Technical note

Sparse spatial filter via a novel objective function minimization with smooth ℓ_1 regularization

Ibrahim Onaran [a,b,*], N. Firat Ince [a,c], A. Enis Cetin [b]

[a] Department of Neurosurgery, University of Minnesota, Minneapolis, MN 55455, USA
[b] Department of Electrical Engineering, Bilkent University, Ankara, Turkey
[c] Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, USA

ARTICLE INFO

ABSTRACT

Common spatial pattern (CSP) method is widely used in brain machine interface (BMI) applications to extract features from the multichannel neural activity through a set of spatial projections. These spatial projections minimize the Rayleigh quotient (RQ) as the objective function, which is the variance ratio of the classes. The CSP method easily overfits the data when the number of training trials is not sufficiently large and it is sensitive to daily variation of multichannel electrode placement, which limits its applicability for everyday use in BMI systems. To overcome these problems, the amount of channels that is used in projections, should be limited to some adequate number. We introduce a spatially sparse projection (SSP) method that exploits the unconstrained minimization of a new objective function with approximated ℓ_1 penalty. Unlike the RQ, this new objective function depends on the magnitude of the sparse filter. The SSP method is employed to classify the multiclass ECoG and two class EEG data sets. We compared our results with a recently introduced sparse CSP solution based on ℓ_0 norm. Our method outperforms the standard CSP method and provides comparable results to ℓ_0 norm based solution and it is associated with less computational complexity. We also conducted several simulation studies on the effect of noisy channel and intersession variability on the performance of the CSP and sparse filters.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

The BMI technology aims to help disabled people to establish communication with their environment solely by their brain signals. With the recent advances in electrode design and recording technology, the number of recording channels used in BMI applications is increasing to capture signals from a larger area of the brain or to get more information from smaller regions using dense electrode grids. Therefore, a dimension reduction algorithm needs to be employed to decrease the correlation between channels and improve the signal to noise ratio (SNR). In this scheme, the CSP algorithm is widely used due to its simplicity and lower computational complexity to extract features from high-density recordings both using noninvasive and invasive modalities [1,2].

Despite the benefits of the CSP method, it also has a number of drawbacks. One major problem of the CSP is that it generally overfits the data when it is recorded from a large number of electrodes and when there is limited number of train trials.

Moreover, the chance that CSP uses a noisy or corrupted channel is linearly increased with increasing number of recording channels. Robustness over time is also a major drawback in CSP applications [3,4]. Since all channels are used in spatial projections of CSP, the classification accuracy may reduce in case the electrode locations slightly change in different sessions. This requires almost identical electrode positions over time, which is difficult to realize [5]. The sparseness of the spatial filter might have an important role to increase the robustness and generalization capacity of the BMI system.

The CSP method minimizes the Rayleigh Quotient (RQ) of the spatial covariance matrices to achieve the variance imbalance between the classes of interest. The RQ is defined as

$$R(w) = \frac{w^T A w}{w^T B w} \tag{1}$$

where A and B are the spatial covariance matrices of two different classes and w is the spatial filter that we want to find. One way to reduce the number of channels used in the projection w, is to transform the CSP algorithm into a regularized optimization problem in the form of

$$L(w) = R(w) + \lambda ||w|| \tag{2}$$

where $R(w)$ is the objective function, $||w||$ is the ℓ_1 norm based penalty and λ is a constant that controls the sparsity of the solution. In the past few years, there is growing interest in using ℓ_1 penalty to construct sparse solutions. However, RQ does not depend on the magnitude of the sparse filter. Therefore, RQ cannot be directly used in a norm based minimization problem, since the optimizer always minimizes the norm along the direction which RQ has been minimized.

A number of studies investigated putting the CSP into alternative optimization forms to obtain a sparse solution for it. In [6] the authors converted CSP into a quadratically constrained quadratic optimization problem with ℓ_1 penalty; others used an ℓ_1/ℓ_2 [3,7] norm based solution. These studies have reported a slight decrease or no change in the classification accuracy while decreasing the number of channels significantly. Recently, in [8] quasi ℓ_0 norm based criterion was used for obtaining the sparse solution which resulted an improved classification accuracy. Since ℓ_0 norm is non-convex, combinatorial and NP-hard, they implemented greedy solutions such as Forward Selection (FS) and Backward Elimination (BE) to decrease the computational complexity. It has been shown that BE was better than FS (less myopic) in terms of classification error and sparseness level but associated with very high complexity making it difficult to use in rapid prototyping scenarios.

