Canonical Correlation Approach to Common Spatial Patterns

Eunho Noh[1] and Virginia R. de Sa[2]

Abstract— **Common spatial patterns (CSPs) are a way of spatially filtering EEG signals to increase the discriminability between the filtered variance/power between the two classes. The proposed canonical correlation approach to CSP (CCACSP) utilizes temporal information in the time series, in addition to exploiting the covariance structure of the different classes, to find filters which maximize the bandpower difference between the classes. We show with simulated data, that the unsupervised canonical correlation analysis (CCA) algorithm is better able to extract the original class-discriminative sources than the CSP algorithm in the presence of large amounts of additive Gaussian noise (while the CSP algorithm is better at very low noise levels) and that our CCACSP algorithm is a hybrid, yielding good performance at all noise levels. Finally, experiments on data from the BCI competitions confirm the effectiveness of the CCACSP algorithm and a merged CSP/CCACSP algorithm (mCCACSP).**

I. INTRODUCTION

Brain computer interfaces (BCIs) are devices that allow interaction between humans and computers using the brain signals of the user. One commonly used method to extract meaningful information in EEG (electroencephalography)-based BCI is to detect event-related desynchronization resulting from motor imagery (MI) of different limbs of the body. One of the most commonly used and effective spatial filtering methods used for feature extraction in MI tasks is common spatial patterns (CSPs) [2]. CSP gives filters which maximize the variance/power for one class while minimizing it for the other which increases the discriminability of the two classes when using bandpower features for classification. However, CSP does not take into account the dependence between time samples. CSP is also sensitive to noise and artifacts which are common in EEG signals [10].

Many variants of CSP have been developed to improve the performance of CSP. Various regularization methods for CSPs (RCSP) have been proposed as reviewed in [9]. Common spatio-spectral patterns (CSSP)[7] uses the temporal structure information to improve CSP. Spectrally weighted common spatial patterns (Spec-CSP)[13] learns the spectral weights as well as the spatial weights in an iterative way. Invariant CSP (iCSP)[3] minimizes variations in the EEG signal caused by various artifacts using a pre-calculated covariance matrix characterizing these modulations. Stationary CSP (sCSP) [11] regularizes CSP filter into stationary subspaces. Local temporal common spatial patterns (LTCSP)[14], [15] uses temporally local variances to compute the spatial filters.

[1]Eunho Noh is with the Department of Electrical & Computer Engineering, UCSD, La Jolla, CA eunoh@ucsd.edu

[2]Virginia R. de Sa is with Faculty of Cognitive Science, UCSD, La Jolla, CA desa@ucsd.edu

In this paper, we propose a method called canonical correlation approach to common spatial patterns (CCACSP) which incorporates the temporal structure of the data to extract discriminative and uncorrelated sources. The method was applied to a simulated dataset as well as two BCI competition datasets to compare our approach to the standard CSP algorithm.

II. METHODS

A. CSP algorithm

CSP finds spatial filters that maximize the variance/power of spatially filtered signals under one condition while minimizing it for the other condition. Let a column vector $x_t \in \mathbb{R}^C$ be the bandpassed EEG signal for time t where C is the number of EEG channels on the scalp and $X = (x_1, ..., x_L) \in \mathbb{R}^{C \times L}$ be a length L sequence of these EEG signals. The estimate of the normalized covariance matrix $\Sigma_y \in \mathbb{R}^{C \times C}$ can be calculated as follows:

$$\Sigma_y = \frac{1}{|\Upsilon_y|} \sum_{i \in \Upsilon_y} \frac{X_i X_i^\top}{tr(X_i X_i^\top)} \quad (1)$$

where $y \in \{1, 2\}$. The optimal set of CSP filters can be found by optimizing the following Rayleigh quotient

$$R(w) = \frac{w^\top \Sigma_1 w}{w^\top (\Sigma_1 + \Sigma_2) w}. \quad (2)$$

The solution can be found by solving the generalized eigenvalue problem given in the form $\Sigma_1 w = \lambda(\Sigma_1 + \Sigma_2)w$ [2]. The generalized eigenvector w_1^* corresponding to the largest eigenvalue maximizes the variance for class 1 while minimizing for class 2. The generalized eigenvector w_C^* corresponding to the smallest eigenvalue maximizes the variance for class 2 while minimizing for class 1.

