6th Annual International IEEE EMBS Conference on Neural Engineering
San Diego, California, 6 - 8 November, 2013

Extraction of Spatially Sparse Common Spatio-Spectral Filters with Recursive Weight Elimination

İbrahim Onaran and Nuri Fırat İnce

Abstract— **The common spatial pattern(CSP) technique linearly combines the channels to filter the neural signal spatially. It operates on data that is band-pass filtered between particular cutoff frequency for all subjects and channels. On the other hand the common spatio-spectral pattern (CSSP) method extends the the traditional CSP technique that combines spectral filtering with the original spatial filtering by using the temporally delayed version of the original data. All recording channels and the delayed versions are combined when extracting the variance as input features for a brain machine interface. This linear combination increases the number of channels extensively and results in overfitting and robustness problems of the constructed system in presence of low number of training trials. To overcome this problem, we proposed spatially sparse CSSP method in which only a subset of all available channels and its temporally shifted versions are linearly combined when extracting the features. We utilized three different versions of the recursive weight elimination (RWE) technique to select a subset of electrodes for spatio-spectral projections. We evaluate the performance of the proposed method to distinguish between the movements of the first three fingers of the hand using electrocorticogram (ECoG) signals of the brain computer interface competition 2005. We observed that spatially sparse CSSP filter outperforms both original CSP and CSSP filter and results in improved generalization in classification.**

I. INTRODUCTION

The functions of human hand such as grasping, lateral hip, pinch, etc. plays a vital role in every aspect of the activities of the human life. Several people loose their hand function due to amputation or interrupted neural pathways. Therefore, they have limitations in the activity of daily living. The brain controlled prosthetic hand, a neuroprosthetics, can help such subjects to regain their hand function and live without any assistance. In this scheme the neural decoding engine of such a prosthetics should be able to extract relevant information from brain such that it can generate necessary control output to replicate complex hand function.

The advances in electrode design and recording technology make it possible to record neurophysiological signals from a high number of channels to increase the spatial density of the recording channels to improve the accuracy and applicability of the neuroprosthetic hands. The increase in number of recording channels also increases the computational complexity of the neural decoding algorithms. To reduce the complexity of the neural signal decoding stage, the number of channels should be decreased by retaining most of the discrimination property of the oscillatory neural signal. The common spatial pattern(CSP) is widely used in BMI applications for reducing the dimensionality of the neural data. It linearly combines the recording channels into a few virtual channels in order to create a variance imbalance between competing classes [1], [2]. To use CSP method, the data needs to be band pass filtered spectrally. On the other hand,, the common spatio-spectral pattern CSSP [3] extends the CSP method in order to obtain channel specific spectral filters. The CSSP and CSP methods are successfully used in BMI applications. However they generally overfit the data when the number of training trials is limited [4], [5]. Sparse CSP method may have an important role to overcome overfitting problem. It has been shown that such methods are superior to their non-sparse counterparts in terms of generalization capability [6]–[9]. Recently, sparsification is extended to the CSSP algorithm by employing greedy solutions to select channel subsets. Due to large dimensionality originating from filter taps in CSSP the complexity of greedy solution was high. Therefore the spectral filters were limited into two taps.

In this paper, we construct spatially sparse spatio-spectral filters with several taps and study the performance of them in a BMI application. In particular, we introduced three different versions of the recently introduce RWE method [9] to select a subset of electrodes with lower complexity. In order to investigate its generalization capability, we used this method as a feature extraction engine for the classification of electrocorticogram (ECoG) related to the movements of three different fingers. Such a decoder is expected to drive a robotic hand with three fingers. We show that spatial sparsification of the spatio-spectral filters increases the classification accuracy dramatically. In the next section we first refer to the details of the traditional CSP method and then show the relation between the spatio-spectral filtering with the original CSP formulation. Next, we describe the RWE method which is used to construct spatially sparse spatio-spectral projections with low complexity. Finally we provide experimental results and our conclusion.