In this paper, we construct a computationally efficient spatially sparse projection (SSP) based on a novel objective function with similar characteristics to RQ. This new objective function can be minimized in the form of (2) to address the drawbacks of regular CSP method. We show that our new objective function has the same minimization solution as RQ and it depends on the magnitude of the spatial filter. The magnitude dependency of our new objective function allows us to use a continuous and differentiable function approximating ℓ_1 norm [9] as regularization term in an unconstrained optimization framework and can be solved using standard algorithms with low complexity. The rest of the paper is organized as follows. In the following section, we describe our novel objective function and its relation to RQ. Then we explain its use in an unconstrained optimization problem. Next, we apply our method on the BCI competition IV ECoG dataset involving individuated movements of five fingers [10] and the BCI competition III EEG dataset IVa [11] involving imaginary foot and hand movements. We also compare our method to standard CSP and the ℓ_0 norm based BE solution given in [8]. Finally, we studied the effect of additional Gaussian noise and simulated channel displacements on the classification accuracy and discuss our results and provide future directions.

2. Materials and methods

2.1. Standard CSP and a new objective function

In the CSP framework, the spatial filters are a weighted linear combination of recording channels, which are tuned to produce spatial projections maximizing the variance of one class and minimizing the other. The spatial projection is computed using

$$X_{CSP} = W^T X \tag{3}$$

where the columns of W are the vectors representing each spatial projection and X is the multichannel ECoG data.

Maximizing the RQ (1) is identical to the following optimization problem.

$$\begin{aligned}\underset{w}{\text{maximize}} \quad & w^T A w \\ \text{subject to} \quad & w^T B w = 1.\end{aligned} \tag{4}$$

After writing this optimization problem in the Lagrange form and taking the derivative with respect to w, we obtain the identical

problem in the form of $Aw = \mu Bw$ which is the Generalized Eigenvalue Decomposition (GED). The solutions of this equation are the joint eigenvectors of A and B and μ is the associated eigenvalue of a particular eigenvector.

The drawbacks of the CSP method that are described earlier lead us to find a way to *sparsify* the spatial filter to increase the classification accuracy and the generalization capability of the method. We assume that the discriminatory information is embedded in a few channels where the number of these channels is much smaller than the actual number of all recording channels. So the discrimination can be obtained with a sparse spatial projection, which uses only informative channels. In this scheme assume that the data was recorded from K channels. We are interested in obtaining a sparse spatial projection using an unconstrained minimization problem in the form of (2), where w has only k nonzero entries, $card(w) = k$ and $k \ll K$. We note that the $R(w)$ does not depend on the magnitude of w, as shown in the following equation. Let $w^* = \alpha w$, then

$$R(w^*) = R(\alpha w) = \frac{\alpha^2 w^T A w}{\alpha^2 w^T B w} = R(w) \tag{5}$$

where α is any scalar which is not equal to zero. Since $R(w)$ does not depend on the gain of w, the optimizer arbitrarily reduces the gain of w to minimize regularization term $\lambda||w||$ after finding the direction that minimizes $R(w)$. Thus, the solution of the optimization problem that uses $R(w)$ as an objective function is essentially the same as the GED solution.

To find a sparse solution we need to have an objective function that depends on the gain of w. In this scheme, we replaced $R(w)$ with the following objective function.

$$G(w) = w^T A w + \frac{1}{w^T B w} \tag{6}$$

This function is bounded from below and has interesting properties. Let us define $a = w^T A w$ and $b = w^T B w$. If we define RQ in terms of a and b such that $R = a/b$ then our new objective function can be expressed as

$$G(w) = a + \frac{1}{b} = \frac{ab}{b} + \frac{1}{b} = Rb + \frac{1}{b} \tag{7}$$

The derivative of $G(w)$ with respect to R is equal to b which is always positive. This indicates that our objective function $G(w)$ decreases with a decrease in R value. After taking the derivative of $G(w)$ with respect to b and solving Eq. 8,

$$\frac{\partial G(w)}{\partial b} = R - \frac{1}{b^2} = 0 \tag{8}$$

we note that b is equal to $\sqrt{R^{-1}}$. By inserting b value into the Eq. 7 we obtain the minimum value of $G(w)$ as $2\sqrt{R}$. This result shows that the direction that minimizes R also minimizes $G(w)$.

We plug $G(w)$ into unconstrained optimization formulation in (2) as the objective function. Rather than working to solve (2) with a non-differentiable ℓ_1 penalty, we replaced it with a twice differentiable smooth version of ℓ_1 (epsL1) which is sufficiently close to minimizing ℓ_1 [9]. The main advantage of this approach is that, since epsL1 and $G(w)$ are both twice differentiable we can directly apply an unconstrained optimization method to minimize $L(w)$ [12]. The epsL1 is defined as

$$||w|| = \sum_{i=1}^{K} \sqrt{w_i^2 + \epsilon} \tag{9}$$

where ϵ is a sufficiently small parameter and K is the dimension of w. The epsL1 approximates the ℓ_1 norm and they are identical when ϵ is equal to zero. Twice differentiability of the epsL1 norm allows us to use it when w_i is equal to zero unlike the regular ℓ_1 norm which is not differentiable at zero.

