REGULARIZATION USING GEOMETRIC INFORMATION BETWEEN SENSORS CAPTURING FEATURES FROM BRAIN SIGNALS

Hiroshi Higashi^{†,‡}

Andrzej Cichocki[‡]

Toshihisa Tanaka^{†,‡}

[†]Tokyo University of Agriculture and Technology, Tokyo, Japan [‡]RIKEN Brain Science Institute, Saitama, Japan Emails: higashi@sip.tuat.ac.jp, cia@brain.riken.jp, tanakat@cc.tuat.ac.jp

ABSTRACT

We propose a regularization based on geometric structure for feature extraction in a sensor array for brain data recordings. The purpose of the study is to add a penalty term using distances between sensors as the geometric information for finding spatial weights. The regularization term is derived under the definition of neighbors of sensors. We evaluate the proposed regularization in common spatial pattern (CSP) which is a well-known feature extraction method for EEG based brain computer interface (BCI). We have demonstrated the CSP procedure with the regularization by simulation for artificial signals. The results show that the proposed method works better than standard CSP in extracting of a component generated in a certain brain spot. Moreover, the classification experimental results using dataset of motor imagery based BCI suggest that the proposed method achieved maximum improvement by 27% in the classification accuracy over the standard CSP in a setting of even when we use only five samples.

Index Terms— sensor arrays, regularization, brain computer interfaces, electroencephalography

1. INTRODUCTION

Regularization is a widely used method to prevent overfitting or to solve an ill-posed problem in signal processing and machine learning [1-3]. The regularization for an optimization problem is to add to an original cost function a penalty term which represents additional information such as smoothness and bounds of the vector norm of optimization parameters. This technique is useful when we process signals observed with a sensor array which is an observation system using several sensors which are installed apart from each other. A sensor array is used to improve the signal to noise ratio (SNR) of signals or to separate some source signals from observations. In 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, sensory arrays are widely used. However, in the case of electrophysiological signals, extracting features with limited number of sensors is highly ill-posed. This motivates us to establish a regularization method that exploits the geometry of sensors.

In this paper, we consider regularization for weights in the following classical problem;

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

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

where M is the number of sensors, $x_i(t)$ is a signal observed at time t in the *i*th sensor that captures signals from brain, w_i is a weight for the *i*the sensor, and y(t) is an extracted feature. In this problem, we seek the weight vector $\boldsymbol{w} = [w_1, \dots, w_M]^T$ to extract certain components observed in the sensor array. This problem involves many signal processing techniques such as principal component analysis (PCA) and independent component analysis (ICA) [4]. In this case, the weight vector w can be regarded as spatial weight. In some situations, the signals measured by close sensors are similar and the observed features are similar. Consider a measurement system of EEG which observes faint electrical difference by electrodes installed in scalp. The EEG reflects the summation of the synchronous activity of thousands or millions of neurons [5,6]. Therefore, compared to separate sensors, close sensors observe activities which are induced from the same neurons. As results, when spatial weight w is designed with the purposes of feature extraction and/or improvement SNR, the weight coefficients corresponding to the close sensors can take similar value. Our proposed method is to use this prior information by introducing a geometric structure based regularization. The proposed regularization term evaluates a value like a second derivative of weight coefficients in close sensors.

Moreover, we apply the proposed regularization for the common spatial pattern (CSP) [7, 8] method. CSP, which uses spatial weights that extract the most discriminative information, is an efficient method for extracting the brain activity for EEG based brain computer interface (BCI) [6]. For the regularization, we define a distance of two electrodes on international 10-20/10-10/10-5 methods. Because the regularization term can be formulated in a quadratic form, the regularized CSP can be solved with a generalized eigenvalue problem. CSP with the proposed regularization has been demonstrated for artificial signals to show close electrodes have similar weight coefficients. The classification experiment for motor imagery based BCI (MI-BCI) dataset has been conducted with comparing an existing regularized CSP [9] and the proposed method demonstrated improvement of classification accuracy in a setting of the small number of samples.

2. REGULARIZATION BASED ON GEOMETRIC STRUCTURE ON HEAD SURFACE

We propose a regularization of using geometric structures for a sensor array on head surface. We first introduce geometric structure on head surface, that is, a distance between electrodes, of EEG measurement system in Sec. 2.1. The regularization is derived with the defined distance in Sec. 2.2.