II. MATERIALS AND METHODS

A. Traditional CSP Method

The CSP method uses a spatial projection which is a linear combination of all recording channels to create variance imbalance of two competing classes A and B. The spatial projection is computed as follows:

$$\mathbf{X}_{CSP} = \mathbf{W}^T\mathbf{X}_i$$

This research was supported in part by The National Science Foundation, award CBET-1067488

İbrahim Onaran and Nuri Fırat İnce are with the Department of Biomedical Engineering, University of Houston, TX 77204 USA onaran@gmail.com and nfince@uh.edu

where the columns of $\mathbf{W}$ are the vectors representing each spatial projection and $\mathbf{X}_i$ is the multichannel ECoG data of the i^{th} trial.

The variance of the spatially projected data can be expressed in terms of the original data. Let the covariance matrix of the original data be $\mathbf{\Sigma}$ and the projection vector be $\mathbf{w}$, then the variance of the projected data expressed as $\mathbf{w}^T\mathbf{\Sigma}\mathbf{w}$. To create a variance imbalance between class A and B, we need to maximize the Rayleigh quotient (RQ) which is defined as follows:

$$\mathrm{RQ}(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{A}\mathbf{w}}{\mathbf{w}^T\mathbf{B}\mathbf{w}}$$

where $\mathbf{A}$ and $\mathbf{B}$ are the covariance matrices of class A and B respectively.

Maximizing the RQ is equivalent to the following optimization problem:

$$\begin{aligned} &\underset{\mathbf{w}}{\text{maximize}} && \mathbf{w}^T\mathbf{A}\mathbf{w} \\ &\text{subject to} && \mathbf{w}^T\mathbf{B}\mathbf{w} = 1. \end{aligned} \tag{1}$$

To solve the optimization problem expressed in (1), we use Lagrange multipliers method to obtain the equivalent problem in the form of $\mathbf{A}\mathbf{w} = \lambda\mathbf{B}\mathbf{w}$ which is generalized eigenvalue decomposition (GED) of the covariance matrices $\mathbf{A}$ and $\mathbf{B}$. The solutions $\mathbf{w}$ to this problem are the joint eigenvectors of $\mathbf{A}$ and $\mathbf{B}$ and λ is the associated eigenvalue for a particular joint eigenvector.

B. Common Spatio-Spectral Filters (CSSP)

The common spatio-spectral pattern (CSSP) method of [3] is an extension of the traditional CSP that filters the data spectrally and spatially at the same time. The multichannel data for the i^{th} trial is denoted as $\mathbf{X}_i \in \mathbb{R}^{C\times N}$ where C is the number of channels and N is the number of samples. A delayed trial is obtained $\mathbf{X}_{i,m}$, where m denotes the delay index. Then the delayed versions of this trial is concatenated to form an enlarged space. The delay index m varies from 0 to M where M is the order of spectral filter. The traditional CSP algorithm is employed on the enlarged space. Let the enlarged trial data be $\mathbf{E}_i \in \mathbb{R}^{C(M+1)\times N}$ and the CSP solution of this enlarged data set be $\mathbf{W} \in \mathbb{R}^{C(M+1)\times D}$, where D is number of projections. The projection of the enlarged trial data can be expresssed in terms of the signal and its delayed versions as follows:

$$\mathbf{Y}_i = \mathbf{W}^T\mathbf{E}_i = \sum_{m=0}^{M} \mathbf{W}_m^T\mathbf{X}_{i,m}$$

Here, $\mathbf{W}_m \in \mathbb{R}^{C\times D}$ denotes the collection of spatial filters that are applied to signal which is delayed with the amount of m. It should be noted that the filter weights in $\mathbf{W}_m$ are obtained from the GED solution of the enlarged covariance matrices. Since each time sample of the spatially filtered signal $\mathbf{Y}_i \in \mathbb{R}^{D\times N}$ is expressed as the linear combination of the original signal and its delayed versions, the whole process described so far can be interpreted as an application of the CSP algorithm to the finite impulse response (FIR) filtered multi-channel neural data, hence, it is named as common spatio-spectral pattern (CSSP). The advantage of this approach is that it extracts the spectral (frequency) information from the data that is specific to the subject and the channel. This advantage on the other hand comes with an additional (delay) parameter, M, to be tuned during the training phase. Moreover, the enlarged space, which results in a much larger covariance matrix of size $C(M+1)$ while solving the GED, causes overfitting and reduces the generalization capability of the classifier. Therefore, in practice a single delay is used which corresponds to a dimensionality of $2C$. As expected, this short filter also yields poor spectral resolution [3].