B. CCA algorithm for separating uncorrelated sources

The goal of the CCA algorithm is to find a pair of projection vectors $v, w \in \mathbb{R}^C$ that maximize the correlation between two signal spaces (in general, the two signal spaces may have different dimensionality) [1]. Let Σ_{12} and Σ_{21} be the cross-covariance matrices of data matrices X_1 and X_2, then the pair of optimal vectors v and w can be found by solving the following eigenvalue problems

$$\Sigma_1^{-1} \Sigma_{12} \Sigma_2^{-1} \Sigma_{21} v = \lambda^2 v \quad (3)$$

$$\Sigma_2^{-1} \Sigma_{21} \Sigma_1^{-1} \Sigma_{12} w = \rho^2 w. \quad (4)$$

In a source separation setting, the two signals chosen for CCA are the original signal X and a time shifted version of the original signal X_Δ. This would be equivalent to maximizing the autocorrelation of each of the projections, hence recovering the original uncorrelated sources [5], [8]. For simplicity, we assume that the two data matrices have the same covariance Σ and that their cross-covariance (shift covariance matrix) is Σ_S. In the source separation setting, v and w are the same. We can find w^* with the highest autocorrelation by maximizing the following Rayleigh quotient

$$R(w) = \frac{w^\top \Sigma_S w}{w^\top \Sigma w}. \tag{5}$$

This problem can be solved with the generalized eigenvalue problem given in the form $\Sigma_S w = \lambda \Sigma w$ [8]. The generalized eigenvector w^* corresponding to the largest eigenvalue gives the projection with the highest autocorrelation. Note that the amount of shift given to X_Δ may prevent sources with the same autocorrelation for the given shift to be separated [5], [8]. For our analysis, the shift was fixed to one sample but multiple shifts or different length shifts can be used. CCA for blind source separation has been used to remove muscle artifacts from EEG data [6].

C. CCACSP algorithm

The CCA algorithm is good at separating independent sources, even in the presence of noise, but has the disadvantage that the separated sources are not weighted towards class discriminability for classification. In order to maintain the source separation properties of CCA while prioritizing separation for class discriminative sources, we search for a projection $w \in \mathbb{R}^C$ which maximizes the shifted variance of signals for one condition and at the same time minimizes the overall variance of the signals. Since this is a supervised method, the shift covariance matrices are calculated for each class. We denote the shift covariance matrix for class y as Σ_{Sy} where $y \in \{1, 2\}$. We define two Rayleigh quotients

$$R_1(w) = \frac{w^\top \Sigma_{S1} w}{w^\top (\Sigma_1 + \Sigma_2) w} \tag{6}$$

$$R_2(w) = \frac{w^\top \Sigma_{S2} w}{w^\top (\Sigma_2 + \Sigma_1) w}. \tag{7}$$

As done for CSP, w_{S1}^* (w_{S2}^*) which maximizes the shifted variance for class 1 (class 2) while minimizing for the overall variance can be found by maximizing the Rayleigh quotient $R_1(w)$ ($R_2(w)$) using the generalized eigenvalue problem $\Sigma_{S1} w = \lambda(\Sigma_1 + \Sigma_2) w$ ($\Sigma_{S2} w = \lambda(\Sigma_2 + \Sigma_1) w$).

III. ANALYSIS ON SIMULATED DATASET

A. Simulated dataset

Artificial datasets were generated by mixing signals from four sinusoidal sources of varying frequency (12, 10, 11.5, and 13 Hz). The artificial data were designed to simulate power differences between two "classes". Sinusoid 1 was set to have a higher magnitude for class 1 than for class 2 (1.3 vs 1). Sinusoids 2 and 3 were set to have a higher magnitude for class 1 (magnitudes of 1.3 vs 1 and 1 vs .7 respectively). Sinusoid 4 was set to have magnitude 4 for both classes. Gaussian noise was added to each of the sinusoidal signals. Finally, the noisy versions of the four sinusoids were mixed according to a linear mixing matrix to generate four electrode signals. Training data were generated using a sampling rate of 200 Hz with 5 seconds of data per class. Test data were generated in a similar manner but without the added noise. Multiple datasets were generated with different noise magnitudes ranging from 0.1 to 10.

