SPARSE SPATIAL FILTER OPTIMIZATION FOR EEG CHANNEL REDUCTION IN BRAIN-COMPUTER INTERFACE

Xinyi Yong, Rabab K. Ward, Gary E. Birch

Department of Electrical and Computer Engineering, University of British Columbia,
2332 Main Mall, Vancouver, BC, Canada V6T 1Z4, email: yongy@ece.ubc.ca

ABSTRACT

Spatial filters are useful in discriminating different classes of electroencephalogram (EEG) signals such as those corresponding to motor activities. In the case of discriminating two classes of signals, EEG signals are projected onto a space where one class of signals is maximally scattered and the other is minimally scattered. This paper finds a minimal number of electrodes that can achieve the discrimination. Applying many electrodes is tedious and time-consuming. To reduce the number of electrodes, we propose inducing sparsity in the spatial filter. We reformulate the optimization problem in Common Spatial Patterns by introducing an ℓ_1-norm regularization term. Experimental results on five subjects show that the proposed method significantly reduces the number of electrodes while generating features with good discriminatory information. The number of electrodes on average, is reduced to 11% (of the 118 electrodes) while the average drop in the classification accuracy is only 3.8%.

Index Terms— Electroencephalogram, Brain-Computer Interface, Common Spatial Patterns, Optimization, Regularization Term

1. INTRODUCTION

EEG signals, recorded using a set of electrodes placed over the scalp, can be used for discriminating different mental tasks. Different signals recorded at different scalp sites, however, do not provide the same amount of discriminatory information [1]. By associating weights w_i to the signals obtained from different electrodes prior to their processing, we essentially perform spatial filtering and can thereby improve the discrimination task.

Common Spatial Patterns (CSP) is a method commonly used to find spatial filters for classification of multichannel EEG signals. After the CSP algorithm's first use in extracting abnormal components from clinical EEG [2], it has been widely used in brain-computer interface (BCI) research to extract features from EEG signals. Ramoser *et al.* [3] demonstrated that spatial filters for multichannel EEG signals, which are derived using CSP, can effectively extract discriminatory information from two classes of EEG signals, namely the left and right hand motor imageries. As an extension, Dornhege *et al.* [4] proposed the CSSSP algorithm, which allows the simultaneous optimization of an FIR filter and a spatial filter for the automatic selection of the subject-specific frequency band, which the CSP operates on.

The CSP algorithm finds the directions where the EEG signals should be projected onto so that the differences between any two classes of EEG signals are maximized (i.e. the variance of one class is minimized while at the same time, the variance of the other class is maximized) [1]. The directions are given by a weight matrix whose rows give the weights of the EEG channels. Useful features can be extracted from the projected EEG signals and then used for classification.

The weights of the spatial filters generated from both the CSP and CSSSP algorithms are dense (not sparse), thus a larger number of electrodes is needed during signal acquisition. For practical applications such as BCI's, applying many electrodes is undesirable since the preparations prior to recording the signals can be time-consuming and troublesome. To reduce the number of channels used in BCIs, Wang *et al.* [5] proposed the use of weight vectors that are obtained from CSP. Four optimal channels were selected using the large coefficients of the CSP weight vectors. Two of these were based on the neurological phenomena of event-related desynchronization (ERD) and the readiness potential respectively. However, by eliminating other channels, the remaining signals can no longer be projected onto the direction that best discriminates the two classes of EEG signals (even though other feature extraction methods can be used for that purpose), the performance will not be optimal. Thus, in this paper, we seek to determine a minimal set of EEG channels (i.e., the smallest number of electrodes) that can be used to discriminate between two classes of EEG signals using spatial filters. The use of a smaller number of electrodes will deteriorate the performance achieved when all EEG channels are used. Thus, the loss in the performance due to using a smaller number of electrodes should also be minimized. In order to achieve this, we modify the CSP algorithm by introducing an ℓ_1-norm regularization term to encourage sparsity in the weight vector of the spatial filter, w.

2. METHODOLOGY

2.1. Data Description

The EEG data used in this study consisted of two classes: right hand and right foot motor imageries. They were provided by Fraunhofer FIRST (Intelligent Data Analysis Group) and Campus Benjamin Franklin of the Charité - University Medicine Berlin (Neurophysics Group) [6]. The EEG signals were recorded from five subjects using 118 electrodes per subject. The extended International 10-20 system at a sampling rate of 1 kHz was employed. During each experiment, the subject was given visual cues that indicated for 3.5s which of the three motor imageries should be performed: left hand, right hand and right foot. The resting interval between two trials was randomized from 1.75 to 2.25 seconds. Only EEG trials for right hand and right foot were provided. Each class of EEG signals consists of 140 trials.