The solution w that minimizes the function $L(w) = G(w) + \lambda||w||$ tends to become sparse as λ gets bigger. The entries of w generally were not exactly equal to zero, so we normalized w to its maximum absolute value and eliminated the weights consequently corresponding channels that do not exceed a predefined threshold ($=10^{-2}$). We used "fminunc" function of Matlab to find the solution of our unconstrained minimization problem. We computed the desired cardinality which is the number of channels to be selected for the spatial projection by implementing a bisection search [13] on the λ. The upper border of λ was determined initially using the $G(w_c)/||w_c||$ ratio where w_c is the full CSP solution. In case the initial upper border results a cardinality larger than the desired value, we kept doubling the λ parameter until we obtained a λ that results a cardinality which is less than or equal to the target value.

Following the above procedure, we computed the first spatial filter w that minimizes the $G(w)$ which also minimizes the $R(w)$. The solution that maximizes $R(w)$ is also a useful spatial filter. Therefore, we interchanged the matrix A and B to find a solution that maximizes $R(w)$. In order to find multiple sparse filters we deflated the covariance matrices with sparse vectors using the Schur complement deflation method described in [14].

2.2. ECoG & EEG datasets

We applied the SSP method on two different datasets, multiclass ECoG and two class EEG of BCI competitions IV and III respectively.

The ECoG data was recorded from three subjects during finger flexions and extensions [10] with a sampling rate of 1 kHz. The electrode grid was placed on the surface of the brain. Each electrode array contained 48 (8×6) or 64 (8×8) 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 3–5 times during the cue period. The ECoG data of each subject was subband filtered in the gamma frequency band (65–200 Hz) as in [15]. We used 1 s data following the movement onset in the analysis. The dataset contains around 146 trials for each subject.

We also used the BCI competition III dataset IVa [11]. The dataset is recorded from five subjects (aa, al, av, aw, ay) who were asked to imagine either right foot or right index finger movements. The sampling rate of the data was 1 kHz and data was recorded from 118 channels. The EEG signal was filtered in the range of 8–30 Hz. There were 140 trials available for each class. Once again, 1 s data following the cue was used in the analysis.

For both ECoG and EEG datasets, the signal was transformed into four spatial filters by taking first and last two eigenvectors for each CSP methods. After computing the spatial filter outputs, we calculated the energy of the signal and converted it to log scale for each sparse filter and we used them as input features to lib-SVM classifier with an RBF kernel [16]. We also investigated the efficacy of using Linear Discriminant Analysis (LDA) classifier [17] which is parameter free decision function.

Since we are tackling a multiclass problem for the ECoG dataset, we used the pairwise discrimination strategy of [2] to apply the CSP to the five-class finger movement data. In other words, we constructed sparse spatial filters tuned to contrast pairs of finger movements such as 1 vs. 2; 1 vs. 3; 2 vs. 4, etc.

In each dataset, we compared the SSP to the standard CSP and to the ℓ_0 norm based BE method of [8] as it provided superior results in terms of classification accuracy and reduced cardinality. We studied the classification accuracy as a function of cardinality. On the training data with the purpose of finding optimum *sparseness* level for the classification, we computed several sparse solutions, with decreasing cardinality. For the ECoG dataset, the sparse CSP methods were employed with $k \in \{40, 30, 20, 15, 10, 5, 2, 1\}$. For the EEG dataset, we computed the sparse filters with $k \in \{80, 60, 40, 30, 20, 15, 10, 5, 2, 1\}$. For each cardinality, we computed the

Table 1
ECoG dataset classification error rates (%) for each subject using SVM classifier.

	Cardinality	Subject 1	Subject 2	Subject 3	Avg
BE	5	19.8	17.1	16.8	18
SSP	5	18.4	13.4	18	17
CSP	All	30.7	26	32.8	30

corresponding RQ value. We studied the inverse of the RQ (IRQ) curve and determined the optimal cardinality where its value suddenly dropped indicating we started to lose informative channels.

For the ECoG dataset, half of the trials were used in training and the remaining half for testing. In average, we used 15 ± 2 train trials per finger (the thumb, index, middle, ring and little fingers respectively). The EEG dataset contains 140 trials per class and subject. We used 70 trials in training to estimate the sparse filters, and 70 trials for testing. In both datasets, the value of the ϵ in epsL1 regularization term was chosen to be 10^{-6}.

3. Results

We observed that for the SSP method, any particular λ value can lead to different cardinality and normalized IRQ values for different subjects as shown in Fig. 1. In particular, this inter subject variability of IRQ did not allow us to use the same λ value for all subjects (See Fig. 1a and b). However, the variability of IRQ values of different subjects was lower when we fixed the cardinality as shown in Fig. 1e and f. Consequently, due to this reduced variability and to compare our method to the BE technique, we studied the classification error as a function of cardinality. In order to decide on the optimal cardinality level to be used on the test data, the IRQ values were computed on the training data, scaled to their maximum value and averaged over subjects. In the following step, we computed the slope of the IRQ curve and normalized it to its maximum value to get an idea about the relative change in the IRQ.

