

Spatial Filtering for Brain Computer Interfaces: A Comparison between the Common Spatial Pattern and Its Variant

He He and Dongrui Wu, *Senior Member, IEEE*

Key Laboratory of the Ministry of Education for Image Processing and Intelligent Control,
School of Automation, Huazhong University of Science and Technology, Wuhan, China

E-mail: hehe91@hust.edu.cn, drwu@hust.edu.cn

Abstract—The electroencephalogram (EEG) is the most popular form of input for brain computer interfaces (BCIs). However, it can be easily contaminated by various artifacts and noise, e.g., eye blink, muscle activities, powerline noise, etc. Therefore, the EEG signals are often filtered both spatially and temporally to increase the signal-to-noise ratio before they are fed into a machine learning algorithm for recognition. This paper considers spatial filtering, particularly, the common spatial pattern (CSP) filters for EEG classification. In binary classification, CSP seeks a set of filters to maximize the variance for one class while minimizing it for the other. We first introduce the traditional solution, and then a new solution based on a slightly different objective function. We performed comprehensive experiments on motor imagery to compare the two approaches, and found that generally the traditional CSP solution still gives better results. We also showed that adding regularization to the covariance matrices can improve the final classification performance, no matter which objective function is used.

Index Terms—Brain computer interface, common spatial pattern, motor imagery, Stiefel manifold, regularization

I. INTRODUCTION

Brain computer interface (BCI) [16] provides a direct communication pathway between a user and an external device, such that the device can recognize the user's brain status and respond accordingly. BCIs have found successful applications in robotics, text input, games, healthcare, etc [13], [15].

The electroencephalogram (EEG) is the most popular form of BCI input as it is easy to acquire (no surgery) and offers high temporal resolution. However, there are still many challenges for wide-spread real-world applications of EEG-based BCIs [9], [11]. An important one is related to the EEG signal quality, as EEG signals can be easily contaminated by various artifacts and noise such as muscle movements, eye blinks, heartbeats, environmental electromagnetic fields, etc. Therefore, it is important to filter the EEG signals both spatially and temporally to increase the signal-to-noise ratio.

This paper focuses on the common spatial pattern (CSP) filters [2], [7], [8], [12], [14], [17], [18] for spatial filtering. CSP was first proposed by Koles et al. [8] to extract discriminative EEG features from two human populations (normal people and patients). Mueller-Gerking et al. [12] extended it to single-trial

classification of motor imagery EEG, which remains the most popular and important application field of CSP filters.

CSP finds a set of spatial filters that can achieve good discrimination among different classes. At the same time, it can reduce the dimensionality of the EEG signals. Different approaches have been proposed in the literature to compute them [2], [8], [12], [17]. In this paper we review some typical approaches, and propose a new approach by solving a slightly different objective function in the optimization. We then use two motor imagery datasets in BCI Competition IV to compare these approaches.

The rest of this paper is organized as follows: Section II introduces existing approaches for computing the CSP filters. Section III introduces the proposed Stiefel manifold based approach. Section IV compares the performance of these approaches in motor imagery applications. Finally, Section V draws conclusions.

II. COMMON SPATIAL PATTERN (CSP)

Without loss of generality, this paper considers binary classification only. However, multi-class classification can be extended from binary classification by the one-versus-one approach or the one-versus-the-rest approach [17].

Let $X \in \mathbb{R}^{C \times T}$ be an EEG epoch, where C is the number of channels and T the number of time samples. Assume Class c has N_c epochs, and $X_{c,i}$ is the i th EEG epoch in Class c . Then, the class mean covariance matrices are:

$$\Sigma_0 = \frac{1}{N_0} \sum_{i=1}^{N_0} X_{0,i} X_{0,i}^T \quad (1)$$

$$\Sigma_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} X_{1,i} X_{1,i}^T \quad (2)$$

CSP identifies a spatial filtering matrix $W \in \mathbb{R}^{C' \times C}$ ($C' < C$) from Σ_0 and Σ_1 , which projects the EEG signals from the original C -dimensional space to a lower C' -dimensional space:

$$X^* = W^T X \in \mathbb{R}^{C' \times T} \quad (3)$$

Each column of W is a spatial filter, and it is designed such that for the projected signal, the variance for one class is maximized, meanwhile the variance for the other class is minimized.

