Bilinear Regularized Locality Preserving Learning on Riemannian Graph for Motor Imagery BCI

Xiaofeng Xie, *Student Member, IEEE,* Zhu Liang Yu, *Member, IEEE,* Zhenghui Gu, *Member, IEEE,*
Jun Zhang, *Member, IEEE,* Ling Cen, *Member, IEEE,* and Yuanqing Li, *Fellow, IEEE*

Abstract—In off-line training of motor imagery-based brain-computer interfaces (BCIs), to enhance the generalization performance of the learned classifier, the local information contained in test data could be used to improve the performance of motor imagery as well. Further considering that the covariance matrices of electroencephalogram (EEG) signal lie on Riemannian manifold, in this paper, we construct a Riemannian graph to incorporate the information of training and test data into processing. The adjacency and weight in Riemannian graph are determined by the geodesic distance of Riemannian manifold. Then, a new graph embedding algorithm, called bilinear regularized locality preserving (BRLP), is derived upon the Riemannian graph for addressing the problems of high dimensionality frequently arising in BCIs. With a proposed regularization term encoding prior information of EEG channels, the BRLP could obtain more robust performance. Finally, an efficient classification algorithm based on extreme learning machine (ELM) is proposed to perform on the tangent space of learned embedding. Experimental evaluations on the BCI competition and in-house datasets reveal that the proposed algorithms could obtain significantly higher performance than many competition algorithms after using same filter process.

Index Terms—Motor imagery, Riemannian manifold, graph embedding, regularization, extreme learning machine

I. Introduction

MOtor imagery is receiving increasing attention in the field of brain computer interfaces (BCIs) owing to its advantage of no external stimulations required [1]–[3]. The trained subject can voluntarily produce electroencephalogram (EEG) signal by imagining movements of different parts of body [4], [5]. In motor imagery system, the signal is recorded with multi-channel electrodes which cover the area of the particular brain activities. One of the major challenges frequently arising in motor imagery system is the high dimensionality of EEG signal.

In the processing of motor imagery EEG signal, common spatial pattern (CSP) is the most effective algorithm for the feature extraction [6], [7]. The CSP not only extracts the most discriminative features from motor imagery signal but also provides the weight knowledge for channel selection [8]. Taking two classes motor imagery as example, the CSP is designed to learn the spatial filters that maximize the

X. Xie, Z. L. Yu, Z. Gu and Y. Li are with the College of Automation Science and Engineering, South China University of Technology, Guangzhou, China, 510641. E-mail: zlyu@scut.edu.cn

J. Zhang and L. Cen are with School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China.

variance for one class of EEG data while minimizing the variance of the others. The CSP well used in motor imagery attributes the success to the spatial filters learned from training data set [9]. In practical motor imagery systems, one purpose is to reduce the training effort, because the tedious and time-consuming training process will limit the application of motor imagery system [10]. However, small training data will easily result in a worse prediction model in the supervised learning method. One potential solution for this problem is semi-supervised learning [11]. It can utilize both training and test data to boost the algorithmic performance. Many related works, like [12]–[14], were proposed to classify EEG based on training and test data in the context of semi-supervised learning. In [12], a self-training support vector machine (SVM) method was proposed to use the test data for updating the parameter of SVM model. In [13], an iterative semi-supervised SVM method was proposed for channel selection and classifier training. In [14], a semi-supervised expectation maximization (EM) was incorporated into CSP to extract more effective features. Recently, since graph can characterize the structure information of training and test data, graph-based semi-supervised learning method has also received wide attention [15]. The graph-based model has elegant mathematical formulation and effectiveness by mining the intrinsic geometrical structure inferred from both training and test data.