These drawbacks led us to find a spatially sparse solution to increase the robustness and generalization capability of the CSSP method with higher order spectral filters. In more detail, only a small number of channels with several temporal delays will be used to extract features. However, finding such a subset is trivial and computationally complex. Consequently, a fast technique is needed to eliminate unnecessary channels. In the past few years, several greedy methods based on ℓ_0 norm such as backward elimination (BE), forward selection (FS) were proposed to find sparse spatial filters [6]. However, they have high computational complexity when the number of channels is high. Therefore, we used a modified version of recursive weight elimination method as described in [9]. It has been shown that the RWE has dramatically lower computationally complexity and provides comparable performance to other ℓ_0 norm based greedy sparse methods [9].

C. The Recursive Weight Elimination (RWE)

The recursive weight elimination (RWE) method is inspired from the support vector machine (SVM) based recursive feature elimination framework (SVM-RFE) which is described in [10]. They eliminated features iteratively corresponding to the minimum entry of the weight vector of the SVM, because the elements of the weight vector is assumed to contribute maximum margin according to their values. Like in [10], the RWE method assumes that the coefficients of the spatio-spectral filters are related to their contribution to the RQ. The RWE algorithm starts with a full size covariance matrices of the traditional CSP method. Initially, RWE solves general CSP problem and finds the weight vector $\mathbf{w}$. The absolute value of the coefficients in $\mathbf{w}$ is sorted and the channel associated with the smallest value of weight vector is eliminated. In the next step, the GED solution is obtained from the remaining channel set. This elimination procedure is iterated until the method reaches the desired cardinality (number of channels used channels in final projections). It is worth noting that using sparseness in CSSP framework was attempted in [11]. We have to emphasize that in [11] sparseness is sought in the length of the spectral FIR filter which was allowed to be arbitrarily long. On the contrary, we are seeking sparseness in the spatial domain.

In our exploration the goal is to analyze the effect of sparseness in CSSP as a function of the number of channels.

For this particular purpose, we modify the RWE search method such that a channel and its delayed versions are eliminated at each step. This is accomplished as follows. After solving the GED in each step on the enlarged covariance matrices, the absolute value of the weight vector is computed. In the following step we study three different methods, to remove a particular channel and its delayed indices. Specifically we compute

i. the value corresponding to the maximum,
ii. the value corresponding to the minimum and
iii. the mean value

for each channel over its delayed weight indices. Next, as in the original RWE, the stored values are sorted in descending order and the channel associated with minimum value is eliminated along with its delays. Similar to SVM-RFE, we estimated the channel with minimum contribution to the RQ using three different measures. The procedure is iterated until desired cardinality over channels is reached. We also investigated the effect of the number of delays (M) of the spectral filters in classification. In our experiments M ranged from 0 (regular sparse CSP method) to 5.

D. ECoG Dataset

We applied the sparse CSSP method on multiclass ECoG of BCI competitions IV. The ECoG data was recorded from three subjects during finger flexions and extensions [12] with a sampling rate of 1 kHz. Each electrode array contained $48(8\times6)$ or $64(8\times8)$ platinum electrodes. The finger index to be moved was shown with a cue on a computer monitor. The subjects moved one of their five fingers three to five times during the cue period. The ECoG data of each subject was sub band filtered in the frequency range of $(40-200$ Hz). We used 1 s data following the movement onset in the analysis. The dataset contains around 146 trials for each subject. In this paper, we applied sparse CSSP filter to discriminate between the movements of three fingers only as they can achieve almost all the functionality that five fingers realize, and therefore widely used in robotic hand design [13] due to light weight.

The signal was transformed into four virtual channels by taking first and last two eigenvectors of the GED solution. After obtaining first and last eigenvector, we deflated the covariance matrices with Schur decomposition using the first and the last eigenvector for the sparse filters [14]. After computing the outputs of these four spatio-spectral filters, we calculated the energy of the output signal and converted it to log scale and used them as input features to lib-SVM classifier with an RBF kernel [15]. Since we are tackling a multiclass problem, we used the pairwise discrimination strategy of [2] for the three-class finger movement data.