Different motivations and approaches for computing W have been proposed in the literature. Some typical ones are introduced next.

A. Approach 1

The earliest CSP approach for EEG processing was proposed by Koles et al. [8]. It first forms a composite covariance matrix:

$$\Sigma = \Sigma_0 + \Sigma_1 \quad (4)$$

which has the following singular value decomposition (SVD):

$$\Sigma = U\Lambda U^T \quad (5)$$

where U is an orthonormal matrix whose columns are normalized eigenvectors of Σ , and Λ is a diagonal matrix, whose diagonal terms are the corresponding eigenvalues.

It then constructs a whitening matrix

$$P = \Lambda^{-\frac{1}{2}} U^T \quad (6)$$

to equalize the variances in the space spanned by the eigenvectors. Applying the whitening transformation to both Σ_0 and Σ_1 , we have

$$S_0 = P\Sigma_0 P^T \quad (7)$$

$$S_1 = P\Sigma_1 P^T \quad (8)$$

$$S_0 + S_1 = I \quad (9)$$

where I is the identify matrix.

It can be shown that S_0 and S_1 share the same eigenvectors, and the sum of the corresponding eigenvalues always equals 1 [6], i.e., if we perform eigen decomposition on S_0 by

$$S_0 = U\Lambda_0 U^T \quad (10)$$

Then we should also have

$$S_1 = U\Lambda_1 U^T \quad (11)$$

$$\Lambda_0 + \Lambda_1 = I \quad (12)$$

where U is an orthogonal matrix whose columns are the normalized eigenvectors, i.e., $UU^T = I$.

Assume the diagonal terms of Λ_0 has been sorted in descending order. Then, the diagonal terms of Λ_1 must be in the ascending order. The first column of U accounts for the maximum variance in S_0 and the least variance in S_1 , so it is very useful in discriminating between S_0 and S_1 . Similarly, the last column of U accounts for the least variance in S_0 and the maximum variance in S_1 , so it is also very useful in discriminating between S_0 and S_1 . In practice we usually select a few eigenvectors corresponding to the maximum eigenvalues, and also a few eigenvectors corresponding to the minimum eigenvalues, to form W .

In summary, Koles et al. [8] used the following procedure to compute the CSP matrix $W \in \mathbb{R}^{C' \times C}$:

- 1) Compute Σ in (4) and its SVD in (5).
- 2) Compute the whitening matrix P in (6).
- 3) Compute S_0 in (7) and its SVD in (10).
- 4) Sort the diagonal terms of Λ_0 in descending order, and adjust the columns of U accordingly. Assemble V as the first and last $C'/2$ columns of U .
- 5) Compute $W = P^T V$.

B. Approach 2

In Approach 1, we need to first calculate the whitening matrix P , then the orthogonal transformation matrix U , assemble V from U , and finally obtain $W = P^T V$. Blankertz et al. [2] proposed a simpler solution to compute W directly from Σ_1 and Σ_2 .

According to [6], we can simultaneously diagonalize the two mean covariance matrices:

$$V^T \Sigma_0 V = \Lambda_0 \quad (13)$$

$$V^T \Sigma_1 V = \Lambda_1 \quad (14)$$

where $\Lambda_0 + \Lambda_1 = I$.

Assume the diagonal terms of Λ_0 has been sorted in descending order. Then, the diagonal terms of Λ_1 must be in the ascending order. The first a few columns of V account for the maximum variance in Σ_0 and the least variance in Σ_1 , so they are very useful in discriminating between Σ_0 and Σ_1 . Similarly, the last a few columns of V account for the least variance in Σ_0 and the maximum variance in Σ_1 , so they are also very useful in discriminating between Σ_0 and Σ_1 .

In summary, Blankertz et al. [2] used the following procedure to compute CSP matrix $W \in \mathbb{R}^{C' \times C}$:

- 1) Solve the following generalized eigenvalue problem

$$\Sigma_0 \mathbf{w} = \lambda \Sigma_1 \mathbf{w} \quad (15)$$

to obtain λ_i and the corresponding \mathbf{w}_i , $i = 1, \dots, C$.

- 2) Sort λ_i and the corresponding \mathbf{w}_i in descending order.
- 3) Construct $W = [\mathbf{w}_1, \dots, \mathbf{w}_{\frac{C'}{2}}, \mathbf{w}_{C-\frac{C'}{2}+1}, \mathbf{w}_C]$, i.e., W uses the first and last $\frac{C'}{2}$ \mathbf{w}_i as its columns.

