Then, we calculate the best recognition rates when these methods apply the corresponding pairs of filters and summarize them in Tables 3 and 4. In addition, the results of the BCI winners are also added as reference resources. It should be mentioned that the values of the BCI winners underlined are the Kappa scores for multi-category classification. Moreover,

**Table 4** Classification accuracies of CSP, ACMCSP, TRCSP, DLCSP and CSP-L21 on subjects of the data set IIa of BCI competition IV without outliers added

Data set	BCI competition IV									
	Data set IIa									
Subject	A01E	A02E	A03E	A04E	A05E	A06E	A07E	A08E	A09E	Mean
CSP	89.58	<b>61.81</b>	<b>96.53</b>	73.61	67.36	65.28	80.56	94.44	91.67	80.09
ACMCSP	89.58	59.03	<b>96.53</b>	72.92	<b>72.92</b>	68.06	78.47	94.44	91.67	80.40
TRCSP	<b>90.97</b>	58.33	<b>96.53</b>	64.58	70.14	68.06	<b>84.72</b>	94.44	90.97	79.86
DLCSP	89.58	61.11	<b>96.53</b>	73.61	68.75	68.06	82.64	95.14	91.67	80.79
CSP-L21	<b>90.97</b>	56.25	<b>96.53</b>	<b>76.39</b>	68.75	<b>68.75</b>	82.64	<b>97.22</b>	<b>93.06</b>	<b>81.17</b>
BCI winner	<u>0.68</u>	<u>0.42</u>	<u>0.75</u>	<u>0.48</u>	<u>0.40</u>	<u>0.27</u>	<u>0.77</u>	<u>0.75</u>	<u>0.61</u>	<u>0.57</u>

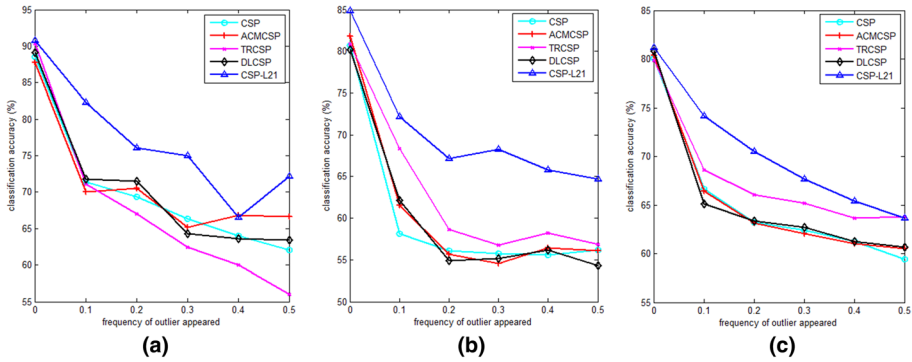
The BCI winner values are the Kappa scores

The bold values represent the highest recognition rates for each subject

the recognition rates, which the BCI winners achieved on the data set IVa of BCI competition III, are obviously higher than the above five methods due to the adoption of a complex ensemble classifier instead of applying a single algorithm. In addition, the BCI winners' preprocessing of the raw signals is also different from the other methods. Therefore, these data are not included in the comparison, only for the integrity of the results.

Although the results of these methods are different throughout the 17 subjects, they all have good performance. This is reflected in the classification accuracies of all individuals, which are all more than 55%. In particular, for some subjects, such as s1, s3, al, aw, A03E, A08E and A09E, the accuracy rates exceed 90%, and some of the rates even reach 100%. Clearly, the classical CSP method also has stable and satisfactory results on subjects s1, s3, al, aw, A02E, A03E and A08E. ACMCSP and DLCSP all work on modifying the conventional covariance matrix with distinct expressions for more distinguished features. This is essentially similar to the idea proposed in our study. The former performs well in subjects al, aw, A03E and A05E, while the latter performs well on subjects s1, s3, ay, A03E and A08E, which proves the effectiveness of the two methods. TRCSP actually adds the L2-regularization constraint to the objective function for the sake of sparsity and robustness of the final results. By observing Tables 3 and 4, we find that this method with regularization has its advantages on the subjects of data set IIa of BCI competition IV. Last but not least, the results of CSP-L21 are superior to the results of the other methods for most instances, and compared with the classical CSP, the mean classification accuracies of CSP-L21 increase by approximately 2.22, 4.25 and 1.08%. CSP-L21 always achieves the highest average recognition rates of all of the comparison methods. In addition, the recognition rate of the subject ay, whose training set included only 28 trials, reaches 87.7%. This shows that CSP-L21 can also be applied to small-sample training sets. This method improves the accuracies of six of the nine subjects. The experiments on the three real data sets without added outliers demonstrate the effectiveness of the CSP-L21 method.