We depicted the change in IRQ values for each cardinality as shown in Fig. 2a and b. As expected, decreasing the cardinality of the spatial projection resulted to a decrease in the IRQ value. To determine the optimum cardinality to be used in classification on the test data, we selected the cardinality that is below 10% of the maximum relative change (see the dashed lines in Fig. 2a and b). For the ECoG dataset, the cardinality value was found to be 5 and for the EEG data set, it was found to be 15 for the SSP method. For the BE method these values were 5 and 10 respectively. These indices perfectly corresponded to the elbow of the IRQ curve, which indicates loss of informative channels. In Tables 1 and 2, we provide the classification results and selected cardinalities for the ECoG and EEG data set using different methods including SSP, CSP and ℓ_0 based greedy solution, BE. In order to give a flavor about the change in error rate versus the cardinality, we provided the related classification error curves in Fig. 2c and d. For the ECoG data selected cardinality provided minimum test errors. However, for the EEG data, although the minimum classification error was obtained at cardinality 5 for the BE method, we noticed that we identified the optimum cardinality as 10 in the training data.

On all subjects we studied, we observed that the SSP method consistently outperformed the CSP method. We noted that the minimum error rate was obtained with SSP method for ECoG data. Both

Table 2
EEG dataset classification error rates (%) for each subject using SVM classifier.

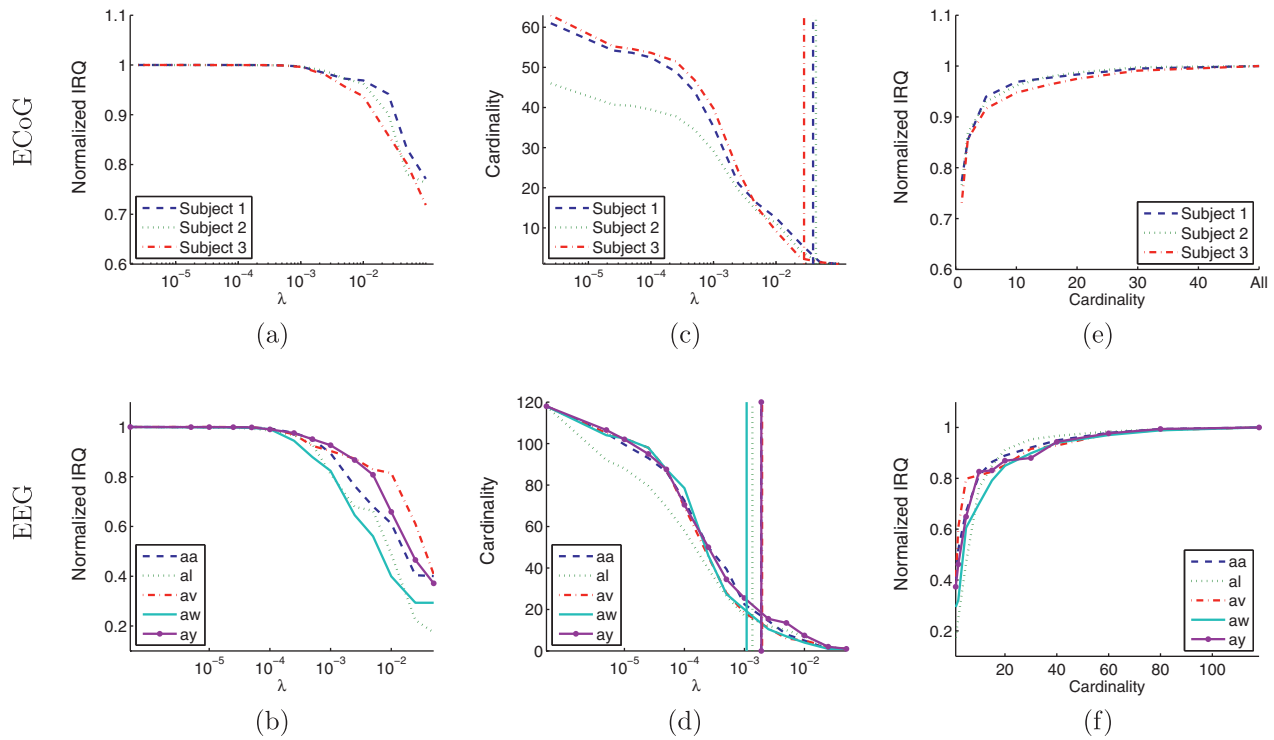|	Cardinality	aa	al	av	aw	ay	Avg
BE	10	13.6	2.9	30.7	2.1	5.0	10.9
SSP	15	19.3	1.4	23.6	4.3	5.7	10.9
CSP	118	23.6	3.6	32.1	2.9	11.4	14.7

Fig. 1. Normalized IRQ values are shown in (a) for ECoG and in (b) for EEG data. The cardinality vs λ value of the minimization function $L(\omega) = G(\omega) + \lambda \| \omega \|$ (c) ECoG and (d) EEG data for each subject. The vertical lines indicate the λ values that are initially chosen for bisection search. The initial data point corresponds to $\lambda = 0$ which produces the regular CSP solution. The normalized IRQ values vs. cardinality for each subject is shown in (e) and (f).