C. Discussions

Although Approaches 1 and 2 use different procedures to compute W , they actually give the same results.

For Approach 1, from (7) and (10) we have:

$$S_0 = P\Sigma_0 P^T = U\Lambda_0 U^T \quad (16)$$

which can be rewritten as:

$$\Sigma_0 P^T U = P^{-1} U \Lambda_0 \quad (17)$$

Since P is the whitening matrix for the composite covariance matrix Σ , i.e.,

$$P\Sigma P^T = I \quad (18)$$

we have

$$P^{-1} = \Sigma P^T \quad (19)$$

Substituting (19) into (17), it follows that:

$$\Sigma_0 P^T U = \Sigma P^T U \Lambda_1 \quad (20)$$

i.e.,

$$\Sigma^{-1} \Sigma_0 (P^T U) = (P^T U) \Lambda_1 \quad (21)$$

So the filtering matrix $P^T U$ consists of the eigenvectors of $\Sigma^{-1} \Sigma_0$, which are identical to the eigenvectors of $\Sigma_1^{-1} \Sigma_0$. $W = P^T V$ consists a subset of $P^T U$.

For Approach 2, (15) can be rewritten as:

$$\Sigma_1^{-1} \Sigma_0 \mathbf{w} = \lambda \mathbf{w} \quad (22)$$

So each \mathbf{w} is also an eigenvector of $\Sigma_1^{-1} \Sigma_0$.

In summary, the above derivations show that W can be constructed from the eigenvectors of $\Sigma_1^{-1} \Sigma_0$, and this is actually the most frequently used approach in computing the CSP in practice.

Finally, if we take a closer look at the rationale for CSP, we can conclude that the CSP filters actually optimize the following objective function:

$$\begin{aligned} Ratio1 = \arg \max_{W=[\mathbf{w}_1, \dots, \mathbf{w}_{C'}]} & \sum_{i=1}^{C'/2} \frac{\mathbf{w}_i^T \Sigma_0 \mathbf{w}_i}{\mathbf{w}_i^T \Sigma_1 \mathbf{w}_i} \\ & + \sum_{i=C'/2+1}^{C'} \frac{\mathbf{w}_i^T \Sigma_1 \mathbf{w}_i}{\mathbf{w}_i^T \Sigma_0 \mathbf{w}_i} \end{aligned} \quad (23)$$

III. A NEW APPROACH FOR COMPUTING THE SPATIAL FILTERS IN THE STIEFEL MANIFOLD

$Ratio1$ in (23) optimizes the sum of ratios. A closely related objective function is to optimize the ratio of sums:

$$\begin{aligned} Ratio2 = \arg \max_{W=[\mathbf{w}_1, \dots, \mathbf{w}_{C'}]} & \frac{\sum_{i=1}^{C'/2} \mathbf{w}_i^T \Sigma_0 \mathbf{w}_i}{\sum_{i=1}^{C'/2} \mathbf{w}_i^T \Sigma_1 \mathbf{w}_i} \\ & + \frac{\sum_{i=C'/2+1}^{C'} \mathbf{w}_i^T \Sigma_1 \mathbf{w}_i}{\sum_{i=C'/2+1}^{C'} \mathbf{w}_i^T \Sigma_0 \mathbf{w}_i} \end{aligned} \quad (24)$$

We are interested in (24) because:

- 1) (24) is very similar to (23). In fact, they are identical when $C' = 2$. So, we would like to investigate whether this new objective function could result in better classification performance.
- 2) Although both Σ_0 and Σ_1 are symmetric, generally $\Sigma_1^{-1} \Sigma_0$ is not symmetric, so its eigenvectors are not mutually orthogonal. In other words, the columns of W obtained from optimizing $Ratio1$ are correlated, which may encode redundant information. On the other hand, when solving $Ratio2$, it is possible to make the first $C'/2$ columns of W orthogonal, and also the last $C'/2$ columns orthogonal (although the first and last $C'/2$ columns are generally not mutually orthogonal). It's interesting to investigate whether this orthogonality can improve the classification performance.

