35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Regularization Using Similarities of Signals Observed in Nearby Sensors for Feature Extraction of Brain Signals

Hiroshi Higashi and Toshihisa Tanaka

***Abstract*— In order to solve uncertainty of spatial weights learned with small amount of training samples for feature extraction from brain signals, a regularization using similarity of signals observed in sensors that are located near each other is proposed. Deriving the regularization is begun defining a distance between the sensors. Under the distance, the proposed regularization works so that the spatial weights extracts similar signals in the nearby sensors. The proposed regularization is applied to the well known common spatial pattern (CSP) method that finds spatial weights for EEG based brain machine interface. In the classification experiment using a dataset of EEG signals during motor imagery, the proposed method achieved maximum improvement by 28% in the classification accuracy over the standard CSP in a setting of even when only five samples are used.**

I. INTRODUCTION

In the fields of brain signal measurement systems such as multichannel electroencephalogram (EEG) system, wireless communication techniques such as multiple input multiple output (MIMO), and so on, sensor arrays are widely used [1]. Improving the signal to noise ratio (SNR) of observed signals and separating some source signals from observations can achieved by weighted-sum of the multiple signals observed in the sensor arrays [1], [2].

In EEG signal processing with the sensor array, the weights for each sensor are often called spatial weights, because each sensor is located in different positions. The spatial weighting is given;

$$y(t) = \sum_{i=1}^{M} w_i x_i(t), \tag{1}$$

where $x_i(t)$ denotes a signal observed in the ith sensor at time t, M denotes the number of the sensors, w_i denotes a spatial weight for the ith sensor, and $y(t)$ denotes an extracted signal. The problem to extract certain components from observations of the sensor array is to find the weight vector denoted by $\boldsymbol{w} = [w_1, \ldots, w_M]^T$ under a certain criterion. For this purpose, learning approaches using observed signals are widely adopted [3], [4]. A number of signal processing techniques such as Wiener filter, principal component analysis (PCA), independent component analysis (ICA) [2], and so forth are involved in this problem.

H. Higashi and T. Tanaka are with the Department of Electronic and Information Engineering, Tokyo University of Agriculture and Technology, Tokyo, Japan, and the Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Saitama, Japan. higashi@sip.tuat.ac.jp, tanakat@cc.tuat.ac.jp

This work is supported in part by KAKENHI, Grant-in-Aid for Scientific Research (B), 21360179 and Grant-in-Aid for JSPS Fellows, 24686.

However, the learning procedure can be ill-posed because of limited numbers of sensors and samples. Hence regularization is widely used to prevent overfitting or to solve an ill-posed problem in signal processing and machine learning for learning parameters [5], [6]. The regularization for an optimization problem is to add to an original cost function a penalty term which represents additional information such as smoothness or bounds of the vector norm of parameters to be optimized. In this way, the regularization can help design more robust spatial weights against ill-posed problems [7].

In some situations, the signals measured by the sensors that are located near each other (the nearby sensors) are similar and also the observed components are similar. To describe the situations, consider a measurement device of EEG where electrodes installed on scalp observe faint electrical difference. The EEG reflects the summation of the synchronous activity of thousands or millions of neurons [8], [9]. Therefore, the nearby sensors likely observe activities which are induced from the same neurons. For the reason, the spatial filters such as the Laplacian filter that averages the signals observed in the nearby sensors are often used for improving SNR in EEG signal processing [10]. The regularizations motivated by the idea of averaging signals observed in the nearby sensors have been proposed in [7], [11]. However, the regularizations will not work appropriate if the amplitudes between sensors are different due to the measurement environments, because the regularizations evaluate the similarities between the weight coefficients. To solve this problem, we propose a regularization that works such that the signal that is observed in the ith channel and weighted by w_i becomes as similar as possible to the weighted signals observed in the sensors that are located near the ith sensor.