There are many traditional methods for constructing graph, e.g., k-nearest neighbor and ε-ball [16]. Both of these methods compute the relationship of data with Euclidean distance. It is easy to be implemented and applied in practice. However, if the data samples do not lie on Euclidean space, the traditional methods cannot guarantee the connectivity of the whole graph and often lead to several separated subgraphs, since the Euclidean distance cannot well characterize the real geometrical relations among samples [17]. An obvious example is the covariance matrices of EEG signal, which commonly used in the learning of spatial filters in CSP. The covariance matrices with form of symmetric positive definite (SPD) matrices lie on the Riemannian manifold [18]. Thus, we can construct a graph based on Riemannian geodesic distance, namely Riemannian graph, for the processing of the covariance matrix. Moreover, we expect to address the problem of high dimensionality with the Riemannian graph.

For the problem of high dimensionality, many graph embedding algorithms, e.g., locally linear embedding (LLE)

[19], locality preserving projection [20], Laplacian Eigenmaps [21], have been proposed to learn a low-dimensional embedding from the graph. All of them attempt to preserve the similarities between the vertex pairs on the graph into embedding. However, the above algorithms need to unfold input data to vector form that will destroy the structural information of data, especially for the SPD matrices on Riemannian manifold. Recently, many bilinear mapping algorithms have been proposed to identify the spatial and temporal projection matrices for the EEG signal [22]–[24]. Taking account both of the locality and structure preserving, we can design a bilinear mapping to seek a low-dimensional embedding from the Riemannian graph.

Further, we expect to exploit the prior information of channels to enhance the performance of dimensionality reduction. The prior information is associated with the weights of channels. Since some channels are more important than the others in motor imagery tasks, it is better to assign different weights for different channels [25]. One potential way to exploit the prior information of channel is regularization technique [26]. It encodes the prior information of channels into a regularization term. As an example, the studies in [26] extended the CSP with a quadratic regularization term to obtain an efficient spatial filter. Similarly, in this paper, we extended the bilinear mapping model on Riemannian graph with a regularization term, called the bilinear regularized locality preserving (BRLP).

After alleviating the over-fitting problem by dimensionality reduction, we expect to design an advanced classifier based on the Riemannian geometry for motor imagery classification. However, since the Riemannian manifold is not a linear vector space, many classical classifiers that work on Euclidean space cannot be implemented directly on Riemannian manifold. One potential solution is to project all data points on Riemannian manifold into its tangent space, which is Euclidean space. The relationship of data points on Riemannian manifold can be faithfully preserved on the tangent space with Riemannian mean used as tangent point [27]. Thus, many classical classifiers, like LDA, SVM, and neural network, can be implemented on such tangent space. For instance, in [18] and [28], the LDA and SVM classifiers have been applied on the tangent space for motor imagery classification, respectively. Although the training of LDA and SVM is more efficient than that of neural network, the neural network is more flexible to handle complex and multivariate data. In this paper, we applied the neural network on the tangent space for classification. Furthermore, to alleviate the expensive training process in the neural network, we used extreme learning machine (ELM) as a fast learning algorithm for the neural network, because the hidden node parameters in ELM can be randomly generated without tuning [29]. The proposed novel classification algorithm is called extreme learning machine on tangent space of regularized embedding (ELM-TS-RE).

The major contributions of this paper are threefold.

1) The Riemannian graph is proposed to model the co-variance matrices of training and test EEG data. It is a significant improved version of traditional graph while the Euclidean distance results in a bias representation for the space of covariance matrices. In Riemannian graph, the Riemannian geodesic distance is used to characterize the neighbor of covariance matrix.

2) A novel BRLP algorithm is proposed to address the problem of high dimensionality frequently arising in motor imagery. In BRLP, the regularization term encoded the prior information of channels is incorporated into the objective function of locality preserving. The main difference of BRLP against others bilinear algorithms is that the BRLP preserves the local structure of data on Riemannian manifold. The BRLP also has a low computational cost in practice, since the optimization function can formulate as an eigenvalue problem.

3) An ELM-TS-RE algorithm is proposed to classify motor imagery EEG signal. The ELM-TS-RE includes a graph-based method for dimensionality reduction, tangent space mapping for feature extraction and ELM network for classification. This overall ensemble is novel and efficient. Moreover, the proposed method is tested on the dataset IIa of BCI competition IV and in-house datasets. The experimental results show that proposed method achieves a higher performance relative to that achieved by competing methods.