The two terms in (24) are independent, so we can compute them separately. Unfortunately, they do not have a closed-form solution. Cunningham and Ghahramani [4] proposed an

approach for solving this problem in the Stiefel manifold (SM), which is the set of ordered tuples of orthonormal vectors. Their algorithm is complex and iterative, and we refer the readers to [4] for it. The authors have also provided their Matlab code¹, which was used in our experiment.

IV. EXPERIMENTS

In this section we compare the SM approach with the traditional CSP approach on two different motor imagery datasets, using two classifiers.

A. Datasets

Both datasets were from BCI Competition IV².

The first is Dataset 1 [3], which was recorded from seven healthy subjects. For each subject two classes of motor imagery were selected from the three classes: left hand, right hand, and foot. Continuous EEG signals were acquired from 59 channels and were divided into three parts: calibration data, evaluation data, and special feature. We only used calibration data in this paper, and each subject had 100 trials in each class.

The second is Dataset 2a, which consists of EEG data from nine subjects. Every subject was instructed to perform four different motor imagery tasks, namely the imagination of movement of the left hand, right hand, both feet, and tongue. The signals were recorded using 22 EEG channels and 3 EOG channels. We only used two classes (left hand and right hand), and each class has 72 trials.

B. Preprocessing and Classifiers

The EEG signals were preprocessed using the Matlab EEGLAB toolbox [5], following the guideline in [2]. First, a band-pass filter (7-30 Hz) was applied to remove muscle artifacts, line-noise contamination and DC drift. Then, we extracted EEG signals between [1, 3.5] seconds after the cue appearance as our trials.

For each subject, we randomly selected 50% trials for training, and the remaining 50% for testing, and repeated this process 30 times to get statistically meaningful results. For a given partition, we computed the spatial filters by the traditional CSP approach, and also the proposed SM approach. We then tested two different classifiers:

- 1) Linear discriminant analysis (LDA), as in [2]. The features for the i th trial were:

$$f_i^j = \log(\mathbf{w}_j^T X_i X_i^T \mathbf{w}_j), \quad j = 1, \dots, C' \quad (25)$$

- 2) Minimum distance to Riemannian mean (MDRM), as in [1]. The features were the covariance matrices of the trials.

C. Experimental Results

The classification accuracies for different subjects, averaged across 30 runs, are shown as the first four bars in Figs. 1 and 2, for different number of spatial filters. The horizontal axis shows the indices of the subjects, and also the average across the subjects. Observe that for both LDA and MDRM, generally

¹<http://github.com/cunni/ldr>

²<http://www.bci.de/competition/iv/>.

the performances of SM were slightly worse than CSP. Also, generally the performance of MDRM was slightly worse than LDA.

Ratio1 and *Ratio2* from the two objective functions are shown in Figs. 3 and 4. Observe that CSP always had higher *Ratio1* than SM, and SM always had higher *Ratio2* than CSP, which are as expected. However, it seems that *Ratio1* is a better objective function, since a higher *Ratio1* usually results in a better classification accuracy.

The average correlation coefficients between the columns of W for different subjects are shown in Tables I and II, for Datasets 1 and 2, respectively. Observe that on average the columns of W computed from the SM approach were less correlated; however, this did not necessarily result in better classification performance.

TABLE I

AVERAGE CORRELATION COEFFICIENTS BETWEEN THE COLUMNS OF W FOR DATASET 1.

Subject	1	2	3	4	5	6	7	Avg
CSP	.1393	.0760	.1064	.0797	.0839	.1122	.1148	.1018
SM	.0907	.0477	.0399	.1373	.0738	.0943	.0886	.0817

TABLE II

AVERAGE CORRELATION COEFFICIENTS BETWEEN THE COLUMNS OF W FOR DATASET 2.

Subject	1	2	3	4	5	6	7	8	9	Avg
CSP	.1807	.3224	.2571	.1776	.1416	.2116	.2419	.1904	.1665	.2100
SM	.1744	.1759	.1522	.1488	.1280	.1481	.1354	.1288	.1521	.1493