SSP and BE methods used cardinality of 5 to achieve the minimum error rate. As expected the full CSP solution did not perform as good as the other sparse methods and likely overfitted the training data. The SSP method improved the classification error rate with an error difference of 13.2%. We obtained comparable results on EEG data using the SSP and BE methods (p-value = 0.5, paired t-test), however BE provided a significant improvement over SSP method on the number of channels (p-value = 0.003) used in spatial projection. The error difference between regular CSP and SSP is less apparent in the EEG dataset where the difference between classification accuracies was 3.8%. This could be due to the high number of training trials used in EEG data. We studied the effect of the amount of training data on the classification accuracy and presented the results in Fig. 4a. When a small number of training trials, as low as 15 are used in the EEG dataset, the difference between the sparse and standard CSP technique was more than 6%. Interestingly, with increasing number of training trials the SSP method consistently provided better results and the difference remained between 3 and 4%. There was no noticeable difference between SSP and BE.

The classification results obtained with LDA classifier are given in Table 3. We observed that the LDA classifier which does not involve parameter selection like SVM, provided slightly higher error rates for the sparse solutions. This could be due to nonlinear decision surface and maximum margin identified by the SVM classifier. Interestingly, in both datasets, the LDA classifier resulted in lower error rates with the non-sparse CSP solution.

Table 3
Average test error rate (%) and corresponding cardinality.

	ECoG			EEG		
	BE	SSP	CSP	BE	SSP	CSP
LDA	19.9	18.9	24.1	11.4	11.1	14.3
SVM	17.9	16.6	29.8	10.9	10.9	14.7
Cardinality	5	5	All	10	15	All

Fig. 3 illustrates the distribution of the spatial filters obtained using SSP and CSP algorithms for all subjects. We observed that the SSP filter coefficients are localized on the left hemisphere and the central area, which is in accordance with the cortical regions related to right hand and the foot movement generation.

Arvaneh et al. [7] used the ℓ_1/ℓ_2 ratio as a penalty term and they applied their algorithm to the BCI competition III EEG dataset IVa [11] which we used in this paper as well. They achieved a mean error rate of $17.7 \pm 15.4\%$ using 22.6 ± 11 channels. Here, we compared our method with the study of Arvaneh et al. by extracting one filter from each end of the sparse solutions. The SSP method achieved a mean error rate of $12 \pm 11.3\%$ with an average number of channels 25.6 ± 2.3. The obtained results indicated that the SSP method provided a significant improvement (p-value = 0.024, paired t-test) over the ℓ_1/ℓ_2 based algorithm on the classification accuracy without any significant difference between number of channels used (p-value = 0.28).

In order to compare the computational complexity of SSP method to the BE, we computed sparse filters with a cardinality of two from an increasing number of recording channels on simulated data. The training was performed on a regular desktop computer with 4 GB of RAM and equipped with a CPU running at 2.66 GHz. The elapsed time per filter computation increased exponentially for the BE method and linearly for the SSP method as shown in Fig. 4b. With 128 channels, the BE algorithm computed a single spatial filter with two nonzero entries in 90 s. For the SSP method with the same setup above, the elapsed time was less than a second. Although, we used the relative change in the IRQ to identify the optimum sparsity level, one can also run a typical k-fold cross validation procedure to identify the optimum level. However, in such a case training the system with BE method will take several hours which may not be feasible for BMI applications. On the other hand with the SSP method training through cross validation can be executed in a few minutes.
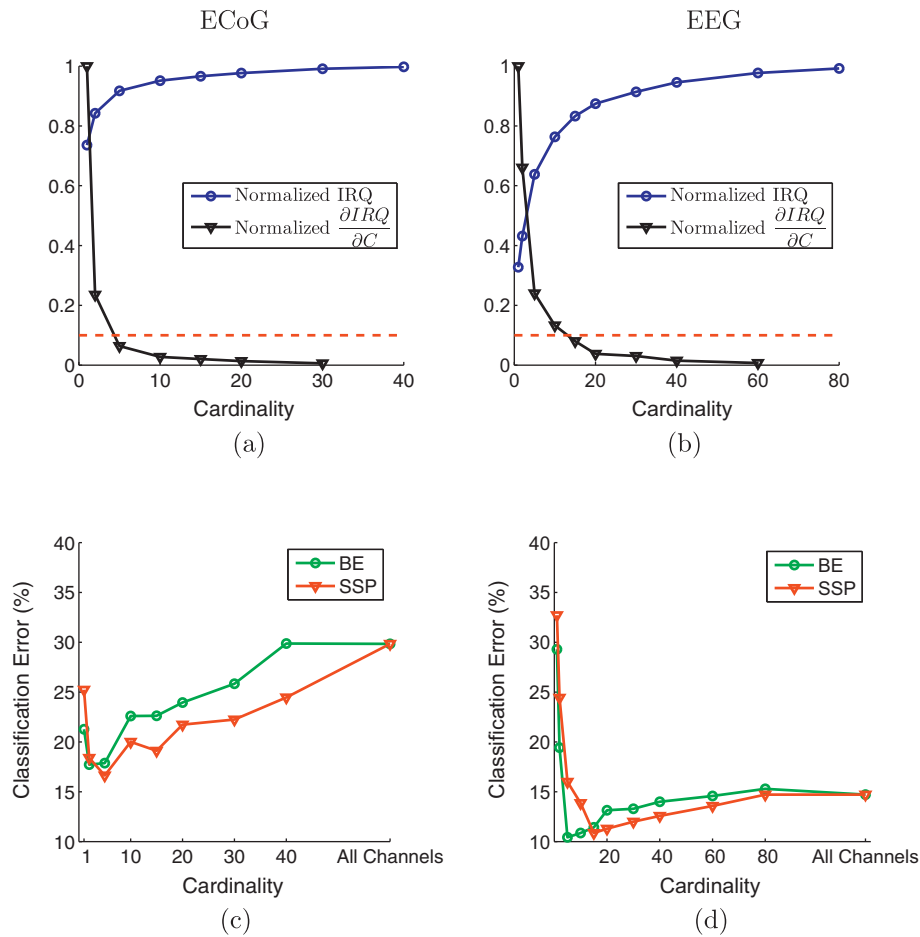
ECoG