B. Results

The three methods (CCA, CSP, and CCACSP) were trained and tested 10,000 times (at each of the additive noise magnitudes). In each training/test run, we used a randomly generated mixing matrix with entries drawn uniformly from $[0, 1]$. In each test run, we computed the correlation coefficients between the separated test sources with the original test sources (before mixing). The average correlations for each of the three discriminative sinusoidal sources (in each class condition) are shown in Figure 1 for different noise magnitudes. In order to distinguish the classes, the algorithm should give a high correlation between the separated sources and the three original discriminable sources. Notice that when the noise magnitude is smaller than the signal magnitude, the CSP algorithm is able to separate the sources well. As the noise magnitude increases, the output of the CSP algorithm shows a rapid decrease in the ability to separate the original sources. The CCA algorithm has a more gentle fall off in performance with noise magnitude but is often not as good at the lowest noise magnitudes. The CCACSP algorithm presents a nice compromise with performance as good or better than CSP at the lowest noise levels and comparable to CCA at the higher noise magnitudes.

The CCA algorithm tries to uncover independent sources while the CSP algorithm does not explicitly try to uncover independent sources. It tries to find mixtures that are discriminative. As the raw signals are discriminative, it has pressure to recover the original discriminative sources.

IV. ANALYSIS ON EEG DATASETS

Evaluation of the CCACSP algorithm was conducted using two publicly available EEG datasets. Analysis on the simulated dataset revealed that while CCACSP was better at recovering the sources in the presence of noise, CSP sometimes gave better results when SNR (signal-to-noise ratio) was high. Hence for evaluation, we included an integration of the CSP and CCACSP methods, namely merged CCACSP (mCCACSP). The two methods were integrated by merging the filters learned from the two algorithms using merging parameter $0 \leq \alpha \leq M$ where $M = \frac{\text{\# of filters}}{2}$. We selected α pair of filters from CSP (filters corresponding to the α largest and α smallest eigenvalues) and $M - \alpha$ pairs from CCACSP ($M - \alpha$ filter pairs corresponding to the largest eigenvalues from the class specific Rayleigh quotients). As is commonly used in MI tasks, $M = 3$ was used [2].
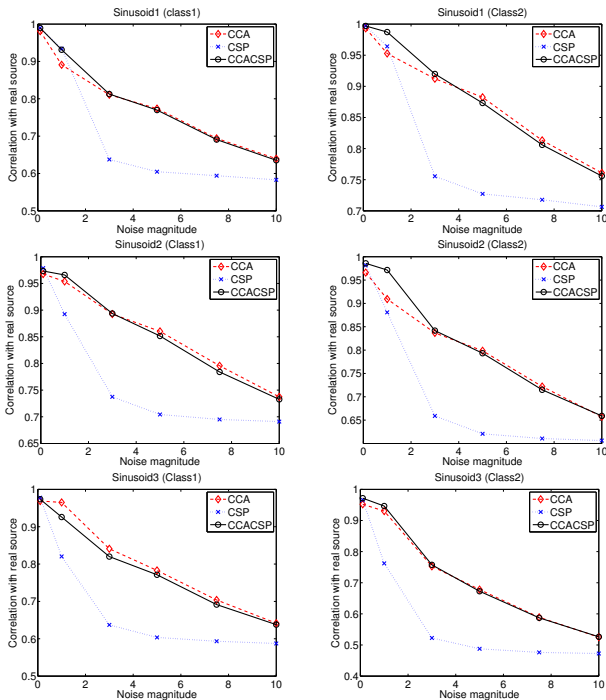
Fig. 1: Average correlation coefficients between the original and recovered discriminative sinusoidal sources as a function of noise magnitude.

A. EEG datasets

The two datasets contain EEG data from 14 subjects while they performed MI tasks. These datasets have been used in previous studies to assess classification performance of various algorithms [9], [11]. Note that the purpose of this analysis is to verify the effectiveness of the proposed feature extraction methods compared to the base case methods.

1) Dataset IVa, BCI competition *III:* This dataset [4] contains EEG signals from 5 subjects where they were instructed to perform right hand and foot MI without feedback. EEG data were recorded from 118 Ag/AgCl electrodes. A subset of 68 channels[1] as given in [11] were selected for further analysis. The size of the training and test sets varied across subjects. The number of training examples available for each class was (80, 88), (112, 112), (42, 42), (30, 26), (18, 10) for subject A1, A2, A3, A4, and A5 respectively. The number of test examples available for each class was (60, 52), (28, 28), (98, 98), (110, 114), (122, 130) for subject A1, A2, A3, A4, and A5 respectively.

2) Dataset IIa, BCI competition *IV:* This dataset [12] contains EEG signals from 9 subjects where they were instructed to perform left hand, right hand, foot, and tongue MI with feedback. Only trials corresponding to left and right hand MI were used for evaluation. EEG data were recorded from 22 Ag/AgCl electrodes. The training and test sets each contained 72 trials per class for all 9 subjects.