We have applied the proposed regularization to the common spatial pattern (CSP) method [3], [4], which is a widely used technique to find effectively spatial weights efficiently that extract the brain activity for an EEG based brain machine/computer interface (BMI/BCI) [9]. The regularized CSP can be solved with a generalized eigenvalue problem, since the regularization term can be formulated in a quadratic form. CSP with the proposed regularization has been demonstrated for artificial signals to show nearby electrodes have similar weight coefficients. The classification experiment of motor imagery based BMI (MI-BMI) dataset has been conducted with comparing an existing regularized CSP [11] and the proposed method demonstrated improvement of classification accuracy in a setting of the small number of samples.

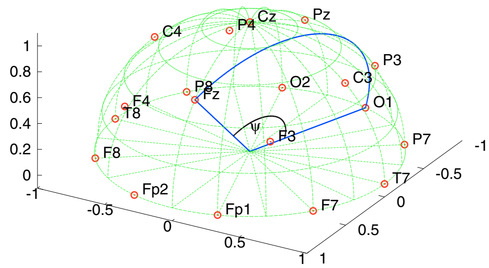


Fig. 1. The electrode arrangement of the international 10-20 method on the orthogonal coordinates. The red circles represent the electrodes.

II. Regularization based on signal similarities in nearby sensors

This section addresses the proposed regularization of using signal similarities in nearby sensors on head surface. A regularization of using signal similarities in nearby sensors a sensor array on head surface is proposed in this section. First, a distance between electrodes of EEG measurement system is defined in Sec. II-A. Then we introduce the regularization derived with the defined distance in Sec. II-B.

A. Distance between electrodes

We define a distance between electrodes on the arrangements used for EEG measurement. International 10-20, 10-10, and 10-5 methods [8], [12], [13] have stood as the defacto standard of electrode arrangement. In these systems, locations on a head surface are described by relative distances between cranial landmarks over the head surface. Under an assumption that the shape of head is a sphere, the locations on head surface can be describe coordinates represented by $\xi = \{x, y, z\}$. We define the coordinates such as the axes of Fig. 1 that illustrates the electrode positions of the international 10-20 method.

Given the positions of two electrodes as $\xi_i = \{x_i, y_i, z_i\}$ and $\xi_j = \{x_j, y_j, z_j\}$. The question arising here is: how to define the distance between two points, on the head. The Euclidean distance defined by $d_{i,j} = \|\xi_i - \xi_j\|$ is a straightforward solution. In this paper, we define the perimeter of a sector the two sides of which are line segments between the origin and two electrode position on the coordinates as the distance between two electrodes. Let ϕ_{ij} be the angle between the line segments between the origin and ξ_i, and the origin and ξ_j. The distance by the perimeter is defined as $d_{i,j} = \nu\phi_{ij}$, where $\nu = \|\epsilon_i\| = \|\epsilon_j\|$. Moreover, because $\cos\phi = \frac{<\xi_i, \xi_j>}{\nu^2}$ and $\nu = 1$, $d_{ij} = \arccos(x_i x_j + y_i y_j + z_i z_j)$. The metric is illustrated in Fig. 1. In the figure, we show the distance between F_z and O_1 as an example. The length of the curve connecting F_z and O_1 is the defined distance by the metric.

B. Regularization

Consider a sensor array consisting of M sensors. A signal sample observed in the ith channel at a time instance is denoted by x_i. A set, $\{x_i\}_{i=1}^{M}$, forms a vector $\boldsymbol{x}$ defined as $\boldsymbol{x} = [x_1, \ldots, x_M]^T$. We obtain d_{ij} for $i, j = 1, \ldots, M$ as the distances between sensors by the metric defined in Sec. II-A. To mainly evaluate the regularization costs between each sensor and its nearby sensors, the Gaussian metric between two points;

$$g_{ij} = \exp\left(-\frac{d_{ij}^2}{2p^2}\right), \tag{2}$$

is employed, where p denotes a parameter to tune the closeness of the two sensors. Then we define the cost;

$$P(\boldsymbol{w}) = E_{\boldsymbol{x}}\left[\sum_{i=1}^{M-1}\sum_{j=i+1}^{M} g_{ij}|w_i x_i - w_j x_j|^2\right], \tag{3}$$

which evaluates the mean of squared error between weighted signals observed in sensors that are located near each other, Note that the cost (3) becomes small as the weighted signals become similar in the nearby sensors.