EEG

(a)

(b)

(c)

(d)

Fig. 2. The average IRQ of all subjects versus cardinality (a) ECoG and (b) EEG data. The red line is the 10 percent threshold that determines the optimum cardinality to be used in the test data. The optimum cardinality levels for ECoG and EEG are five and 15 respectively. The classification error curves of SSP and BE methods versus the cardinality are given in (c) for ECoG and in (d) for EEG The last data point corresponds to the results obtained from standard CSP which uses all channels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

In order to evaluate the effect of noise and intersession variability on the performance of our approach we studied two different controlled experiments:

i Adding Gaussian noise to a randomly selected channel in ECoG and EEG data.
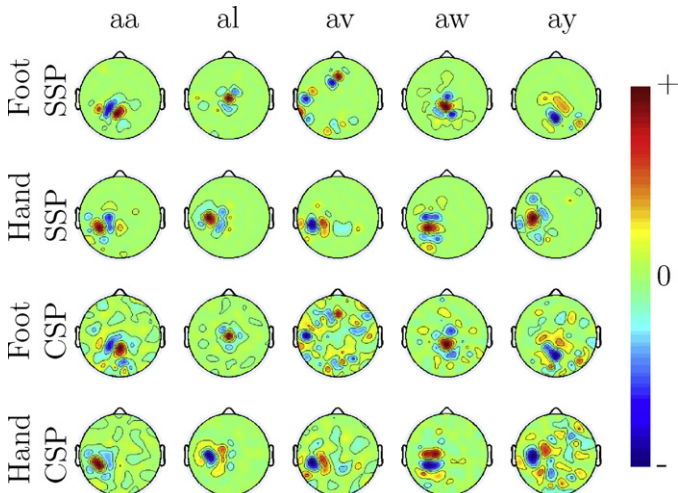

Fig. 3. The CSP and SSP filters for hand and foot movement imagination.

ii Simulation of electrode displacement in EEG data.

During the first experiment, we added zero mean Gaussian noise to one of the channels and calculated the resulting classification error to obtain final classification accuracy for the noisy data. This experiment was repeated for all channels and then the resulting average classification error was computed. The variance of the additional noise was increased in a controlled manner and proportional to the average variance of all channels. The ratio of the noise to the variance of the original data was named as noise to signal plus noise ratio (NSNR) since the original signal already contaminated from different noise sources. In Fig. 5, we depicted the classification accuracy vs. NSNR which was expressed in Decibel scale. It was noted that an increase in NSNR caused error to increase for all methods, however the increase in CSP method was more than the sparse filters. While using the sparse methods in the ECoG data, the classification error reached a plateau after 5 dB whereas the standard CSP error increased monotonically. A similar behavior was observed in the EEG data.