Figure 1: Electrode arrangement of the international 10-20 method on the orthogonal coordinates. The red circles represent the electrodes.

2.1. A distance between electrodes

We introduce the electrode arrangements widely used for EEG measurement and define a distance between electrodes on the arrangements. When spatial high resolution EEG measurement is required, international 10-20, 10-10, and 10-5 methods [5, 10, 11] have stood as the de-facto standard of electrode arrangement. These systems describe head surface locations via relative distances between cranial landmarks over the head surface. In the international 10-20 method, the position determined as follows. Reference locations are nasion, level with the eyes, and inion [5]. From these points, the skull perimeters are measured in the transverse and median planes. Electrode locations are determined by dividing these perimeters into 10 % and 20 % intervals. Although the shape of head depends on the difference among individuals, the shape of head is supposed as a sphere, and we define coordinates to describe an electrode arrangement. The electrode positions defined the international 10-20 method are illustrated in Fig. 1. The position on head surface can be represented as $\{x, y, z\}$.

We want to know 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. Given the position of the two electrodes as $\{x_1, y_1, z_1\}$ and $\{x_2, y_2, z_2\}$. The angle between the line segments between the origin and the electrode positions can be given as $\psi = \arccos(x_1x_2 + y_1y_2 + z_1z_2)$, because $\sqrt{x^2 + y^2 + z^2} = 1$. As the unit of ψ is radian, the distance of the two electrodes given by $d = 2\pi\psi/2\pi = \psi$. The distance between the *i*th and the *j*th electrodes is represented as d_{ij} . Figure 1 illustrates this metric by showing 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.

2.2. Regularization

Suppose a sensor array consists of M sensors. By the metric defined in Sec. 2.1, we obtain d_{ij} for i, j = 1, ..., M as the distances between sensors. We next define the transform function for d_{ij} as

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

where p is a parameter to decide close sensors of a sensors. In order to evaluate the weight differences with considering geometric

feature, we define the cost:

$$P(\boldsymbol{w}) = \sum_{i=1}^{M} \left| \sum_{j=1}^{M} g_{ij}(w_i - w_j) \right|^2.$$
 (3)

We can transform a term of cost as $\sum_{j=1}^{M} g_{ij}(w_i - w_j) = (\sum_{j=1, j \neq i}^{M} g_{ij})w_i - \sum_{j=1, j \neq i}^{M} g_{ij}w_j$, so we can regard the term as a Laplacian filter for *i*th sensor under an assumption of a uniform sensor arrangement.

Eq. (3) can be transformed to matrix vector form as follows. A matrix, G, and a diagonal matrix, D, are defined as

$$[\mathbf{G}]_{ij} = g_{ij}, \quad [\mathbf{D}]_{ii} = \sum_{k=1}^{M} g_{ik}, \quad i, j = 1, \dots, M.$$
 (4)

By using them, (3) can be represented in matrix-vector form as

$$P(\boldsymbol{w}) = \boldsymbol{w}^T (\boldsymbol{D} - \boldsymbol{G}) (\boldsymbol{D} - \boldsymbol{G})^T \boldsymbol{w}.$$
 (5)

3. COMMON SPATIAL PATTERN WITH THE GEOMETRIC STRUCTURE BASED REGULARIZATION

BCI is an interface using brain signals as inputs. The input signals are classified to a class corresponding to a mental task or an external stimulus in BCIs. CSP is an effective method for the feature extraction and classification in two class MI-BCI [7,8]. In this section, we propose a method of CSP with the regularization explained in Sec. 2. We first review basic CSP algorithm and add the regularization into CSP.