Equation (3) can be transformed to matrix vector form as

$$P(\boldsymbol{w}) = E_{\boldsymbol{x}}[\boldsymbol{w}^T\boldsymbol{D}_x(\boldsymbol{C}-\boldsymbol{G})\boldsymbol{D}_x\boldsymbol{w}] = \boldsymbol{w}^T\boldsymbol{Q}\boldsymbol{w}, \tag{4}$$

where $\boldsymbol{C}$ and $\boldsymbol{D}_x$ are diagonal matrices defined as $[\boldsymbol{C}]_{ii} = \sum_{k=1}^{M} g_{ik}$, $[\boldsymbol{D}_x]_{ii} = x_i$, $i = 1, \ldots, M$, each element of $\boldsymbol{G} \in \mathbb{R}^{M\times M}$ is defined as $[\boldsymbol{G}]_{ij} = g_{ij}$, $i, j = 1, \ldots, M$, and $\boldsymbol{Q} = E_{\boldsymbol{x}}[\boldsymbol{D}_x(\boldsymbol{C}-\boldsymbol{G})\boldsymbol{D}_x]$. To take expectation over $\boldsymbol{x}$ for obtaining $\boldsymbol{Q}$, we can use the sample average of observed signals.

III. CSP with the signal similarities based regularization

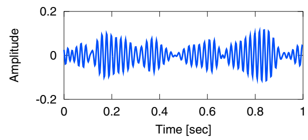
The CSP method [3], [4] is effective in feature extraction and classification for two-class MI-BMI. In this section, we first review the standard CSP method. Then, we exhibit how to apply the regularization in described Sec. II for finding CSP.

A. Common spatial pattern (CSP) [3], [4]

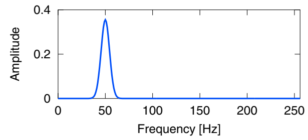
CSP is a set of spatial weights extracting a signal from multichannel signals [3], [4]. The problem using labeled training samples to design the spatial weights can be formulated as follows. Let $\boldsymbol{X} \in \mathbb{R}^{M\times N}$ be a matrix representing the observed signals, where M is the number of channels and N is the number of time instances. Denote the components (vectors) of $\boldsymbol{X}$ by $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]$, where $\boldsymbol{x}_n \in \mathbb{R}^M$ and n is the time index. CSP, $\boldsymbol{w} \in \mathbb{R}^M$, is found in such a way that the variance of a signal extracted by linear combination of $\boldsymbol{X}$ and $\boldsymbol{w}$ is minimized in a class [4]. The time variance of the extracted signal of $\boldsymbol{X}$ is given by

$$\sigma^2(\boldsymbol{X}, \boldsymbol{w}) = \frac{1}{N}\sum_{n=1}^{N}|\boldsymbol{w}^T(\boldsymbol{x}_n - \boldsymbol{\mu})|^2, \tag{5}$$

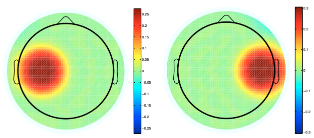
where $\boldsymbol{\mu} = N^{-1}\sum_{n=1}^{N}\boldsymbol{x}_n$. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be sets of the training data. The set, $\mathcal{C}_d$, contains the signals belonging to class d, d represents a class label chosen in $\{1, 2\}$, and $\mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$. We choose c as a class label and CSP is given as the generalized eigenvector corresponding to the smallest



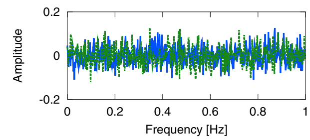
(a) Source signal.