[1]Electrodes F, FFC, FC, CFC, C, CCP, CP, PCP, P,PPO, and PO with numbers smaller than 7 according to the International 10-20 system are used [11].

B. Preprocessing and classification

The preprocessing procedure used in [9] and [11] was applied to both datasets. Time segments between 0.5-2.5 s after cue presentation were extracted and bandpass filtered between 8-30 Hz using a fifth order Butterworth filter. No trial rejection was performed.

For CSP, 3 filters were selected from each class (each corresponding to the largest and smallest eigenvalues) for feature extraction [2]. For CCA, 6 filters corresponding to the largest eigenvalues were selected[2]. For CCACSP, 3 filters corresponding to the largest eigenvalues were selected from the two Rayleigh quotients (equations 6 and 7) resulting in a total of 6 filters. For mCCACSP, α filter pairs were selected from the CCACSP results and $3-\alpha$ filter pairs were selected from the CSP results (all of which corresponded to the largest and smallest eigenvalues). Hence the filters were equivalent to CSP and CCACSP when $\alpha = 0$ and $\alpha = 3$ respectively. The merging parameter α was determined by cross-validation on the training data. Finally, log bandpower of the filtered EEG signals were used as inputs to the LDA (linear discriminant analysis) classifier.

C. Results

Table I gives the classification errors obtained by classifying the test sets of the two datasets. In order to verify whether the proposed algorithms showed significant improvement from the base case results, paired t-tests were conducted between the error rates given by the different methods. The base case classification results using CSP were slightly different from those given in [9] and [11]. However, the average error rate was significantly lower than [9] ($p = 0.039$) and comparable to [11] which verified that our base case results agreed with previous results. The CCACSP results were not significantly better than the CSP results ($p = 0.061$). However, the mCCACSP results were significantly better than both CSP ($p = 0.015$) and CCACSP ($p = 0.023$) results.

mCCACSP provides flexibility to choose between the CSP and CCACSP algorithms. For each of the 14 subjects, mCCACSP performance on the test data is never worse than the best performance of the CSP and CCACSP algorithms. In the 5 subjects where CCACSP and CSP give identical test set performance, mCCACSP only used the CSP filters. In the 5 subjects where mCCACSP selected filters from both CSP and CCACSP, the final test performance was always better than the test performance of the two algorithms (except in the one case where the CSP algorithm had ceiling performance).

In order to verify whether the improvement of mCCACSP was related to individual performance in the base case, the dataset was divided into two groups. The first group (group 1) consisted of half of the subjects with low base case error rate (average error rate of 8 %) and the second group (group

[2]Since CCA is an unsupervised procedure, filters corresponding to the largest eigenvalues may not give the most discriminative features. Hence, an alternate method was examined where 6 filters with the lowest error on classification of training data were selected. However, the results were similar to the previous analysis suggesting that the sources separated by CCA do not effectively distinguish the two classes.

TABLE I: Comparison of classification error rates (%) for datasets IV*a* (A1-A5) and II*a* (B1-B9) from BCI competitions III and IV, respectively. Smallest values in each column are given in bold.

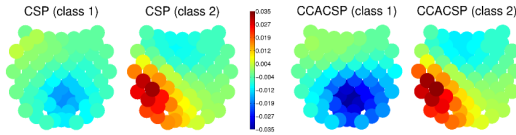| Subject	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	B6	B7	B8	B9	Mean	STD	Median
CCA	53.6	23.2	30.6	37.9	25.0	20.6	**44.4**	29.9	53.4	47.4	38.0	47.1	**3.7**	12.3	33.4	15.3	34.3
CSP	28.6	**0.0**	41.8	10.7	26.6	**9.9**	46.5	**2.9**	28.4	45.9	**35.2**	19.3	4.5	**8.5**	22.1	16.3	22.9
CCACSP	**18.8**	3.6	32.1	**7.6**	21.8	**9.9**	**44.4**	**2.9**	28.4	48.1	**35.2**	**18.6**	4.5	**8.5**	20.3	15.3	18.7
mCCACSP	**18.8**	**0.0**	27.6	**7.6**	11.5	**9.9**	**44.4**	**2.9**	25.9	44.4	**35.2**	**18.6**	4.5	**8.5**	**18.5**	**15.0**	**15.1**
α (merging param)	3	2	2	3	2	0	3	0	2	2	0	3	0	0	-	-	-

Fig. 2: Mappings of the spatial patterns produced by CSP and CCACSP for subject *A1* in dataset IV*a* from BCI competition III. The patterns corresponding to the most discriminative filters for each class are illustrated.