3.1. Common spatial pattern (CSP) [7,8]

CSP is a method that designs spatial weights extracting a signal of which a variance is different between BCI classes (e.g. left hand and right hand classes) [7,8]. The problem using labeled learning samples to design the spatial weights can be formulated as follows. Let $\boldsymbol{X} \in \mathbb{R}^{M imes N}$ be a matrix representing observed signals, where M is the number of channels and N is the number of samples. CSP finds a spatial weight vector, $\boldsymbol{w} \in \mathbb{R}^M$, in such a way that the variance of a signal extracted by linear combination of X and w is minimized in a class [8]. Denote the components (vectors) of X by $\boldsymbol{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_N]$, where $\boldsymbol{x}_n \in \mathbb{R}^M$ and n is the time index. The time average of the observed signal is given by $\boldsymbol{\mu} = N^{-1} \sum_{n=1}^{N} \boldsymbol{x}_n$. Then, the time variance of the extracted signal of \boldsymbol{X} is given by $\sigma^2(\boldsymbol{X}, \boldsymbol{w}) = N^{-1} \sum_{n=1}^{N} |\boldsymbol{w}^T(\boldsymbol{x}_n - \boldsymbol{\mu})|^2$, where \cdot^T denotes the transpose of a vector or matrix. We assume that sets of the learning data are represented as C_1 and C_2 , where 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 finds w_c by solving the following optimization problem [7,8];

$$\min_{\boldsymbol{w}} \quad E_{\boldsymbol{X} \in \mathcal{C}_c}[\sigma^2(\boldsymbol{X}, \boldsymbol{w})],$$

subject to
$$\sum_{d=1,2} E_{\boldsymbol{X} \in \mathcal{C}_d}[\sigma^2(\boldsymbol{X}, \boldsymbol{w})] = 1,$$
 (6)

where $E_{\boldsymbol{X} \in \mathcal{C}_d}[\cdot]$ denotes the expectation over \mathcal{C}_d . Then, (6) can be rewritten as

$$\min_{\boldsymbol{w}} \quad \boldsymbol{w}^T \boldsymbol{\Sigma}_c \boldsymbol{w}, \quad \text{subject to} \quad \boldsymbol{w}^T (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2) \boldsymbol{w} = 1, \quad (7)$$

where Σ_d are defined as $\Sigma_d = E_{X \in C_d} [N^{-1} \sum_{n=1}^N (x_n - \mu)(x_n - \mu)^T]$, for d = 1, 2. The solution of (7) is given by the generalized

eigenvector corresponding to the smallest generalized eigenvalue of the generalized eigenvalue problem described as

$$\boldsymbol{\Sigma}_c \boldsymbol{w} = \lambda (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2) \boldsymbol{w}. \tag{8}$$

Though the solution of (7) is given by the eigenvector corresponding to the smallest eigenvalue in (8), we can use the other eigenvectors for feature extraction [7,8]. The M eigenvectors can be obtained by solving (8) as $\hat{w}_1, \ldots, \hat{w}_M$, where \hat{w}_i is the eigenvector corresponding to the *i*-th largest eigenvalue of (8). We assume that the 2r eigenvectors are used to classify an unlabeled data, X. Then we obtain the feature vector, $y \in \mathbb{R}^{2r}$, from X defined as $y = [\sigma^2(X, \hat{w}_1), \ldots, \sigma^2(X, \hat{w}_r), \sigma^2(X, \hat{w}_{M-r+1}), \ldots, \sigma^2(X, \hat{w}_M)]^T$.

3.2. Regularized common spatial pattern

We add the regularization defined in (5) into CSP. G and D used in P(w) are calculated by the distance defined in Sec. 2.1 with the parameter, r. The optimization problem with the regularization are defined as

$$\min_{\boldsymbol{w}} \quad \boldsymbol{w}^T (\boldsymbol{\Sigma}_c + \gamma (\boldsymbol{D} - \boldsymbol{G}) (\boldsymbol{D} - \boldsymbol{G})^T) \boldsymbol{w},$$

subject to $\boldsymbol{w}^T (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2) \boldsymbol{w} = 1,$ (9)

where γ is a combination coefficient. When the matrices of $\Sigma_c + \gamma (\boldsymbol{D} - \boldsymbol{G})(\boldsymbol{D} - \boldsymbol{G})^T$ and $\Sigma_1 + \Sigma_2$ are nonsingular, (9) can be solved by the generalized eigenvalue problem:

$$(\boldsymbol{\Sigma}_c + \gamma (\boldsymbol{D} - \boldsymbol{G}) (\boldsymbol{D} - \boldsymbol{G})^T) \boldsymbol{w} = \lambda (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2) \boldsymbol{w}.$$
 (10)

The feature vector is extracted as follows. In each case of c = 1 and c = 2, we solve (10), and we get 2M eigenvectors as $\hat{w}_i^{(1)}$ and $\hat{w}_i^{(2)}$ (i = 1, ..., M) in each eigenvalue problem. By using the weight vectors, the feature vector is defined as $\boldsymbol{y} = [\sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_1^{(1)}), ..., \sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_r^{(1)}), \sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_1^{(2)}), ..., \sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_r^{(2)})]^T$.