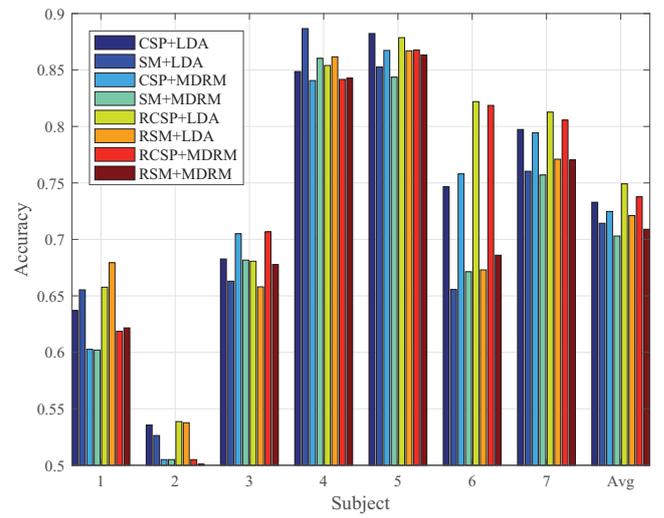
D. Discussions

The above results showed that generally a larger *Ratio1* results in better classification performance. We would like to study if this is always true. Lotte and Guan [10] showed that regularization on the traditional CSPs can improve the classification performance. In this subsection we study if regularization can also improve *Ratio1*, i.e., the improved performance is due to the increased *Ratio1*.

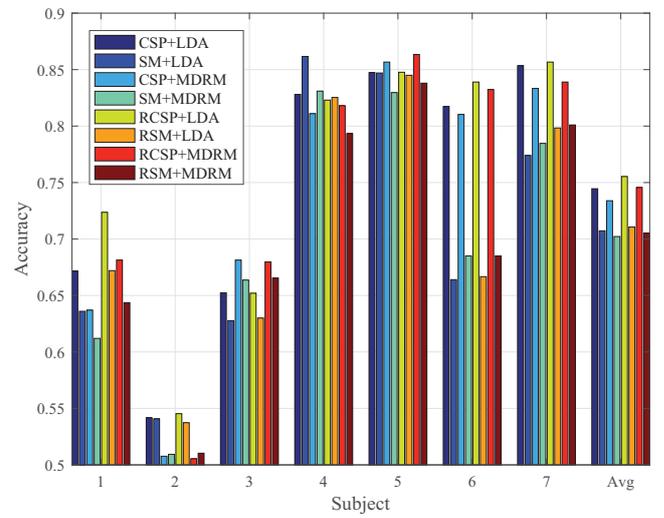
Several different regularized CSP (RCSP) approaches have been proposed in [10]. In this paper we compute the first $C'/2$ columns of W in the RCSP from the eigenvectors of $(\Sigma_1 + \lambda I)^{-1}\Sigma_0$, and the last $C'/2$ columns from the eigenvectors of $(\Sigma_0 + \lambda I)^{-1}\Sigma_1$, where λ is an adjustable parameter identified by cross-validation on the labeled data [10]. This approach has showed good performance in [10].

Similarly, we also develop a regularized SM (RSM) approach, where the first $C'/2$ columns of W in the RSM are computed from maximizing $\frac{\sum_{i=1}^{C'/2} \mathbf{w}_i^T \Sigma_0 \mathbf{w}_i}{\sum_{i=1}^{C'/2} \mathbf{w}_i^T (\Sigma_1 + \lambda I) \mathbf{w}_i}$, and the last $C'/2$ columns are computed from maximizing $\frac{\sum_{i=C'/2+1}^{C'} \mathbf{w}_i^T \Sigma_1 \mathbf{w}_i}{\sum_{i=C'/2+1}^{C'} \mathbf{w}_i^T (\Sigma_0 + \lambda I) \mathbf{w}_i}$, in which λ is again an adjustable parameter identified by cross-validation on the labeled data.

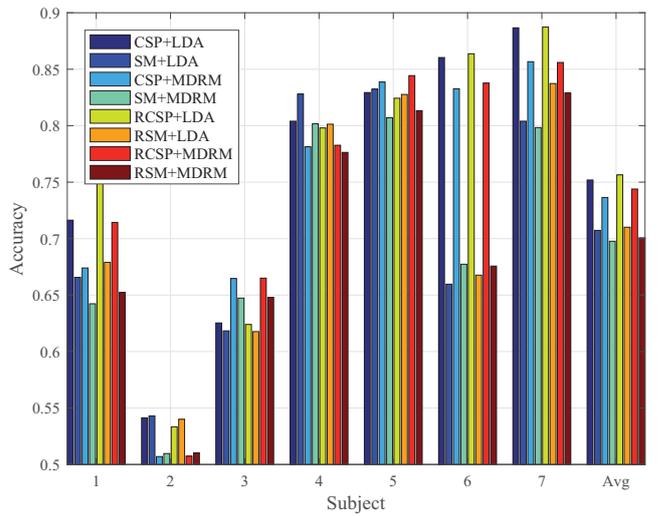
The classification accuracies for different subjects, averaged across 30 runs, are shown as the last four bars in Figs. 1 and 2, for different number of spatial filters. Observe that:



(a)

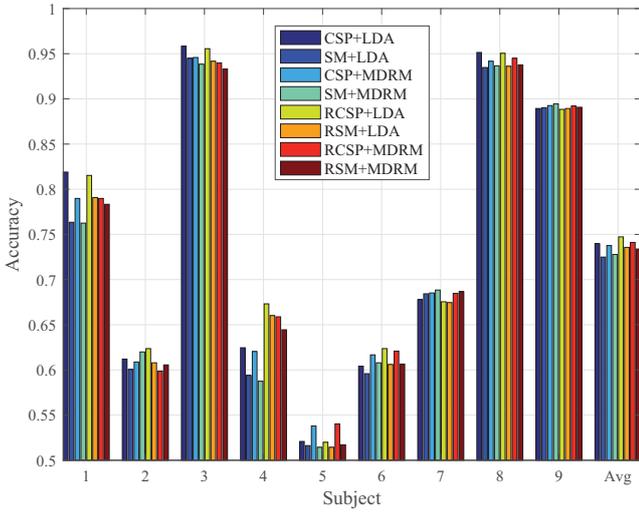


(b)

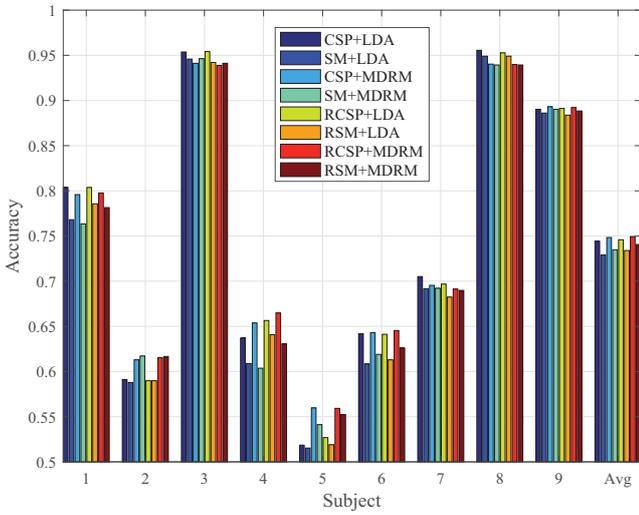


(c)

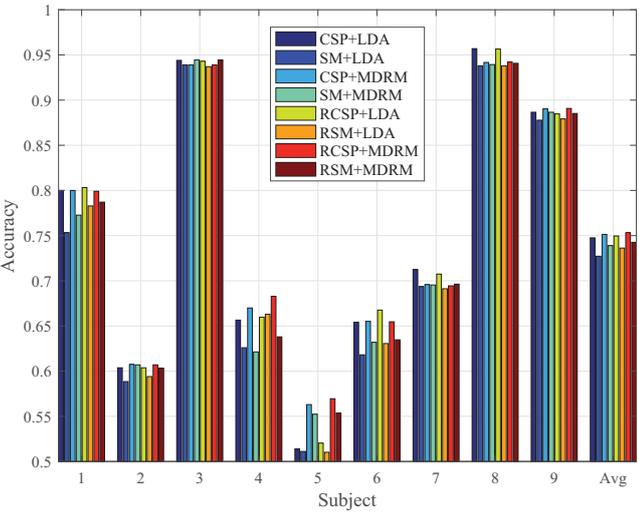
Fig. 1. Classification accuracies on Dataset 1. (a) $C' = 4$; (b) $C' = 6$; (c) $C' = 8$.