The rest of the paper is organized as follows. In Section II, we review the space of SPD matrices and spatial filter of CSP. In Section III, we first introduce the construction of Riemannian graph. Then we propose BRLP for graph embedding and ELM-TS-RE for classification. Extensive experimental results are given in Section IV to show the effectiveness of the proposed method. Finally, in Section V, some conclusions are given.

II. RELATED WORKS

In this section, the space of SPD matrices and the spatial filter of CSP are briefly reviewed to provide some basic knowledge for the proposed methods.

A. Space of Symmetric Positive Definite Matrices
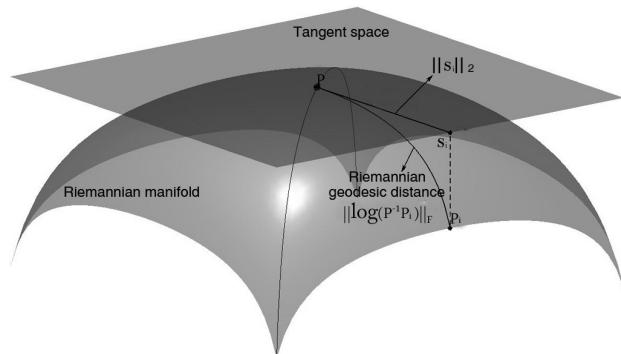
Figure 1. An illustration of Riemannian manifold and tangent space.

Denoting the space of symmetric matrices

$$\mathcal{S}(N) = \left\{ \mathbf{P} \in \mathbb{R}^{N \times N}, \mathbf{P} = \mathbf{P}^T \right\} \tag{1}$$

and the space of positive-definite matrices

$$\mathcal{P}(N) = \left\{ \mathbf{P} \in \mathbb{R}^{N \times N}, \mathbf{u}^T \mathbf{P} \mathbf{u} > 0, \forall \mathbf{u} \in \mathbb{R}^N \right\}, \tag{2}$$

the space of SPD matrices is defined as

$$\mathcal{SPD}(N) = \mathcal{S}(N) \cap \mathcal{P}(N). \tag{3}$$

The SPD matrices lie on a differentiable Riemannian manifold [30]. Hence, all the mathematical tools defined in Riemannian geometry can be applied to $\mathcal{SPD}(N)$.

The Riemannian distance between two matrices $\mathbf{P}_1, \mathbf{P}_2 \in \mathcal{SPD}(N)$ is defined as

$$\delta_R(\mathbf{P}_1, \mathbf{P}_2) = \left\| \log(\mathbf{P}_1^{-1} \mathbf{P}_2) \right\|_F = \left[\sum_{i=1}^{N} \log^2 \eta_i \right]^{\frac{1}{2}} \tag{4}$$

where $||\cdot||_F$ is the Frobenius norm of a matrix and η_i is the i-th real eigenvalue of $\mathbf{P}_1^{-1}\mathbf{P}_2$. The Riemannian distance $\delta_R(\mathbf{P}_1, \mathbf{P}_2)$ is the minimum length of curves connecting $\mathbf{P}_1$ and $\mathbf{P}_2$ on Riemannian manifold [31]. The Riemannian distance poses three fundamental properties of metric space, i.e., positivity, symmetry and triangle inequality [30].

Tangent space, as a Euclidean space, is an important space in the analysis of Riemannian manifold. The tangent space $\mathcal{T}(N)$ at $\mathbf{P}$ is defined as [27]

$$\mathcal{T}(N) = \left\{ \mathbf{s}_i = upper(\mathbf{P}^{-\frac{1}{2}} Log_{\mathbf{P}}(\mathbf{P}_i) \mathbf{P}^{-\frac{1}{2}}) \in \mathbb{R}^{N(N+1)/2} \right\} \tag{5}$$

where upper($\cdot$) operator is to keep the upper triangular part of matrix and vectorize it. The logarithmic mapping operator is denoted as $Log_{\mathbf{P}}(\mathbf{P}_i) = \mathbf{P}^{\frac{1}{2}} \log(\mathbf{P}^{-\frac{1}{2}} \mathbf{P}_i \mathbf{P}^{-\frac{1}{2}}) \mathbf{P}^{\frac{1}{2}}$. The relationship between Riemannian manifold and tangent space is shown in Fig. 1.