4. EXPERIMENTS

We demonstrated the regularized CSP in artificial signals and realworld EEG signals. In the simulation of artificial signals, we evaluate the ability of extracting a local feature. Moreover, we classify EEG signals by spatially weighting of the proposed method.

4.1. Simulation using artificial signals

We used artificial signals including source signals to be extracted. We know the spatial distributions of the source signals to compare the spatial weights given by CSP to true distributions. The artificial signals were generated as follows. Given $s_1[t]$ and $s_2[t]$ as source signals which are assumed to be related to two BCI classes, where t is the discrete time index. The signals belonging to class 1 and 2 are generated by

$$\boldsymbol{x}_{1}[t] = \boldsymbol{a}_{1}s_{1}[t] + 2\boldsymbol{a}_{2}s_{2}[t] + \beta\eta_{1}[t], \qquad (11)$$

$$\boldsymbol{x}_{2}[t] = 2\boldsymbol{a}_{1}\boldsymbol{s}_{1}[t] + \boldsymbol{a}_{2}\boldsymbol{s}_{2}[t] + \beta\eta_{2}[t], \tag{12}$$

where $x_i[t]$ is observed signals in all channel at the time index t, a_i is a spatial distribution of a source signal, β is a coefficient to determine SNR, and $\eta_i[t]$ is a noise signal. The simulation settings are as follows. Figure 2a shows $s_1[t]$ and $s_2[t]$. As the signal length was set to 1 second with sampling frequency of 512 Hz, the number of samples in the signal was 512. As shown in Fig. 2b, the source



(a) Source signals (solid: source 1, (b) Spectrum of source signals dashed: source 2). (solid: source 1, dashed: source 2).



(c) Spatial distributions of source sig- (d) Examples of observed signals. nals.

Figure 2: The artificial signals.



Figure 3: Topographical maps of the spatial weights in the experiment using the artificial signals.

signals have same frequency spectra but differ in phase for each frequency. The international extended 10-20 method was adopted as the electrode arrangement and the number of the electrodes was 118. The spatial distributions represented by a_1 and a_2 were based on Gaussian distributions and the topographically showed spatial distributions are shown in Fig. 2c. $\eta_1[t]$ and $\eta_2[t]$ were random values generated from a Gaussian distribution (mean 0, variance 1) and β was set to 0.01. Examples of observed signals are shown in Fig. 2d.

Figure 3 shows the topographically plotted spatial weights given by CSP and the proposed method. The proposed method implemented under the parameters of p = 0.05 and $\gamma = 10^{10}$. Comparing with CSP, in the proposed method, the large weight electrodes concentrate at the certain spots. Moreover we can observe in Fig. 3b that the topographical map of the spatial weights given by the proposed method is close to the true distribution map shown in Fig 2c.

4.2. Classification of real-world EEG signals

We compare performance in classifying EEG signals during motor imagery using the proposed method to those using standard CSP and the spatially regularized CSP (SRCSP) [9], respectively. The EEG signals were classified to two classes by spatially weighting and a classifier.

4.2.1. 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

Table 1: Accuracy [%] given by 100 learning samples per a class.

	aa	al	av	aw	ay	Ave.
CSP	81.7	94.7	54.2	93.6	89.1	82.7
SRCSP	82.5	95.2	67.4	95.0	92.3	86.5
Proposed	82.6	95.4	67.5	95.1	92.3	86.6

Table 2: Accuracy [%] given by 5 learning samples per a class.

	aa	al	av	aw	ay	Ave.
CSP	52.6	65.6	51.3	62.3	51.3	56.6
SRCSP	59.4	76.9	55.0	75.8	77.6	68.9
Proposed	60.2	77.7	55.0	76.4	78.9	69.6

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, we applied to this data the lowpass filter whose the cutoff frequency is 50 Hz, and downsampled to 100 Hz. The dataset for each subject consisted of signals of 140 trials per a class. A signal of one trial was measured for 3.5 seconds.

4.2.2. Results

Before we apply CSP and the proposed method, the signals were bandpass filtered with the passband of 7–30 Hz. The feature vectors extracted by CSPs were classified by linear discriminant analysis [3].