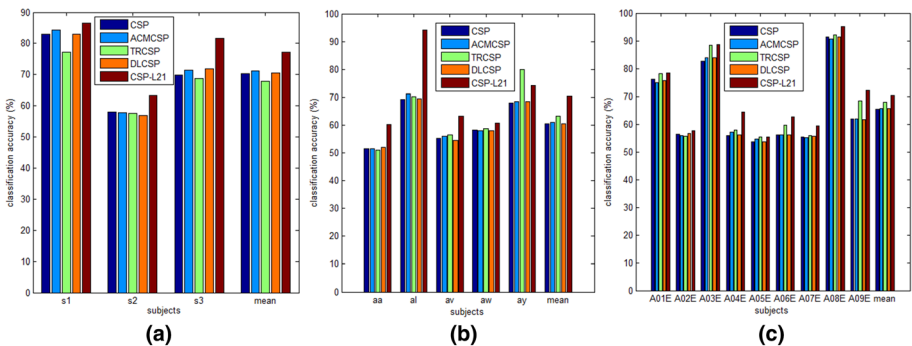
Afterwards, to check the robustness of CSP-L21, outliers varied between 0.1~0.5 by step=0.1 are added to the raw signals of the three real data sets. For each data set, we draw the curve of the average recognition rates of the subjects varying with the outliers' frequencies in Fig. 3. We observe that the performance of these methods deteriorates as the frequency of the occurring outliers increases. However, it is shown that CSP-L21 always



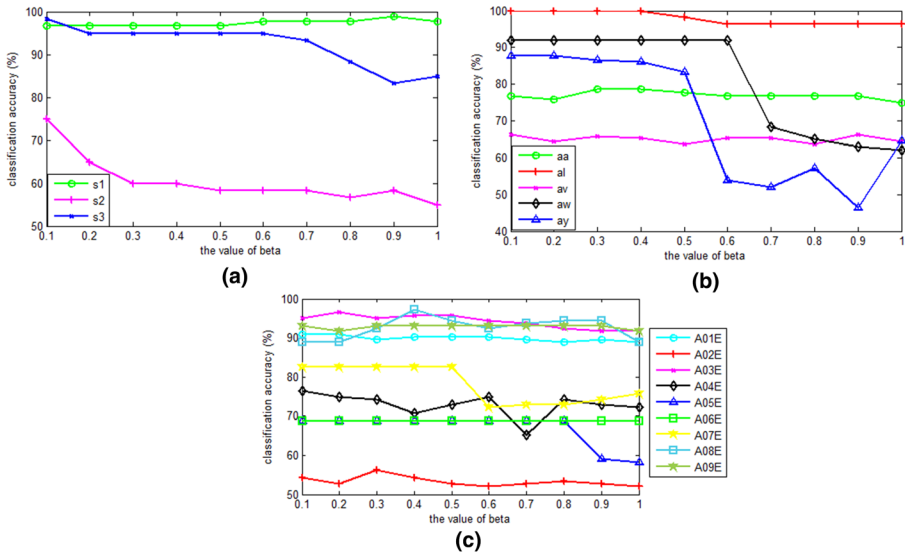
**Fig. 3** Average classification accuracies of CSP, ACMCSP, TRCSP, DLCSP and CSP-L21 for the subjects of the three real EEG data sets with outliers added. The frequencies of the outliers are 0.1, 0.2, 0.3, 0.4 and 0.5. **a** Data set IIIa of BCI competition III. **b** Data set IVa of BCI competition III. **c** Data set IIa of BCI competition IV