The Riemannian mean of SPD matrices also plays an important role in classification and is defined as the point $\mathbf{P}_R \in \mathcal{SPD}(N)$, which has a minimum sum of the squared distances to all SPD matrices in dataset $\mathcal{C}$

$$\mathbf{P}_R = \arg \min_{\mathbf{P} \in \mathcal{SPD}(N)} \sum_{\mathbf{P}_i \in \mathcal{C}} \delta^2_R(\mathbf{P}, \mathbf{P}_i). \tag{6}$$

B. Spatial Filter for EEG Signal

EEG signal is recorded from multiple channels and represented as

$$\mathbf{X}(t) = [\mathbf{x}(t), ..., \mathbf{x}(t + L - 1)] \in \mathbb{R}^{N \times L} \tag{7}$$

where N and L denote the number of channels and sampled points, and $\mathbf{x}(t) = [x_1(t), ..., x_N(t)]^T \in \mathbb{R}^{N \times 1}$ is the snapshot vector. The spatial covariance matrix of $\mathbf{X}(t)$ is represented by

$$\mathbf{P}(t) = \frac{1}{L-1} \mathbf{X}(t) \mathbf{X}^T(t) \in \mathcal{SPD}(N). \tag{8}$$

The well applied spatial filter for EEG signal, CSP, is learned from the covariance matrices. The CSP aims to find a spatial filtering matrix $\mathbf{W} \in \mathbb{R}^{N \times N_s}$ that maximize the variance of one-hand trials while minimizing the variance of the other [6], where N_s is number of the spatial filters. Generally, the matrix $\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2]$ contains two submatrices $\mathbf{W}_1, \mathbf{W}_2 \in \mathbb{R}^{N \times \frac{N_s}{2}}$, which can be obtained by maximizing and minimizing the following cost function, respectively [22]:

$$J(\tilde{\mathbf{W}}) = \frac{tr(\tilde{\mathbf{W}}^T \mathbf{C}_1 \tilde{\mathbf{W}})}{tr(\tilde{\mathbf{W}}^T \mathbf{C}_2 \tilde{\mathbf{W}})} \tag{9}$$

where $\tilde{\mathbf{W}} \in \mathbb{R}^{N \times \frac{N_s}{2}}$ and the $\mathbf{C}_l$ is the arithmetic mean of covariance matrices of EEG signals belonging to class l. The optimal matrix $\mathbf{W}$ is composed of the eigenvectors of $\mathbf{C}_2^{-1} \mathbf{C}_1$, which correspond to its first $\frac{N_s}{2}$ largest and $\frac{N_s}{2}$ smallest eigenvalues. The i-th column $\mathbf{w}_i \in \mathbb{R}^{N \times 1}$ of $\mathbf{W}$ is called a spatial filter for feature extraction.

III. PROPOSED METHODS

In this section, we propose the BRLP to learn the low-dimensional embedding from Riemannian graph and the ELM-TS-RE for classification on the learned embedding.

A. Riemannian Graph

To learn low-dimensional embedding of Riemannian manifold, a Riemannian graph is firstly constructed for the data points on the Riemannian manifold. The Riemannian graph $G_R = (\mathcal{V}, \mathcal{E})$ is composed of vertices $\mathcal{V}$ and edges $\mathcal{E}$ with weight u_{ij}. Thus, the constructions of adjacency and weights are two main steps for Riemannian graph. First, the adjacency of Riemannian graph is designed by k-nearest neighbors. For data point $\mathbf{P}_i \in \mathcal{V}$, we select its neighbors according to Riemannian geodesic distance. This key step ensures that the neighbors of Riemannian graph can really reflect the local structure of data on Riemannian manifold. Secondly, the weight between two adjacent points $\mathbf{P}_i, \mathbf{P}_j \in \mathcal{V}$ is given by [32]

$$u_{ij} = \begin{cases} e^{\frac{-d_{ij}^2}{2\sigma^2}} & \text{if } \mathbf{P}_i \text{ and } \mathbf{P}_j \text{ are neighbors,} \\ 0 & \text{otherwise} \end{cases}$$

where $d_{ij} = \delta_R(\mathbf{P}_i, \mathbf{P}_j)$ and the parameter σ is a scaling factor.