In the second experiment, in order to study intersession variability we simulated electrode displacements by interpolating the EEG test data at different positions. Since electrode locations for the ECoG data were not available, we conducted this experiment with the EEG data only. We randomly determined the direction and the amount of displacement of the new electrode locations. To be more realistic, we introduced displacement which was uniformly distributed over 118 electrodes varying between 0 and 50% of the
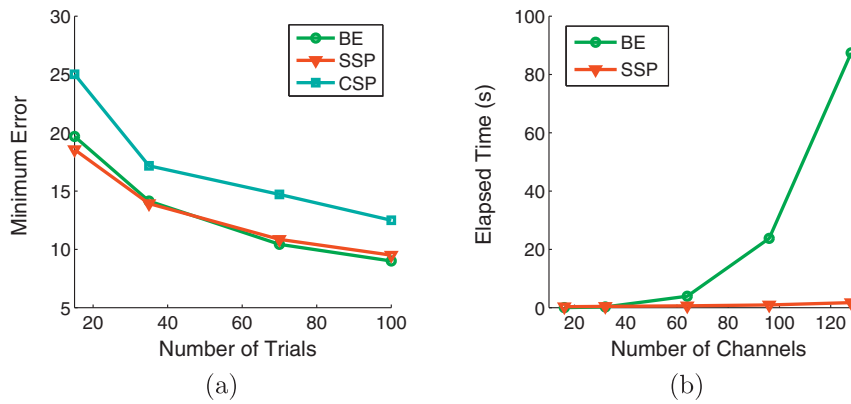
Fig. 4. (a) The minimum error vs. the number of trials. (b) The average elapsed time to estimate a spatial filter with a cardinality of two vs. the number of total recording channels.
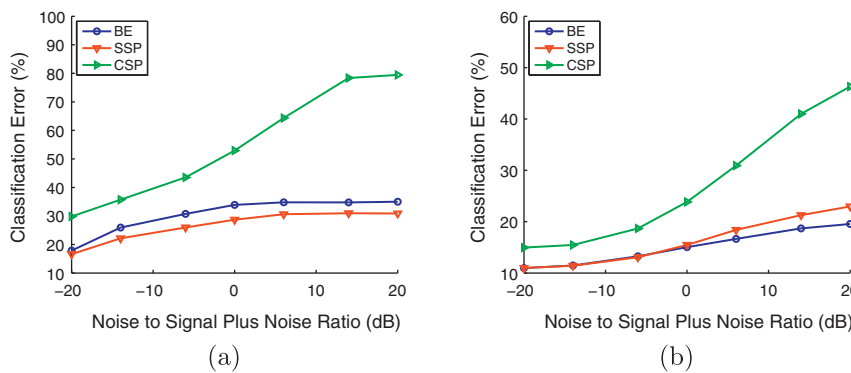

Fig. 5. The noise to signal plus noise ratio (NSNR) vs. classification accuracy for (a) ECoG and (b) EEG sets.

distance to the nearest electrode as shown in Fig. 6a. The testing of algorithms was conducted on the interpolated EEG data on the displaced electrode locations. The classification results vs. cardinality is shown in Fig. 6b. It was noted that the error change for CSP, BE and SSP methods are 13%, 12% and 10%, respectively. The increase in error rate was not as severe as in the previous noise contaminated channel experiment. The sparse methods had slightly lower error increase than the standard CSP. Initially, it was expected that the sparse methods would not be this vulnerable to displacement in the electrode locations. At this point, we speculate that the smoothness of the projections have a certain advantage on the

displaced electrodes. For instance, due to sparseness, it is likely that the shifted electrode locations are weighted with zero. This makes the sparse projections discard the shifted channels that fall outside the projection zone. In contrary, the standard CSP is associated with some sort of smoothness and many crucial channels, although displaced, could still be projected with similar weights due to the correlated distribution. Consequently, it should be noted that not only sparseness but also smoothness of projections could be an important parameter contributing to the generalization capability of the methods. Never the less, once again we need to highlight that, the increase in error rate was lower for the sparse methods.
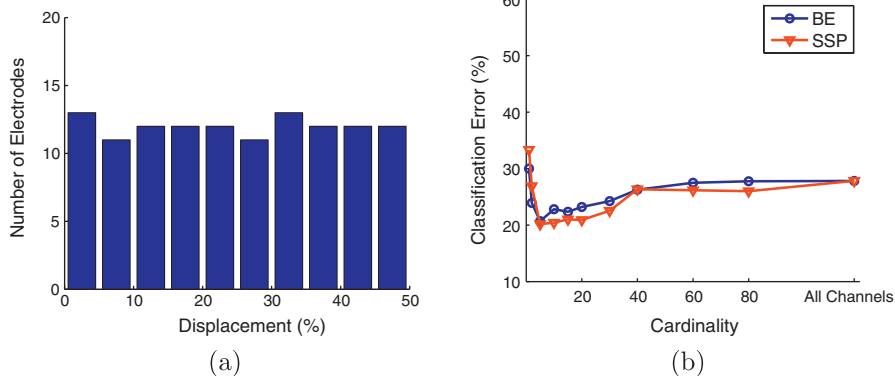

Fig. 6. (a) The histogram of the number of electrodes with respect to the displacement induced on the test data. (b) The cardinality vs. classification error on the test data with displaced electrode locations.