2.2. Data Preprocessing

We used EEG data that were downsampled to 100 Hz. The data were then band-pass filtered to the 8–35 Hz frequency band. This band encompasses the Mu and Beta rhythms which have been reported to desynchronize during motor imagery [7]. These neurological phenomena have been used successfully in BCI systems to classify EEG signals [1, 3, 5].

2.3. Problem Formulation

The data consist of $N_e = 118$ EEG channels. There are two classes of EEG signals: Class 1 (right hand) and Class 2 (right foot), with each class containing M trials. We seek to find a spatial filter or weight vector such that the signals can be projected onto a 1-dimensional space where one class of signals is maximally scattered and the other is minimally scattered. High variance of the signals indicates strong rhythms whereas low variance indicates attenuated rhythms [1].

For example, during a right hand imagined movement, ERD occurs and the motor rhythms are attenuated. Hence, we can find a spatial filter such that the EEG signals corresponding to a right hand motor imagery has minimal variance. This can be done by solving an optimization problem. The criterion or the cost function used in this optimization problem is the variance of the projected Class 1 signals. It is minimized while keeping the sum of the variances of both signal classes fixed [1]. A sparse solution of the weight vector is desired. (Note: During a right foot movement, ERD also occurs and we can then find a spatial filter such that the EEG signals corresponding to a right foot motor imagery has minimal variance. We can achieve this by maximizing the same cost function mentioned above.)

Let $S = \{S_1, S_2, \ldots, S_{2M}\}$ where $S_i \in \mathbb{R}^{N_e \times N}$ denotes the filtered i-th trial EEG signal and N the number of samples in the signal (350 in this study). The optimization problem is expressed as:

$$\begin{aligned} \underset{w}{\text{minimize}} \quad & \sum_{i \in \mathcal{C}_1} \text{var}(w^T S_i) \\ \text{subject to} \quad & \sum_{i=1}^{2M} \text{var}(w^T S_i) = 1 \\ & \|w\|_0 \equiv \sum_{i=1}^{N_e} |w_i|^0 \leq k \end{aligned} \tag{1}$$

where $\mathcal{C}_1$ represents all Class 1 EEG trials and $w \in \mathbb{R}^{N_e}$ is the unknown weight vector of the spatial filter.

With only the first constraint in (1), the optimization problem reduces to the original CSP [1]. The second constraint was introduced in this study to minimize the number of electrodes (less than a positive integer, k). $\|w\|_0$ (ℓ_0-norm of w) is a pseudo-norm giving the number of non-zero elements in the vector with

$$|w_i|^0 = \begin{cases} 0 & w_i = 0 \\ 1 & \text{otherwise} \end{cases}$$

We can express the cost function in (1) using the definition of variance, i.e.,

$$\begin{aligned} \text{var}(w^T S_i) &= w^T E\{(S_i - \text{E}\{S_i\})(S_i - \text{E}\{S_i\})^T\} w \\ &= w^T \Sigma_i w \end{aligned}$$

where Σ_1 and Σ_2 are the mean covariance matrices for the signals belonging to sets $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively. Solving a problem with the ℓ_0-norm is combinatorial in nature and computationally prohibitive to solve for large problems. Thus, the ℓ_0-norm is replaced by an ℓ_1-norm ($\|w\|_1 \equiv \sum_{i=1}^{N_e} |w_i|$), which also promotes sparsity [8]. The formulation of the proposed sparse CSP (SCSP) problem can then be rewritten as

$$\begin{aligned} \underset{w}{\text{minimize}} \quad & w^T \Sigma_1 w + \rho \|w\|_1 \\ \text{subject to} \quad & w^T (\Sigma_1 + \Sigma_2) w = 1 \end{aligned} \tag{2}$$

where $\rho > 0$ is an appropriate regularization parameter that controls the sparsity of the solution. When $\rho = 0$, the solution is essentially the same as the one obtained using CSP.

Now, letting $w = w^+ - w^-$ where $w^+ = \max(w, 0)$ and $w^- = \max(-w, 0)$, we can express (2) as a Quadratically Constrained Quadratic Programming (QCQP) problem expressed by

$$\begin{aligned} \underset{\tilde{w}}{\text{minimize}} \quad & \tilde{w}^T D \Sigma_1 D^T \tilde{w} + \rho c^T \tilde{w} \\ \text{subject to} \quad & \tilde{w}^T D(\Sigma_1 + \Sigma_2) D^T \tilde{w} = 1 \\ & \tilde{w} \geq 0 \end{aligned} \tag{3}$$

where $\tilde{w}$, c and D are defined as:-

$$\tilde{w} = \begin{pmatrix} w^+ \\ w^- \end{pmatrix}; \quad c = \begin{pmatrix} \mathbf{1} \\ \mathbf{1} \end{pmatrix}; \quad D = \begin{pmatrix} I \\ -I \end{pmatrix}$$

This is a non-convex programming problem because of the quadratic equality constraint. It can be solved using several methods such as sequential quadratic programming (SQP) and augmented Lagrangian methods. The QCQP problem (3) was solved using the software package NPSOL, available in TOMLAB (http://tomopt.com/tomlab/). The software uses SQP to solve nonlinear programming problems.