B. Bilinear Regularized Locality Preserving on Riemannian Graph

A bilinear mapping $\mathbf{V} \in \mathbb{R}^{M \times N}$ with constraint $\mathbf{V}\mathbf{V}^T = \mathbf{I}_M$ is used to learn a low-dimensional embedding ($M < N$) from the Riemannian graph. The learned low-dimensional embedding $\mathbf{P}_e = \mathbf{V}\mathbf{P}\mathbf{V}^T \in \mathcal{SPD}(M)$ is expected to preserve the local structure of Riemannian graph and impose the prior knowledge of EEG channels.

A reasonable bilinear mapping respecting the locality preserving can be obtained by solving the following objective function:

$$\min_{\mathbf{V}} \sum_{i,j=1}^{|\mathcal{C}|} \| \mathbf{V}\mathbf{P}_i\mathbf{V}^T - \mathbf{V}\mathbf{P}_j\mathbf{V}^T \|_F^2 u_{ij}. \tag{10}$$
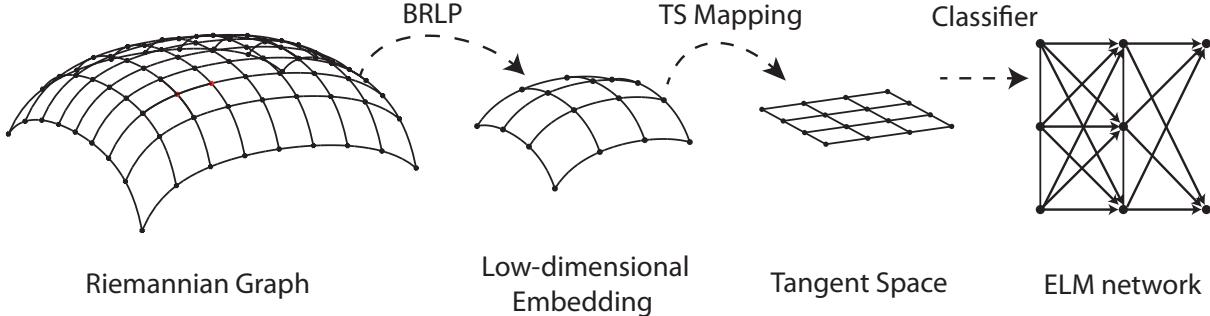
Figure 2. The graphical illustration of the ELM-TS-RE algorithm. The overall ensemble includes a graph-based approach for dimensionality reduction, tangent space (TS) mapping for feature extraction, and extreme learning machine (ELM) for classification.

It attempts to ensure that if $\mathbf{P}_i$ and $\mathbf{P}_j$ in high-dimensional manifold are "close" then $\mathbf{V}\mathbf{P}_i\mathbf{V}^T$ and $\mathbf{V}\mathbf{P}_j\mathbf{V}^T$ in low-dimensional embedding are "close" as well. There is no closed-form solution for (10). In this paper, we employ an alternating iterative strategy to learn the bilinear mapping matrix in (10). The matrix $\mathbf{V}$ is firstly initialized as $\mathbf{V}_0 = [\mathbf{I}_M, \mathbf{0}]$, and the new optimal $\mathbf{V}_{t+1}$ of the $(t+1)$th-iteration is learned from the following eigen-decomposition problem [33], [34] once $\mathbf{V}_t$ is given

$$\mathbf{V}_{t+1} = \arg\min_{\mathbf{V}} tr(\mathbf{V}\mathbf{L}\mathbf{V}^T) \qquad (11)$$

where $\mathbf{L} = \sum_{i,j=1}^{|\mathcal{C}|} (\mathbf{P}_i - \mathbf{P}_j)\mathbf{V}_t^T\mathbf{V}_t(\mathbf{P}_i - \mathbf{P}_j)^T u_{ij}$.