We studied the classification accuracy as a function of cardinality over channels and the amount of delay of the spectral filter. On the training data with the purpose of finding optimum sparseness level for the classification, we computed several sparse solutions with cardinalities 40, 30, 20, 15, 10, 5, 2, 1. A five fold cross validation was applied to the dataset to study the generalization accuracy. The classification accuracies are averaged over subjects.

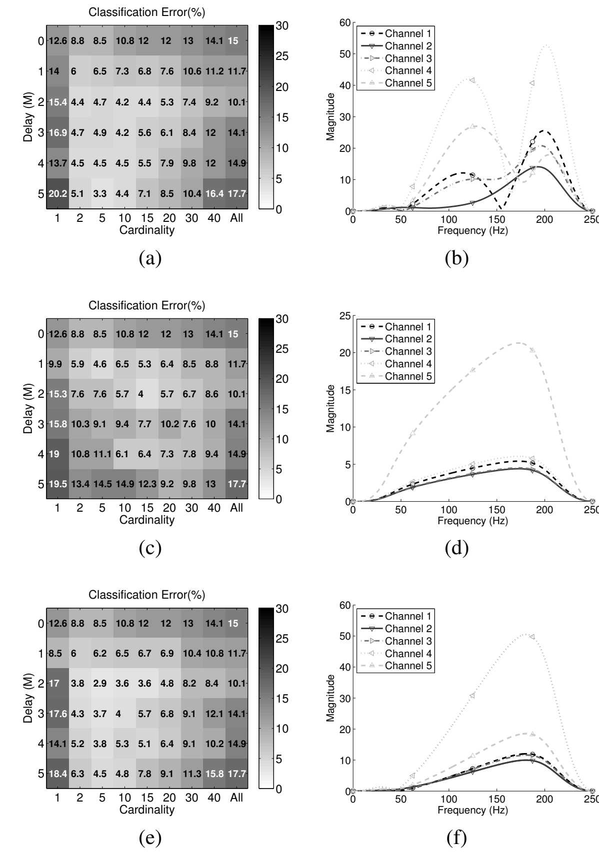

Fig. 1. The average classification errors over three subjects for different cardinalities and delays obtained using the max (a), min (c) and average (e) RWE. The last column reflects the errors for full CSSP method. Also note that the last column of the first row is the accuracy of traditional CSP method. The corresponding spectral filters of the sparse CSSP solution are plotted on the right side in figures (b),(d) and (f). (b) 6 taps (M=5) for cardinality 5. (d) 3 taps (M=2) for cardinality 5. (f) 3 taps (M=2) for cardinality 5. The effect of 40-200 Hz band filtering also included.

III. RESULTS

In Fig. 1 the classification errors for different cardinality and filter lengths are given on the left side. The last error points correspond to the errors that are obtained by the original CSSP method working on all channels. The first row reflects the results that are obtained from spatial filters, only. Consequently, the error rate of on the last column of the first row corresponds to the traditional CSP method. We observed that the minimum classification error 2.9% was obtained by average RWE method with a cardinality of 5 and delay of 2. For maximum and minimum RWE methods we obtained 3.3% and 4% error with cardinalities 5 and 15 and delays 5 and 2 respectively. Using sparseness in the spatial domain improved the CSSP classification performance considerably, especially when the spatial filters are highly to moderately sparse. The best performances are obtained by spatial filters that are considerably sparse. Nevertheless, the spatially sparse CSSP method provided lower classification error rates for all delays. The standard CSP and CSSP methods had a minimum error rate of 15% and 10.1% (with a delay 2) respectively.

The frequency response of one of the 6 taps filter obtained by spatially sparse CSSP method at cardinality 5 with maximum RWE is shown in Fig. 1b. Each line represents the frequency response of a different channel. The effect of 40-200 Hz band filtering also included in the plot. The spectral filters that are formed by sparse CSSP are simply band-pass filters. The filters suppress the signal around 50 Hz and 160 Hz. This is not the case for average and minimum RWE, since they have produce a lower order spectral filters. The filters tend to select higher bands in 40-200 Hz range.