maintains the best discrimination (almost more than 65%) on the EEG data sets, while the other methods almost cease to be effective, especially on the second data set with five subjects. Similarly, the role of Fig. 4 is to verify the robustness of the CSP-L21 method again by showing the average recognition rates for each subject and depicting the average classification accuracies across the different subjects of each EEG data set. ACMCSP, TRCSP and DLCSP behave well in some cases, such as for subject ay and A09E, which illustrates that the extensions to CSP in the previous articles exhibit good performance to some extent. However, CSP-L21 clearly has the best performance among all compared methods. The accuracies of the proposed method for subjects s1, al, A03E and A08E reach approximately 90%, and some accuracies are close to 100%. According to the figures and analysis above, this paper concludes that the CSP-L21 method is valid and able to effectively reduce the impact of outliers.

Next, we discuss some details. In the non-greedy iterative algorithm from Table 1, line search parameter  $\beta$  belonging to the interval 0 and 1 is uncertain. For simplicity, we set the



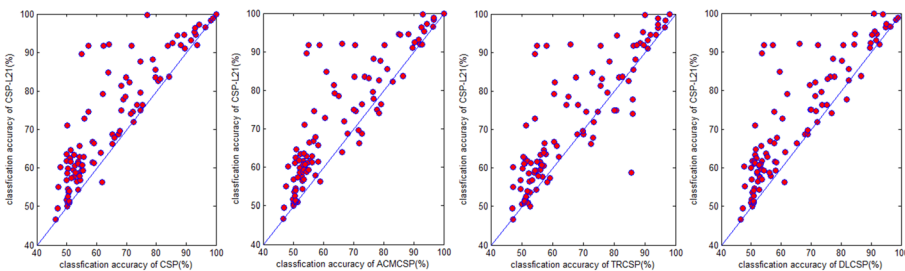
**Fig. 4** Classification accuracies of CSP, ACMCSP, TRCSP, DLCSP and CSP-L21 on the subjects of the three real EEG data sets with outliers added. The last group in each plot depicts the average classification accuracies across the different subjects of each EEG data set. **a** Data set IIIa of BCI competition III. **b** Data set IVa of BCI competition III. **c** Data set IIa of BCI competition IV



**Fig. 5** Classification accuracies of CSP-L21 for each subjects from the three real data sets vary with the value of line search parameter  $\beta$ . **a** Data set IIIa of BCI competition III. **b** Data set IVa of BCI competition III. **c** Data set IIa of BCI competition IV

list {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1} for beta and draw the curve of classification accuracies for each subject of the three real data sets varying with the value of  $\beta$  in Fig. 5. This fully proves that the state of the brain wave varies greatly between individuals, which leads to the difference in the selection of the optimal value of beta for each subject. In most cases, the accuracies do not change widely as the value of  $\beta$  changes.

Moreover, we compare these methods in general and evaluate the significant differences between them. Figure 6 shows the scatter plots [27] of the classification accuracies with and without added outliers. There are 102 cases in total. Each point denotes the classification accuracies of CSP, ACMCSP, TRCSP, DLCSP and CSP-L21 on one subject. In these four figures, the y-axis represents the recognition rates of CSP-L21. Most of the points are above the diagonal line, indicating that CSP-L21 gains the highest scores on most of the subjects and outperforms the other four methods.



**Fig. 6** Classification accuracies of CSP, ACMCSP, TRCSP and DLCSP compared with CSP-L21 on the real EEG data sets with and without outliers added. Each solid point denotes the classification accuracies of two methods on one subject. The proposed CSP-L21 method outperforms the other methods for the points above the diagonal line