In order to impose the prior knowledge of EEG channels, we further introduce a regularization term to (11). The new method can be formulated as

$$\mathbf{V}_{t+1} = \arg\min_{\mathbf{V}} tr(\mathbf{V}\mathbf{L}\mathbf{V}^T + \lambda R(\mathbf{V})) \qquad (12)$$

where λ represents the regularization parameter and $R(\mathbf{V})$ is the regularization term. Since the regularization term with quadratic form could lead to computational efficiency, in this paper, we define the regularized term as $R(\mathbf{V}) = \mathbf{V}\mathbf{D}_w\mathbf{V}^T$, where the diagonal matrix $\mathbf{D}_w$ encodes the penalty on each EEG channel. The regularized term is used to penalize solutions that do not satisfy a given channels prior. Similarly, we also employ an iterative strategy to learn mapping matrix $\mathbf{V}$ in (12). The pseudo-code of the proposed BRLP algorithm is shown in Algorithm 1.

The problem remaining here is how to design the penalty matrix $\mathbf{D}_w$. Since the spatial filter of CSP carries the channel weighting information and is widely used for channel selection [26], in this paper, we recommend setting the penalty level of channel according to the information contained in the CSP spatial filters. The penalty matrix is given as the inverse of the average absolute values of N_s normalized spatial filters,

$$\mathbf{D}_w = diag(\frac{1}{N_s}\sum_{i=1}^{N_s}|\frac{\mathbf{w}_i}{\| \mathbf{w}_i \|}|)^{-1}. \qquad (13)$$

It is clear that the higher absolute value of filter weight leads to less penalty on the corresponding channel.

Algorithm 1 Bilinear Regularized Locality Preserving (BRLP)

Input: Given SPD matrix set $\mathcal{C}$ with samples $\mathbf{P}_i \in \mathbb{R}^{N \times N}, i = 1, \cdots, |\mathcal{C}|$, dimensions of embedding M, iterative number N_{ite}, stop threshold τ;

1: Construct a Riemannian graph G_R over all samples based on Riemannian geodesic distances (Sec. III-A);
2: Calculate the penalty matrix $\mathbf{D}_w$ by (13);
3: Initialize: $\mathbf{V}_0 = [\mathbf{I}_M, \mathbf{0}]$;
4: **for** t=1:1:(N_{ite}), **do**;
    Calculate the matrix $\mathbf{L}$ in (11);
    Obtain mapping matrix $\mathbf{V}_{t+1} \in \mathbb{R}^{M \times N}$ by (12);
    If $\| \mathbf{V}_{t+1} - \mathbf{V}_t \|_F^2 \leq \tau$
      break;
    end If
  end for
5: Construct $M \times M$ dimensional embedding $\mathbf{P}_{ei} = \mathbf{V}_{opt}\mathbf{P}_i\mathbf{V}_{opt}^T, i = 1, \cdots, |\mathcal{C}|$ where $\mathbf{V}_{opt}$ is convergent matrix in step 4;

Output: Learned SPD matrix samples on low dimensional embedding $\mathbf{P}_{ei} \in \mathbb{R}^{M \times M}, i = 1, \cdots, |\mathcal{C}|$.

C. Classification on Low-Dimensional Embedding

In this paper, we designed an ELM classifier performed on the tangent space of learning regularized embedding. Fig. 2 shows the graphical illustration of the ELM-TS-RE. It includes a graph-based method for dimensionality reduction, tangent space mapping for feature extraction and ELM network for classification. The proposed method enhances the existing methods [18] from two perspectives. Firstly, the local characteristics of Riemannian manifold and prior knowledge of EEG channels are condensed in a low-dimensional embedding to avoid over-fitting. Secondary, the ELM classifier is more flexible and powerful than LDA classifier.