4. Conclusion

The need for the sparse filters is apparent when there is large number of recording electrodes and insufficient amount of training data. To minimize overfitting on the training data and eliminate noisy channels, we introduced a spatially sparse projection technique (SSP) based on a novel objective function. Unlike the RQ, this new objective function has a dependency on the filter magnitude. By using an approximated ℓ_1 norm, we computed the sparse spatial filters through an unconstrained minimization formulation with standard optimization algorithm. We applied our method to ECoG and EEG datasets and compared its efficiency to standard CSP, and to a ℓ_0 norm based greedy technique. The SSP method outperformed the standard CSP on both datasets and provided comparable results to ℓ_0 norm based method, which is associated with higher computational complexity. On the ECoG data, the SSP method provided 44% decrease in the error rate compared to standard CSP method and used only five channels in each spatial projection. The error difference between regular CSP and SSP is less apparent in the EEG dataset as SSP method provided 26% decrease in the error rate. In contrary to the ECoG data, we also observed that more channels were used to achieve minimum classification accuracy in the EEG dataset. This could be due the low spatial resolution originating from the volume conduction and low SNR of the EEG. Nevertheless, the SSP algorithm was able to reach a minimum error rate with only 15 channels. Our results indicate that SSP method can be effectively used to extract features from both EEG and ECoG datasets with large number of recording channels. We note that the sparse methods provided superior results compared to the standard CSP when there is a noisy/corrupted channel in the test data. In another setup where displaced channels were used to simulate intersession variability, we note that the sparse methods had slightly better robustness than the standard CSP. These observations indicate that the sparse spatial projection framework can be effectively used as a robust feature extraction engine of future BCI systems.

Acknowledgments

References

[1] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K.-R. Muller, Optimizing spatial filters for robust EEG Single-trial analysis, signal, Processing Magazine, IEEE 25 (2008) 41–56.

[2] N.F. Ince, R. Gupta, S. Arica, A.H. Tewfik, J. Ashe, G. Pellizzer, High accuracy decoding of movement target direction in non-human primates based on common spatial patterns of local field potentials, PLoS One 5 (2010) e14384.

[3] J. Farquhar, N.J. Hill, T.N. Lal, B. Schlkopf, Regularised CSP for sensor selection in BCI, in: 3rd International Brain-Computer Interface Workshop and Training Course, 14-15 Aug 2006, Graz, Austria, 2006, http://eprints.pascal-network.org/archive/00002709/

[4] B. Reuderink, M. Poel, Robustness of the Common Spatial Patterns algorithm in the BCI-pipeline, 2008.

[5] H. Ramoser, J. Muller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, IEEE Transactions on Rehabilitation Engineering 8 (2000) 441–446.

[6] X. Yong, R. Ward, G. Birch, Sparse spatial filter optimization for EEG channel reduction in brain–computer interface, in: Acoustics, Speech and Signal Processing, IEEE International Conference on ICASSP 2008, 2008, pp. 417–420.

[7] M. Arvaneh, C. Guan, K.K. Ang, C. Quek, Optimizing the channel selection and classification accuracy in EEG-based BCI, IEEE Transactions on Biomedical Engineering 58 (2011) 1865–1873.

[8] F. Goksu, N. Ince, A. Tewfik, Sparse common spatial patterns in brain computer interface applications, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011, pp. 533–536.

[9] S.-i. Lee, H. Lee, P. Abbeel, A.Y. Ng, Efficient L1 Regularized Logistic Regression, in: In AAAI.

[10] K. J. Miller, G. Schalk, Prediction of Finger Flexion 4th Brain–Computer Interface Data Competition, 2008.

[11] G. Dornhege, B. Blankertz, G. Curio, K.-R. Müller, Bci Competion III, Dataset IVa, 2005.

[12] M. Schmidt, G. Fung, R. Rosales, Fast Optimization Methods for L1 regularization: A Comparative Study and Two New Approaches, 2009.

[13] R. Burden, J. Faires, Numerical Analysis, 8th ed., Thomson Brooks/Cole, 2005.

[14] L. Mackey, Deflation Methods for Sparse pca, in: D. Koller, D. Schuurmans, Y. Bengio, L. Bottou (Eds.), in: Advances in Neural Information Processing Systems, vol. 21, 2009, pp. 1017–1024.

[15] I. Onaran, N.F. Ince, A. E. Cetin, Classification of Multichannel ECoG Related to Individual F inger Movements with Redundant Spatial Projections, in: International IEEE EMBS Conference, 2011, http://www.ieeexplore.info/search/searchresult.jsp?newsearch=true&queryText=Classification+of+Multichannel+ECoG+Related+to+Individual&x=34&y=18

[16] C.-C. Chang, C.-J. Lin, A library for Support Vector Machines, 2001.

[17] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2nd ed., John Wiley & Sons, Inc., New York, N.Y., 2000.