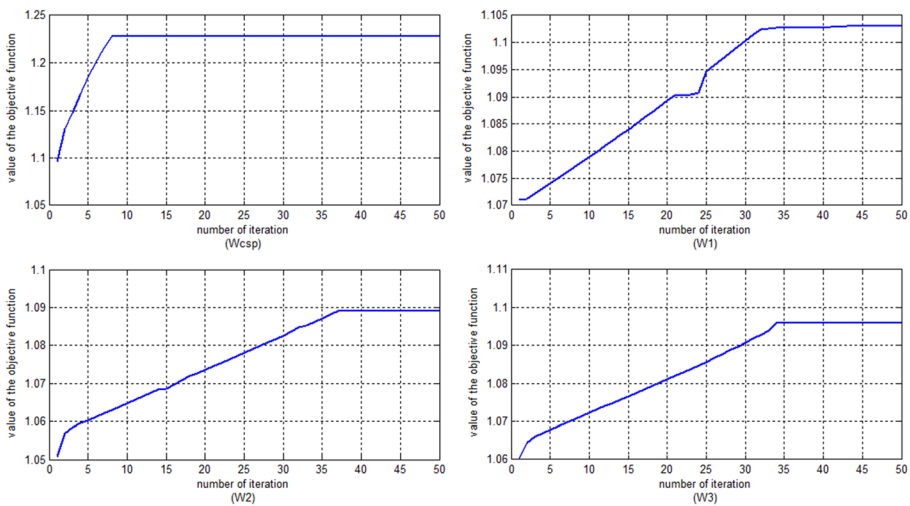
**Table 5** The  $p$ -values and  $h$ -values of the Wilcoxon signed-rank test

Method		$p$ -value( $\alpha=0.05$ )	$h$ -value
CSP-L21	CSP	0.0004	1
	ACMCSP	0.0006	1
	TRCSP	0.0045	1
	DLCSP	0.0004	1

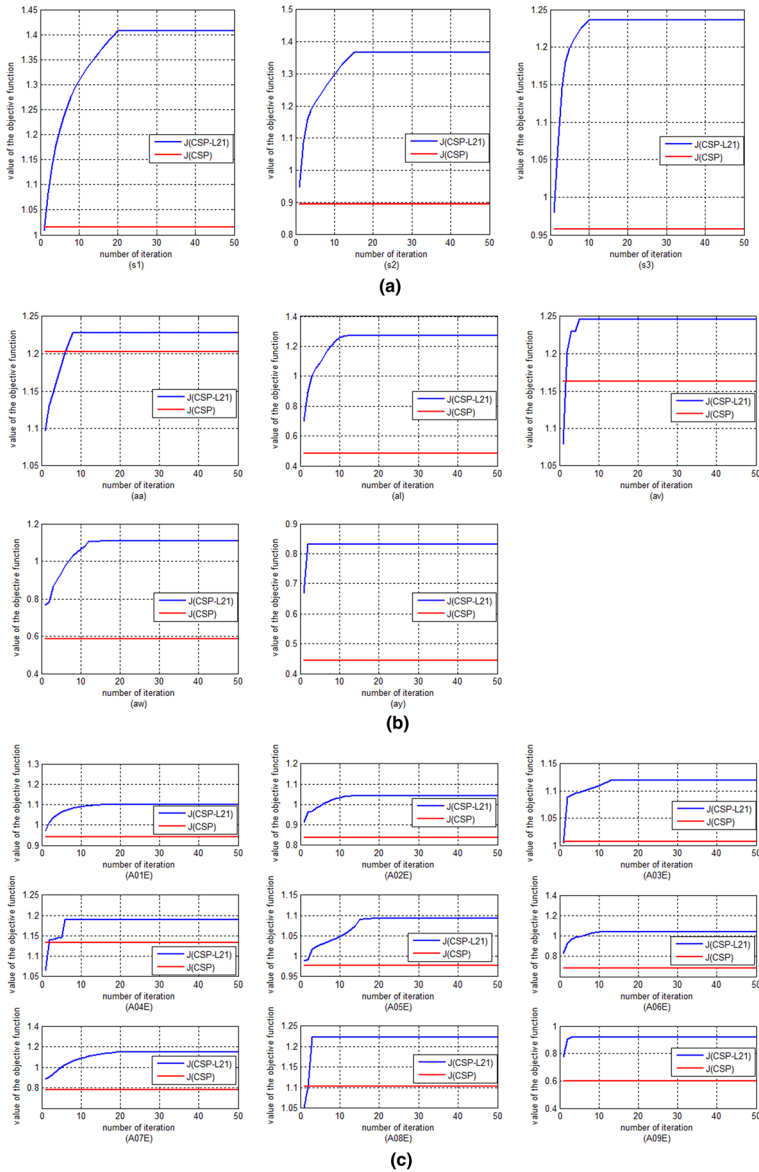
The following values denote statistical significance at the 5% significance level