The classification accuracy was given by learning using randomchosen 100 samples and testing using the remaining samples. The accuracy shown in Table 1 is an average accuracy over 100 times of this procedure. For SRCSP and the proposed method, we chose the parameters out of $p \in \{0.025, 0.05, 0.075, 0.1\}$ and $\gamma \in \{10^0, 10^1, \ldots, 10^{30}\}$. The dimension of the feature vector 2r was set to 2. The best accuracy among the parameters for each subject is shown in Table 1. For *aa*, *al*, *av*, *aw*, and *ay*, $\{r, \gamma\}$ are set to $\{0.025, 10^{19}\}$, $\{0.05, 10^{10}\}$, $\{0.025, 10^{25}\}$, $\{0.05, 10^8\}$, and $\{0.05, 10^{10}\}$, respectively, in SRCSP. In the proposed method, $\{r, \gamma\}$ are set to $\{0.05, 10^{12}\}$, $\{0.05, 10^{16}\}$, $\{0.05, 10^{14}\}$, $\{0.075, 10^7\}$, and $\{0.05, 10^{16}\}$, respectively. The both of regularized CSP algorithms slightly outperform the standard CSP in the classification accuracy for all subjects.

Table 2 also shows classification accuracy, however when the number of learning samples is considerably reduced to only 5 samples. As the same as in Table 1, the parameters performing the best classification accuracy are chosen out of the candidates. For *aa*, *al*, *av*, *aw*, and *ay*, $\{r, \gamma\}$ are set to $\{0.05, 10^{9}\}$, $\{0.025, 10^{25}\}$, $\{0.05, 10^{11}\}$, $\{0.05, 10^{10}\}$, and $\{0.05, 10^{10}\}$, respectively, in SRCSP. In the proposed method, $\{r, \gamma\}$ are set to $\{0.05, 10^{15}\}$, $\{0.05, 10^{16}\}$, $\{0.05, 10^{16}\}$, $\{0.05, 10^{16}\}$, $\{0.05, 10^{16}\}$, respectively. We can observe significant improvement of accuracy for subject *ay*. The result suggests that the proposed regularization can improve the accuracy even if the number of available learning samples is small.

Figure 4 shows the topographically plotted spatial weights for subject *ay*. The spatial weights were found by using all samples of the dataset. The parameters of the proposed method, *c* and γ , were set to 0.05 and 10¹⁵, respectively. Comparing with CSP, the electrodes which have large coefficients do not be scattered spatially in the proposed method.



Figure 4: Topographical maps of the spatial weights for subject *ay* (EEG signals).

5. CONCLUSION

We have proposed the regularization based on the geometric information for feature extraction problem in an EEG sensor array on head surface. We illustrated the regularization procedure which was used for CSP for artificial signals and MI-BCI dataset. The results demonstrated that the proposed method improves classification accuracy in a setting of the small number of samples. Although the regularization works under the assumption of a uniform sensor arrangement, the arrangement is not spatially uniform always. For future works, we have to solve the problem.

6. REFERENCES

- A. N. Tikhonov and V. Y. Aresenin, Solutions of Ill-Posed Problems. Washington, D.C.: Winston, 1977.
- [2] T. Poggio and F. Girosi, "Regularization algorithms for learning that are equivalent to multilayer networks," *Science*, vol. 247, pp. 978–982, 1990.
- [3] C. M. Bishop, Pattern Recognition and Machine Learning. New York, Springer, 2006.
- [4] A. Cichocki and S. Amari, Adaptive Blind Signal and Image Processing. Hoboken, NJ: John Wiley & Sons, 2002.
- [5] J. Malmivuo and R. Plonsey, Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields. Oxford and New York: Oxford University Press, 1995.
- [6] J. R. Wolpaw, N. Bribaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, pp. 767– 791, 2002.
- [7] J. Müller-Gerling, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clin. Neurophysiol.*, vol. 110, pp. 787–789, 1999.
- [8] H. Ramoser, J. Müller-Gerling, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 4, pp. 441– 447, 2000.
- [9] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, 2011.
- [10] M. R. Nuwer, "Recording electrode site nomenclature," *Journal of Clinical Neurophysiology*, vol. 4, pp. 121–133, 1987.
- [11] R. Oostenveld, P. Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements," *Clinical Neurophysiology*, vol. 112, pp. 713–719, 2001.