2) consisted of half of the subjects with high base case error rate (average error rate of 36.1 %). The paired t-test was conducted separately for the two groups. It was found that the improvements for group 2 were significant ($p = 0.019$) while the improvements for group 1 were not ($p = 0.13$). Since a couple of subjects in group 1 already had ceiling performance in the base case, this result may not be surprising. However, it is in accordance with the results found in Section III.

It was revealed that the filters learned by CSP and CCACSP showed similar patterns. However, in cases where CCACSP outperformed CSP, the CCACSP produced patterns which had sharper lateralized contrasts between the two classes. Figure 2 illustrates the patterns corresponding to the filters produced by the two methods for subject A1. The most discriminative filters for class 1 and 2 were selected from each method for illustration. We can observe that while both methods give patterns with lateral differences, the difference is more distinctive for the CCACSP algorithm.

V. CONCLUSIONS

Unlike some variations of CSP where separate recording sessions are needed (iCSP) or optimization is performed in multiple steps (Spec-CSP), CCACSP can be represented in a single objective function which can be formulated as a generalized eigenvalue problem. CCACSP incorporates the class labels to design spatial filters to increase discriminability but also uses more information from the temporal signal to help reduce noise sensitivity. Even though CCACSP results were not significantly better than the base case results, they were marginally significant ($p = 0.061$). mCCACSP outperformed both the CSP and CCACSP methods. The advantage of mCCACSP comes from the fact that it is able to select the most discriminative filters from the CSP and CCACSP results. This procedure provides the classifier with useful features while discarding less discriminative ones.

REFERENCES

[1] T. D. Bie, N. Cristianini, and R. Rosipal, "Eigenproblems in pattern recognition," in *Handbook of Geometric Computing: Applications in Pattern Recognition, Computer Vision, Neuralcomputing, and Robotics*. Springer, 2005, pp. 129–170.

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

[3] B. Blankertz, M. Kawanabe, R. Tomioka, F. Hohlefeld, V. Nikulin, and K.-R. Müller, "Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing," in *Adv. Neural Inform. Process. Syst. 20*, J. Platt, D. Koller, Y. Singer, and S. Roweis, Eds. Cambridge, MA: MIT Press, 2008, pp. 113–120.

[4] B. Blankertz, K.-R. Müller, G. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlgl, G. Pfurtscheller, J. del R. Millán, M. Schrder, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, pp. 153–159, 2006.

[5] M. Borga and H. Knutsson, "A canonical correlation approach to blind source separation," IEEE PAMI, Vol, Tech. Rep., 2001.

[6] W. D. Clercq, A. Vergult, B. Vanrumste, W. V. Paesschen, and S. V. Huffel, "Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram." *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2583–2587, 2006.

[7] S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 52, pp. 1541–1548, 2005.

[8] W. Liu, D. P. Mandic, and A. Cichocki, "Analysis and online realization of the CCA approach for blind source separation," *IEEE Trans. Neural Networks*, vol. 18, no. 5, pp. 1505–1510, 2007.

[9] F. Lotte and C. Guan, "Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms," *IEEE Trans. Bio-Med. Eng.*, vol. 58, no. 2, pp. 355–362, Feb. 2011.

[10] B. Reuderink and M. Poel, "Robustness of the common spatial patterns algorithm in the BCI-pipeline," University of Twente, Tech. Rep., July 2008.

[11] W. Samek, C. Vidaurre, K.-R. Müller, and M. Kawanabe, "Stationary common spatial patterns for braincomputer interfacing," *J. Neural Eng.*, vol. 9, no. 2, p. 026013, 2012.

[12] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. Miller, G. Mueller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the BCI competition IV," *Frontiers in Neuroprosthetics*, vol. 6, pp. 1–31, 2012.

[13] R. Tomioka, G. Dornhege, G. Nolte, B. Blankertz, K. Aihara, and K.-R. Müller, "Spectrally weighted common spatial pattern algorithm for single trial EEG classification," The University of Tokyo, Tech. Rep. 40, July 2006.

[14] H. Wang and W. Zheng, "Local temporal common spatial patterns for robust single-trial EEG classification." *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 2, pp. 131–9, 2008.

[15] H. Wang, "Discriminant and adaptive extensions to local temporal common spatial patterns," *Pattern Recognition Letters*, vol. 34, no. 10, pp. 1125 – 1129, 2013.