To show the statistical significance of the results, the Wilcoxon signed-rank test, which is shown in Table 5, is used to compare the significance of the differences in results between CSP-L21 and CSP, ACMCSP, TRCSP and DLCSP at a significance level of 0.05. Clearly, the situation in which the  $p$ -value is less than 0.05 and the  $h$ -value equals 1 indicates a significant difference between the two methods.

Finally, in order to demonstrate the convergence of the proposed algorithm, we draw the convergence curves with the different random initial values of  $W$ . Take the subject aa as an example, Fig. 7 shows values of the objective function changing with the number of iterations when the projection matrix  $W$  takes different initial values. What calls for special attention is that one initial value of  $W$  is set as the basic spatial filter obtained by the traditional CSP algorithm. It is observed that the objective function is increasing and can always converge within limited steps. Particularly, we find that the curve converges quickly and stably with the initial  $W$  provided by the classical CSP. Therefore, we take this set in our experiments. Figure 8 shows the convergence curves of all subjects in the three real data sets, in which the red curves represent the objective function of CSP for reference. It is indicated that the proposed algorithm has good performance of convergence.



**Fig. 7** Values of the objective function of CSP-L21 vary with the number of iterations on the subject aa when the projection matrix  $W$  takes different initial values



**Fig. 8** Values of the objective function of CSP-L21 vary with the number of iterations on each subjects in the three real data sets. **a** Data set IIIa of BCI competition III. **b** Data set IVa of BCI competition III. **c** Data set IIa of BCI competition IV

Through the above experiments and analysis, it can be concluded that the proposed CSP-L21 method is powerful for robust modeling.

## 5 Conclusion

In this paper, we propose the L21-norm-based common spatial pattern, termed as CSP-L21. The new approach is obtained by rewriting the formulation of the conventional CSP using the L21-norm rather than the L2-norm. The advantages of CSP-L21 are that it alleviates the influence of outliers, has rotation invariance and good geometric structure. In addition, we design a non-greedy iterative algorithm for the optimal spatial filter matrix. The effectiveness and robustness of the proposed CSP-L21 method are confirmed by classification experiments on a toy example utilizing three real EEG data sets. However, the line search parameter  $\beta$  needs to be tuned theoretically and practically for the stability of the program.

**Acknowledgements** The authors would like to extend sincere gratitude to the reviewers for their thoughtful comments and suggestions. This work is supported by the National Natural Science Foundation of China under Grant 61773114.