(b) Spectrum of source signal.



(c) Spatial distributions of source signal.



(d) Examples of the observed signals.

Fig. 2. The artificial signals.

generalized eigenvalue of the generalized eigenvalue problem described as

$$\mathbf{\Sigma}_c \boldsymbol{w} = \lambda(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2)\boldsymbol{w}, \tag{6}$$

where λ is the generalized eigenvalue, $\mathbf{\Sigma}_d$ are defined as

$$\mathbf{\Sigma}_d = E_{\boldsymbol{X} \in \mathcal{C}_d}\left[\frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{x}_n - \boldsymbol{\mu})(\boldsymbol{x}_n - \boldsymbol{\mu})^T\right] \tag{7}$$

for $d = 1, 2$, and $E_{\boldsymbol{X} \in \mathcal{C}_d}[\cdot]$ denotes the expectation over $\mathcal{C}_d$.

B. Regularized CSP

By adding the regularization term given as (3), the modified regularized optimization problem is defined as

$$\min_{\boldsymbol{w}} \ \boldsymbol{w}^T(\mathbf{\Sigma}_c + \gamma \boldsymbol{Q})\boldsymbol{w}, \ \text{subject to } \boldsymbol{w}^T(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2)\boldsymbol{w} = 1, \tag{8}$$

where γ is a combination coefficient. If the matrices of $\mathbf{\Sigma}_c + \gamma \boldsymbol{Q}$ and $\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2$ are nonsingular, (8) is equivalent to the generalized eigenvalue problem:

$$(\mathbf{\Sigma}_c + \gamma \boldsymbol{Q})\boldsymbol{w} = \lambda(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2)\boldsymbol{w}. \tag{9}$$

IV. Experiments

We illustrate the ability of the proposed regularized CSP method for extracting a local feature. The proposed method is demonstrated with artificial signals in Sec. IV-A. Furthermore, the result of classification of real-world EEG signals by spatially weighting of the proposed method is shown in Sec. IV-B.

A. Artificial signals

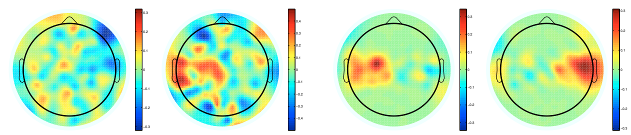
An analysis of the proposed method by a toy experiment with artificial signals is given. We used the mixture of synthetic source and noise signals. We assume to know the spatial distributions of the source signals. The spatial weights derived by the CSP method and the regularized CSP method were compared with the true distribution.

We assumed a 2-class BMI where observed EEG signals are modeled by a mixture of narrow-band signals. In this model, two signals, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, belonging to class 1 and class 2, respectively, are given by $\boldsymbol{x}_1[n] = \boldsymbol{a}_1[n]s[n] + \boldsymbol{\eta}$, and $\boldsymbol{x}_2[n] = \boldsymbol{a}_2[n]s[n] + \boldsymbol{\eta}$, for $n = 1, \ldots, N$, where $\boldsymbol{x}_1[n], \boldsymbol{x}_2[n] \in \mathbb{R}^M$ denote vectors representing a signal observed at discrete time instance n, N denotes the number of time instances, M denotes the number of channels, $s[n] \in \mathbb{R}$ denotes a source signal of feature component, $\boldsymbol{a}_1, \boldsymbol{a}_2 \in \mathbb{R}^M$ denote vectors defined by $\boldsymbol{a}_i = [a_{i1}, \ldots, a_{iM}]^T$, $a_{im} \in \mathbb{R}$ denote an amplitude of the source at the mth channel for class i, and $\boldsymbol{\eta} \in \mathbb{R}^M$ denote a stochastic noise.