IV. CONCLUSION

The CSP and CSSP methods suffer from overfitting in the presence of high density recordings with small amount of training data. Since each delay expands the dimensionality of the covariance matrices used in GED, the CSSP method tends to overfit the training data more than CSP. Here, we constructed spatially sparse CSSP method in which only a subset of all available channels is linearly combined when extracting the features. To select a subset of spatio-spectral projections, we utilized three versions of a modified recursive weight elimination (RWE) technique which is recently introduced. We evaluate the performance of the proposed method to distinguish the movement of the first three fingers of the hand using electrocorticogram (ECoG) signals of the BCI competition 2005. We observe that spatially sparse spatio-spectral filters are superior to both original CSP and CSSP.

REFERENCES

[1] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing Spatial Filters for Robust EEG Single-Trial Analysis," *Signal Processing Magazine, IEEE*, vol. 25, no. 1, pp. 41 –56, 2008.

[2] N. F. İnce, R. Gupta, S. Arica, A. H. Tewfik, J. Ashe, and G. Pellizzer, "High Accuracy Decoding of Movement Target Direction in Non-Human Primates Based on Common Spatial Patterns of Local Field Potentials," *PLoS ONE*, vol. 5, no. 12, p. e14384, December 2010.

[3] S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller, "Spatio-Spectral Filters for Improving the Classification of Single Trial EEG," *Biomedical Engineering, IEEE Transactions on*, vol. 52, no. 9, pp. 1541 –1548, September 2005.

[4] J. Farquhar, N. J. Hill, T. N. Lal, and B. Schlkopf, "Regularised CSP for Sensor Selection in BCI," in *In Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course*, 2006.

[5] B. Reuderink and M. Poel, "Robustness of the Common Spatial Patterns Algorithm in the BCI-Pipeline," Univ. of Twente, Enschede, July 2008.

[6] F. Göksu, N. F. İnce, and A. Tewfik, "Sparse Common Spatial Patterns in Brain Computer Interface Applications," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, May 2011, pp. 533 –536.

[7] İ. Onaran, N. F. İnce, and A. E. Çetin, "Sparse Spatial Filter via a Novel Objective Function Minimization with Smooth ℓ_1 Regularization," *Biomedical Signal Processing and Control*, no. 3, pp. 282 – 288, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1746809412001103

[8] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing the Channel Selection and Classification Accuracy in EEG-Based BCI," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 6, pp. 1865 –1873, June 2011.

[9] F. Göksu, N. F. İnce, and İ. Onaran, "Sparse Common Spatial Patterns with Recursive Weight Elimination," in *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*, November 2011, pp. 117 –121.

[10] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene Selection for Cancer Classification Using Support Vector Machines," *Machine Learning*, vol. 46, no. 1-3, pp. 389–422, 2002. [Online]. Available: http://citeseer.ist.psu.edu/guyon02gene.html

[11] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Müller, "Combined Optimization of Spatial and Temporal Filters for Improving Brain-Computer Interfacing," *Biomedical Engineering, IEEE Transactions on*, vol. 53, no. 11, pp. 2274 –2281, November 2006.

[12] K. J. Miller and G. Schalk, "Prediction of Finger Flexion 4th Brain-Computer Interface Data Competition," 2008, http://www.bbci.de/competition/iv/desc_4.pdf.

[13] S.-y. Jung, S. kyun Kang, M.-J. Lee, and I. Moon, "Design of Robotic Hand with Tendon-Driven Three Fingers," in *Control, Automation and Systems, 2007. ICCAS '07. International Conference on*, 2007, pp. 83–86.

[14] L. Mackey, "Deflation Methods for Sparse PCA," in *Advances in Neural Information Processing Systems 21*, D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, Eds., 2009, pp. 1017–1024.

[15] C.-C. Chang and C.-J. Lin, "A Library for Support Vector Machines." 2001, http://www.csie.ntu.edu.tw/~cjlin/libsvm.