(a)

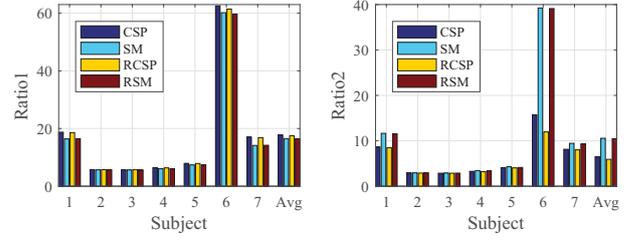


(b)

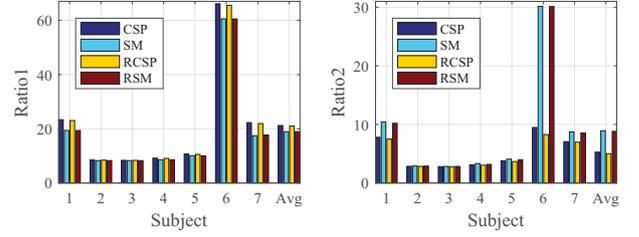


(c)

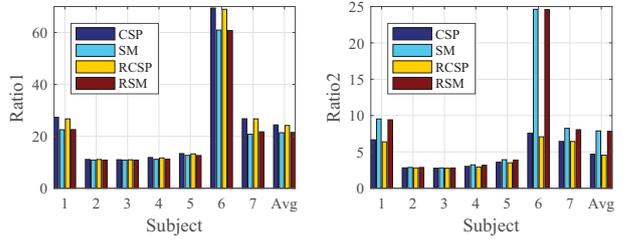
Fig. 2. Classification accuracies on Dataset 2. (a) $C' = 4$; (b) $C' = 6$; (c) $C' = 8$.



(a)



(b)



(c)

Fig. 3. *Ratio1* and *Ratio2* on Dataset 1. (a) $C' = 4$; (b) $C' = 6$; (c) $C' = 8$.

- 1) Generally the performance of RCSP was better than CSP, and the performance of RSM was better than SM, for both LDA and MDRM. This confirmed the observations in [10].
- 2) For both LDA and MDRM, generally the performances of RSM were slightly worse than RCSP. This pattern was also observed between SM and CSP.
- 3) For both RCSP and RSM, generally the performance of MDRM was slightly worse than LDA. Again, this pattern was observed before for CSP and SM.

Next we check if the improved performance of RCSP and RSM over CSP and SM was indeed due to increased *Ratio1*. For this purpose, we plot *Ratio1* and *Ratio2* from RCSP and RSM as the last four bars in Figs. 3 and 4. Observe that:

- 1) Although RCSP had higher classification accuracy than CSP, its *Ratio1* and *Ratio2* were slightly worse than those from the unregularized CSP. A similar pattern can also be observed from RSM and SM. This is reasonable, as CSP directly optimizes *Ratio1*, whereas the objective function for RCSP is slightly different.
- 2) RCSP always had higher *Ratio1* than RSM, and RSM always had higher *Ratio2* than RCSP. This pattern was similar to what we have observed on CSP and SM.

From all these observations, we can reach the following two conclusions:

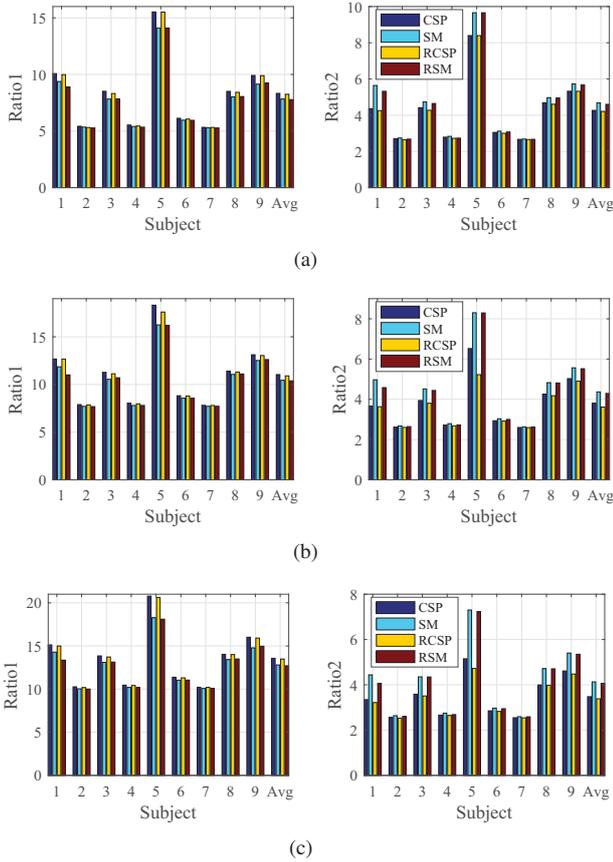


Fig. 4. *Ratio1* and *Ratio2* on Dataset 2. (a) $C' = 4$; (b) $C' = 6$; (c) $C' = 8$.