TABLE I

The Settings for Generating Artificial Signals

Parameter	Value and distribution
Number of channels	118
Electrodes arrangement	Int'l extended 10-20
Number of samples	512
Sampling frequency	512
Spectrum of the source signals	Fig. 2(b)
Distributions	Fig. 2(c)
Noise $[\boldsymbol{\eta}]_m$	$\mathcal{N}(0, 0.1)$



(a) The standard CSP (b) CSP with the proposed method

Fig. 3. Topographical maps of the spatial weights that are as the eigenvectors corresponding to the largest eigenvalues of (6) and (9) ($c = 1$ (left) and $c = 2$ (right)) in the experiment using the artificial signals.

The simulation settings for generating artificial signals were shown in Table I. The observed signal in 2 channels are shown in Fig. 2(d).

The topographically plotted spatial weights given by by the standard CSP and the proposed methods are shown in Fig. 3. The parameters for the proposed method are 0.05 for p and 10^8 for γ. Compared to the standard CSP, the weight designed by the proposed method resulted in the large weight coefficients concentrated at the certain spots. Moreover we can observe in Fig. 3(b) that the topographical maps of the spatial weights given by the proposed method are similar to the true distribution maps shown in Fig 2(c).

B. Real-world EEG signals

We compared performance in a two-class classification of EEG signals during motor imagery using the proposed method to those using the standard CSP and the spatially regularized CSP (SRCSP) [11], respectively.

1) Data description: We used dataset IVa from BCI competition III (for details of the dataset, see http://www.bbci.de/competition/iii/). This dataset consists of EEG signals during right hand and right foot motor-imageries. The EEG signals were recorded from five subjects labeled *aa*, *al*, *av*, *aw*, and *ay*. The measured signal was bandpass filtered with the passband of 0.05–200 Hz, and then digitized at 1000 Hz.

Moreover, the lowpass filter whose cutoff frequency is 50 Hz was applied to recorded signals and the filtered signals was donwsampled to 100 Hz. Furthermore, the signals were bandpass filtered with the passband of 7–30 Hz that is a

TABLE II

ACCURACY [%] GIVEN BY 100 TRAINING SAMPLES PER A CLASS.

	Subject					
	aa	*al*	*av*	*aw*	*ay*	Ave.
CSP	81.4	94.8	53.1	92.9	89.6	82.3
SRCSP	82.2	95.2	64.2	94.3	92.7	85.7
Proposed	82.0	95.4	66.2	94.6	93.0	86.3

band including mu and beta rhythms. The dataset for each subject consisted of signals of 140 trials per a class. The signal length for each trial is 3.5 seconds.

C. Features for classification

The following feature vector was used for classification. In each case of $c = 1$ and $c = 2$, we solve (6) or (9), and then we got the eigenvectors corresponding to the largest eigenvalues in each eigenvalue problem defined by $\hat{\boldsymbol{w}}_1$ and $\hat{\boldsymbol{w}}_2$, respectively. By using the weight vectors, the feature vector was defined as $\boldsymbol{y} = [\sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_1), \sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_2)]^T$.

D. Results and discussions

Linear discriminant analysis [6] were used for classifying the extracted feature vectors. For the proposed methods, we used the signals that were observed in the intervals between the tasks to form the regularization term (3) for each subject.

The classification accuracy was given by training for the spatial weights and the classifier with randomly chosen 100 samples, and testing with the remaining samples. An average accuracy over 100 times of this procedure is shown in in Table II. The parameter in (2) were set to $p = 0.05$ for SRCSP and the proposed method, The parameter was chosen out of $\gamma \in \{10^0, 10^{0.1}, \ldots, 10^{30}\}$. The best accuracy among the parameters for each subject is shown in Table II. In the result of Table II, for *aa*, *al*, *av*, *aw*, and *ay*, γ were set to $10^{9.4}$, 10^{12}, $10^{11.1}$, 10^{10}, and $10^{12.5}$, respectively, in SRCSP. In the proposed method, γ were set to $10^{9.9}$, $10^{14.3}$, $10^{12.7}$, $10^{11.1}$, and $10^{14.4}$, respectively. The both of the regularized CSP slightly outperform the standard CSP method in the classification accuracy for all subjects.