For the optimization of (3), the cost function's gradient and Hessian are required. These are given by $2D\Sigma_1 D^T\tilde{w}+\rho c$ and $2D\Sigma_1 D^T$ respectively. In addition, w needs to be appropriately initialized in the iterative optimization algorithm. In this study, the starting point used is a feasible point, which is the optimal value obtained from the CSP algorithm ($\rho = 0$). Furthermore, ρ that controls the sparsity of the solution (w) has to be appropriately chosen. It is varied from 0.0 to 1.0. It is subject specific and is chosen manually in this study based on the number of electrodes reduced and the classification accuracy achieved.

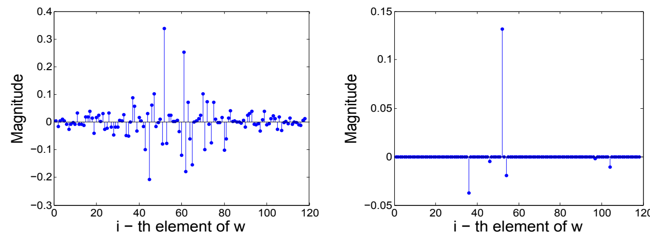
2.4. Feature Extraction and Classification

Two spatial filters were obtained by minimizing and maximizing the cost function in (3). Minimizing the variance of Class 1 reflects the neurological phenomenon of event-related desynchronization (ERD) of the motor rhythms whereas maximizing the variance of Class 1 reflects the synchronization of the motor rhythm [1]. From the preliminary results, the spatial filter obtained by minimizing the variance of Class 1 yielded a good discriminating feature. However, the second spatial filter obtained by maximizing the variance of Class 1 did not give good discriminating features for any of the subjects. Therefore, only the first spatial filter is used in extracting features from the data.

The spatial filter is used to project the signals and the variance of the projected signals is the only feature used in the classification. Linear Discriminant Analysis (LDA) is used for classification. The number of non-zero elements in the weight vector of the spatial filter and the averaged 10×10 fold cross-validation classification accuracy are used as performance metrics. Note that no artifact rejection algorithm was used in this study.

3. EXPERIMENTAL RESULTS

In this study, the regularization parameter ρ is selected manually. As ρ increases, the solution becomes more sparse. Fig. 1 shows the values of the elements of w when CSP and our proposed SCSP were used in estimating the weight vector. The proposed SCSP successfully produced a sparse weight vector (with only 7 non-zero elements) when $\rho = 0.47$ was used. However, there is a trade-off between ρ and the classification accuracy. Therefore, ρ has to be chosen carefully to produce reasonable results (acceptable accuracy with a minimal set of electrodes selected). This varies between subjects. For example, in subject *aw*, with $0.0 < \rho < 0.3$, the classification accuracy obtained lies in the range of 80% to 89%. Using $\rho = 0.32$, 13 out of 118 electrodes are selected with an accuracy of approximately 84.4%.



(a) (b)

Fig. 1. Weight vector, w obtained using: (a) CSP; (b) the proposed SCSP when $\rho = 0.47$.

Table 1 presents the number of non-zero elements and the averaged 10×10 cross-validation classification accuracy achieved using three different methods in feature extraction: CSP, the proposed SCSP and CSPv. CSPv uses CSP to select two optimal channels (the two largest coefficients in the weight vector) as suggested by [5]. The features used are the variance of the EEG signals corresponding to the optimal channels. For CSP and the proposed SCSP, only one feature is used in the classification.

Table 1. Performance comparison of CSP, our proposed SCSP and CSPv. (Sbj: Subject; Ave: Average.)