- 1) Although the SM approach optimizes an objective function very similar to the objective function of the traditional CSP, it generally results in slightly worse classification performance. Meanwhile, the SM approach does not have a closed-form solution, and hence requires much higher computational cost. For these two reasons, the new objective function and optimization method are not recommended in designing spatial filters.
- 2) *Ratio1* in general is a reliable performance measure for the spatial filters. However, because Σ_0 and Σ_1 may be noisy, adding regularization to the covariance matrices can improve the final classification performance, although it may reduce *Ratio1* very slightly. This suggests that maybe it is possible to define an improved objective function for the spatial filters, which is better correlated with the final classification performance. This is one of our future research directions.

V. CONCLUSIONS

CSP is a popular spatial filtering approach for EEG-based BCIs, especially for motor imagery applications. It is used to increase the signal-to-noise ratio of EEG signals before they are fed into a classifier. Its main idea is to project the EEG signals from the original sensor space into a low-dimensional space which maximizing the variance for one

class while minimizing it for the other. Different motivations and approaches have been proposed to compute the CSP filters. In this paper, we gave an overview of some typical approaches, and showed that they lead to the same closed-form solution. We also proposed a new objective function, which closely resembles the objective function of the traditional CSP, for developing spatial filters. Experimental results on two Motor Imagery datasets showed that the new objective function results in slightly worse spatial filters than the CSP filters, so the traditional CSP approach is still preferred in practice. Moreover, we also confirmed that adding regularization on the covariance matrices can improve the classification performance, no matter which objective function is used.

REFERENCES

- [1] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain-computer interface classification by Riemannian geometry," *IEEE Trans. on Biomedical Engineering*, vol. 59, no. 4, pp. 920–928, 2012.
- [2] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2008.
- [3] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Muller, 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.
- [4] J. P. Cunningham and Z. Ghahramani, "Linear dimensionality reduction: survey, insights, and generalizations," *Journal of Machine Learning Research*, vol. 16, no. 1, pp. 2859–2900, 2015.
- [5] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, pp. 9–21, 2004.
- [6] K. Fukunaga, *Introduction to statistical pattern recognition*. Academic press, 1972.
- [7] H. He and D. Wu, "Transfer learning enhanced common spatial pattern filtering for brain computer interfaces (BCIs): Overview and a new approach," in *Proc. 24th Int'l. Conf. on Neural Information Processing*, Guangzhou, China, November 2017.
- [8] Z. J. Koles, M. S. Lazar, and S. Z. Zhou, "Spatial patterns underlying population differences in the background EEG," *Brain Topography*, vol. 2, no. 4, pp. 275–284, 1990.
- [9] B. J. Lance, S. E. Kerick, A. J. Ries, K. S. Oie, and K. McDowell, "Brain-computer interface technologies in the coming decades," *Proc. of the IEEE*, vol. 100, no. 3, pp. 1585–1599, 2012.
- [10] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2011.
- [11] S. Makeig, C. Kothe, T. Mullen, N. Bigdely-Shamlo, Z. Zhang, and K. Kreutz-Delgado, "Evolving signal processing for brain-computer interfaces," *Proc. of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1567–1584, 2012.
- [12] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clinical Neurophysiology*, vol. 110, no. 5, pp. 787–798, 1999.
- [13] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.
- [14] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.
- [15] J. van Erp, F. Lotte, and M. Tangermann, "Brain-computer interfaces: Beyond medical applications," *Computer*, vol. 45, no. 4, pp. 26–34, 2012.
- [16] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.
- [17] D. Wu, J.-T. King, C.-H. Chuang, C.-T. Lin, and T.-P. Jung, "Spatial filtering for EEG-based regression problems in brain-computer interface (BCI)," *IEEE Trans. on Fuzzy Systems*, 2017, in press.
- [18] D. Wu, V. J. Lawhern, B. J. Lance, S. Gordon, T.-P. Jung, and C.-T. Lin, "EEG-based user reaction time estimation using Riemannian geometry features," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, 2017, in press.