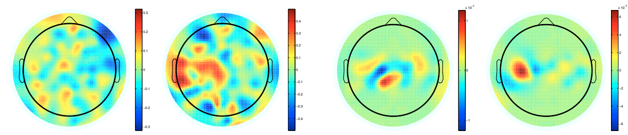
Table III also shows classification accuracy, however when the number of the training samples is considerably reduced to only five samples. As the same as in Table II, the parameters performing the best classification accuracy were chosen out of the candidates. For *aa*, *al*, *av*, *aw*, and *ay*, γ were set to $10^{10.4}$, $10^{10.5}$, $10^{12.5}$, $10^{10.8}$, and $10^{11.4}$, respectively, in SRCSP. In the proposed method, γ were set to $10^{13.6}$, $10^{13.5}$, $10^{16.9}$, $10^{13.9}$, and $10^{15.0}$, respectively. We can observe significant improvement of the accuracy rates for subjects *al* and *ay* by the regularizations. The results suggest that the proposed regularization can improve the accuracy even if the number of training samples available is small.

The topographically plotted spatial weights for subject *ay* is shown in Fig. 4. All samples in the dataset were used to find the spatial weights. The parameters of the proposed method, p and γ, were set to 0.05 and 10^{15}, respectively. Comparing to the standard CSP, the electrodes which have large coefficients do not be scattered spatially in the proposed method.

TABLE III

ACCURACY [%] GIVEN BY 5 TRAINING SAMPLES PER A CLASS.

	Subject					
	aa	*al*	*av*	*aw*	*ay*	Ave.
CSP	52.6	67.0	50.3	61.2	51.0	56.4
SRCSP	58.4	77.8	54.2	71.6	78.6	68.1
Proposed	59.2	81.3	54.4	71.3	79.3	69.1



(a) The standard CSP (b) CSP with the regularization

Fig. 4. Topographical maps of the spatial weights, $\hat{\boldsymbol{w}}_1^{(1)}$ and $\hat{\boldsymbol{w}}_1^{(2)}$, for subject *ay* in the experiment using the real-world EEG signals.

V. CONCLUSION

We have proposed the regularization based on the similarity of observed signals in the nearby sensors for feature extraction problem in an EEG sensor array. Moreover, we have illustrated how to apply the proposed regularization to the CSP method. The experimental results demonstrated that the proposed regularization improves classification accuracy in a setting of the small number of samples.

REFERENCES

[1] H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[2] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning and Applications*. Wiley-Blackwell, 2002.

[3] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clinical Neurophysiology*, vol. 110, no. 5, pp. 787–798, 1999.

[4] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. on Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, 2000.

[5] A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill-posed problems*. V. H. Winston & Sons, 1977.

[6] C. M Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[7] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. on Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, 2011.

[8] J. Malmivuo and R. Plonsey, *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, 1995.

[9] J. R Wolpaw, N. Birbaumer, D. J McFarland, G. Pfurtscheller, and T. M Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[10] S. Sanei and J. Chambers, *EEG Signal Processing*. Wiley-Interscience, 2007.

[11] H. Higashi, A. Cichocki, and T. Tanaka, "Regularization using geometric information between sensors capturing features from brain signals," in *Proc. of 2012 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 721–724.

[12] M. R Nuwer, "Recording electrode site nomenclature," *Journal of Clinical Neurophysiology*, vol. 4, no. 2, pp. 121, 1987.

[13] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements," *Clinical Neurophysiology*, vol. 112, no. 4, pp. 713–719, 2001.