Sbj	**# non-zero elements of w (ρ)**			**Classification accuracy (%)**		
	CSP	SCSP	CSPv	CSP	SCSP	CSPv
aa	118	17 (0.02)	2	58.6	57.5	51.0
av	118	5 (0.44)	2	52.5	54.4	52.4
al	118	9 (0.40)	2	92.1	86.9	82.5
aw	118	13 (0.32)	2	91.8	84.4	66.1
ay	118	20 (0.18)	2	91.4	84.3	71.3
Ave	118	13	2	77.3	73.5	64.7

CSP and the proposed SCSP successfully extracted features from the EEG signals except for subject *aa* and *av*. CSP yields higher classification accuracy but it requires the use of all 118 electrodes in obtaining the features. On the other hand, the proposed SCSP requires significantly fewer electrodes (ranging from 5 to 20) even though the classification accuracy on average drops from 77.3% to 73.5%, i.e., by 3.8% only. Applying many electrodes in practical BCI applications is undesirable because it is time-consuming and troublesome. Using fewer electrodes speeds up the preparation process and benefits both the operator and the user.

The accuracy of the proposed SCSP is higher than CSPv as only two electrodes were used in CSPv. Even if the number of optimal electrodes were increased to, for example, 13 in subject *aw*, the accuracy obtained was 71.4%, which is still less than that of the proposed SCSP (84.4%). This shows that all projected signals provide better discriminating features.

Fig. 2 shows that the proposed method is successful in clearly discriminating between the right hand and right foot motor imageries EEG signals. The ERD plot of the projected signals shows that the right hand EEG signals has the lower variance and the right foot EEG signals has the higher variance.

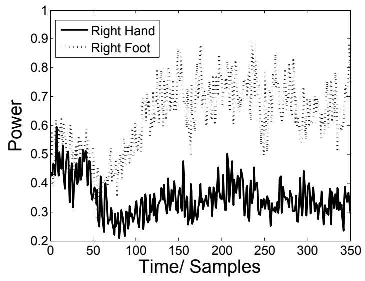


Fig. 2. The averaged ERD time course for the projected right hand and right foot EEG signals using the proposed SCSP method.

The proposed SCSP not only extracts discriminatory information from the signals, the non-zero elements of w obtained from the method can possibly provide useful spatial information about the cortex area that best discriminates any two classes of signals. This is subject-specific. For example, in subject *al*, *aw* and *ay*, the maximum value of the w corresponds to the electrode that lies on the left side of the primary motor cortex area of the brain (C3, CCP5 and PCP3). This agrees with findings in the literature that the contralateral motor cortex area is activated during hand motor imagery [7].

4. CONCLUSIONS

In this study, we focused on demonstrating how to obtain a minimal set of electrodes that can be used to find the direction that best discriminates between two classes of EEG signals. This is achieved by introducing an ℓ_1-norm regularization term in the CSP algorithm, which encourages sparsity in the weights of the spatial filter. The spatial filter can then project the EEG signals onto a space where one class of signals is minimally scattered and the other is maximally scattered. The variance of the resulting signals carries good discriminatory information.

It was demonstrated that a suitably chosen regularization parameter ρ can produce a sparser solution with only a small effect on the generalization ability of the classifier. The number of electrodes, on average, is successfully reduced by 89%, i.e., from 118 electrodes to 13 electrodes even though the average drop in the classification accuracy is only 3.8%. Hence, only a small number of electrodes is required in the signal acquisition and the processing of EEG signals. This can benefit both the operator and the user. Furthermore, the non-zero elements (signals corresponding to the selected electrodes) are deemed relevant in providing us useful subject-specific spatial information of the EEG sources. The ρ value is selected manually in this study. In the future, we plan to perform a more detailed analysis on the effect of ρ on the performance of the system and devise an automatic algorithm to select a subject-specific ρ.

Our proposed SCSP method gives the flexibility of choosing the number of electrodes desired by adjusting the value of the regularization parameter. This method can potentially be used in conjunction with other feature extraction methods to improve the classification accuracy.

5. REFERENCES

[1] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K-R. Müller, "Optimizing spatio-temporal filters for improving Brain-Computer Interfacing," in *Advances in Neural Inf. Proc. Systems (NIPS05)*, J. Platt, Ed., Vancouver, Canada, December 2005, vol. 18.

[2] Z. J. Koles, "The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 79, no. 6, pp. 440–447, 1991.

[3] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–447, 2000.

[4] 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," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 11, pp. 2274–2281, 2006.

[5] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selection in motor imagery based Brain-computer Interface," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the IEEE*, 2005, pp. 5392–5395.

[6] "Data Set IVa for the BCI Competition III," http://ida.first.fraunhofer.de/projects/bci/competition_iii/desc_IVa.html.

[7] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical Neurophysiology*, vol. 110, pp. 1842–1857, 1999.

[8] Y. Tsaig and D. L. Donoho, "Breakdown of equivalence between the minimal ℓ^1-norm solution and the sparsest solution," *Signal Processing*, vol. 86, pp. 533–548, 2006.