## References

1. Blankertz B, Tomioka R, Lemm S et al (2008) Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Process Mag* 25(1):41–56
2. Chaudhary U, Birbaumer N, Ramos-Murguialday A (2016) Brain-computer interfaces for communication and rehabilitation. *Nat Rev Neurol* 12(9):513–525
3. Grosse-Wentrup M, Liefhold C, Gramann K et al (2009) Beamforming in noninvasive brain-computer interfaces. *IEEE Trans Biomed Eng* 56(4):1209–1219
4. Dornhege G, Blankertz B, Curio G et al (2004) Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms. *IEEE Trans Biomed Eng* 51(6):993–1002
5. Wang H, Tang Q, Zheng W (2012) L1-norm-based common spatial patterns. *IEEE Trans Biomed Eng* 59(3):653–662
6. Fang N, Wang H (2017) Generalization of local temporal correlation common spatial patterns using Lp-norm ( $0 < p < 2$ ). In: *International Conference on Neural Information Processing*, pp 769–777
7. Wang H, Zheng W (2008) Local temporal common spatial patterns for robust single-trial EEG classification. *IEEE Trans Neural Syst Rehabil Eng* 16(2):131–139
8. Ding C, Zhou D, He X et al (2006) R1-PCA: rotational invariant L1-norm principal component analysis for robust subspace factorization. In: *International Conference on Machine Learning*, pp 281–288
9. Yang Y, Shen H, Ma Z et al (2011) L2, 1-norm regularized discriminative feature selection for unsupervised learning. In: *International Joint Conference on Artificial Intelligence*, pp 1589–1594
10. Liao S, Gao Q, Yang Z et al (2018) Discriminant analysis via joint Euler transform and  $\ell_{2,1}$ -norm. *IEEE Trans Image Process* 27(11):5668–5682
11. Liu Y, Gao Q, Gao X et al (2018) L2,1-norm discriminant manifold learning. *IEEE Access* 6:40723–40734
12. Nie F, Huang H, Cai X et al (2010) Efficient and robust feature selection via joint  $\ell_2, 1$ -norms minimization. In: *Proceedings of the Conference on Neural Information Processing Systems*, pp 1813–1821
13. Luo M, Nie F, Chang X et al (2016) Avoiding optimal mean robust PCA/2DPCA with non-greedy L1-norm maximization. In: *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp 1802–1808
14. Blankertz B, Kawanabe M, Tomioka R et al (2008) Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing. In: *Proceedings of the Conference on Neural Information Processing Systems*, pp 113–120
15. Lotte F, Guan C (2011) Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. *IEEE Trans Biomed Eng* 58(2):355–362
16. Kawanabe M, Vidaurre C (2009) Improving BCI performance by modified common spatial patterns with robustly averaged covariance matrices. In: *World Congress on Medical Physics and Biomedical Engineering*, pp 279–282

17. Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. *J Multivar Anal* 88(2):365–411
18. Li X, Wang H (2013) Smooth spatial filter for common spatial patterns. In: *International Conference on Neural Information Processing*, pp 315–322
19. Li X, Lu X, Wang H (2016) Robust common spatial patterns with sparsity. *Biomed Signal Process Control* 26:52–57
20. Wang H, Li X (2016) Regularized filters for L1-norm-based common spatial patterns. *IEEE Trans Neural Syst Rehabil Eng* 24(2):201–211
21. Deng Y, Li Z, Wang H et al (2020) Local temporal joint recurrence common spatial patterns for MI-based BCI. In: *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference*, pp 813–816
22. Yong X, Ward R, Birch G (2008) Robust common spatial patterns for EEG signal preprocessing. In: *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp 2087–2090
23. Aljalal M, Djemal R, Ibrahim S (2019) Robot navigation using a brain computer interface based on motor imagery. *J Med Biol Eng* 39(4):508–522
24. Webb A (1999) *Statistical Pattern Recognition*. United Kingdom, London
25. Wolpaw J, Birbaumer N, McFarland D et al (2002) Brain-computer interfaces for communication and control. *Clin Neurophysiol* 113(6):767–791
26. Parra L, Spence C, Gerson A et al (2005) Recipes for linear analysis of EEG. *Neuroimage* 28(2):326–341
27. Pfurtscheller G, Silva F (1999) Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin Neurophysiol* 110(11):1842–1857
28. Zhao Y, Han J, Chen Y et al (2018) Improving generalization based on L1-norm regularization for EEG-based motor imagery classification. *Front Neurosci* 12:272
29. Tang Q, Wang J, Wang H (2014) L1-norm based discriminative spatial pattern for single-trial EEG classification. *Biomed Signal Process Control* 10(3):313–321
30. Huang L, Gu J, Li R et al (2013) A novel BCI classifier based on autoregressive model and support vector machine. *Adv Mater Res* 694–697:2522–2525
31. Aboy M, McNames J, Márquez O et al (2004) Power spectral density estimation and tracking nonstationary pressure signals based on Kalman filtering. *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 1:156–159
32. Adeli H, Zhou Z, Dadmehr N (2003) Analysis of EEG records in an epileptic patient using wavelet transform. *J Neurosci Methods* 123(1):69–87
33. Delorme A, Sejnowski T, Makeig S (2007) Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage* 34(4):1443–1449

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.