To keep this paper self-contained, we present the ELM classifier [29] as follows. We design a three-layer ELM network, which includes input, hidden and output layer. The output weights $\boldsymbol{\beta} \in \mathbb{R}^{N_h \times N_o}$ between hidden layer

and output layer are obtained by

$$\min_{\boldsymbol{\beta}} \frac{1}{2}||\boldsymbol{\beta}||^2 + \frac{\theta}{2}\sum_{i=1}^{|\mathcal{C}|}||\mathbf{h}^T(\mathbf{s}_i)\boldsymbol{\beta} - \mathbf{y}_i||^2 \qquad (14)$$

where θ is a parameter to balance the loss and $L2$ regularization term in ELM, $\mathbf{s}_i \in \mathbb{R}^{\frac{M(M+1)}{2} \times 1}$ is the constructed feature vector of data points on tangent space of low-dimensional embedding, and $\mathbf{h}(\mathbf{s}_i) \in \mathbb{R}^{N_h \times 1}$ is output vector of the hidden layer with N_h neural elements. The output $\mathbf{y}_i \in \mathbb{R}^{1 \times N_o}$ is the label vector corresponding to $\mathbf{s}_i$ (N_o is the number of output neurons). The key principle of ELM [35] is that the parameters $(\mathbf{a}, b)$ in sigmoid mapping function $h(\mathbf{s}_i) = 1/(1 + \exp(-\mathbf{a}^T\mathbf{s}_i - b))$ are randomly generated within the range of $[-1, 1]$. The pseudo-code of proposed classification method is given in Algorithm 2.

Algorithm 2 ELM on Tangent Space of Regularized Embedding (ELM-TS-RE)

Input: Training and test SPD datasets $\mathbf{P}_{Tr}, \mathbf{P}_{Te}$, label of training data $\mathbf{y}_{Tr}$, dimensions of embedding M, iterative number N_{ite}, stop threshold τ, number of hidden neuron N_h, coefficient θ;

Output: Label of test data;

1: Obtain the optimal mapping matrix by BRLP, $[\mathbf{V}] = \text{BRLP}(\mathbf{P}_{Tr}, M, N_{ite}, \tau)$;
2: Map data onto the learned embedding with the size of $M \times M$ as $\mathbf{P}_{eTr} = \mathbf{V}\mathbf{P}_{Tr}\mathbf{V}^T, \mathbf{P}_{eTe} = \mathbf{V}\mathbf{P}_{Te}\mathbf{V}^T$;
3: Calculate the Riemannian mean of all data points as $\mathbf{P}_R = \arg\min_{\mathbf{P}} \sum_{\mathbf{P}_i} \delta^2_R(\mathbf{P}, \mathbf{P}_i)$, $\mathbf{P}_i \in \mathbf{P}_{eTr} \bigcup \mathbf{P}_{eTe}$;
4: Project data onto the tangent space of learned embedding
   $\mathbf{s}_{Tr} = \text{upper}(\mathbf{P}_R^{-\frac{1}{2}} Log(\mathbf{P}_{eTr})\mathbf{P}_R^{-\frac{1}{2}}) \in \mathbb{R}^{\frac{M(M+1)}{2} \times 1}$
   $\mathbf{s}_{Te} = \text{upper}(\mathbf{P}_R^{-\frac{1}{2}} Log(\mathbf{P}_{eTe})\mathbf{P}_R^{-\frac{1}{2}}) \in \mathbb{R}^{\frac{M(M+1)}{2} \times 1}$;
5: Train the output weight $\boldsymbol{\beta}$ of ELM network by feeding $\mathbf{s}_{Tr}, \mathbf{y}_{Tr}$ into (14);
6: Calculate the label of test data based on $\mathbf{s}_{Te}$: $\mathbf{y}_{Te} = \text{sign}(\mathbf{h}(\mathbf{s}_{Te})\boldsymbol{\beta})$;

IV. NUMERICAL RESULTS

A. Experimental Setup

Data Description: the dataset IIa of BCI competition IV and in-house motor imagery datasets were used to validate the effectiveness of the proposed methods.

1) Dataset IIa of BCI competition IV was recorded from 9 subjects (S01-S09) who performed four types of motor imagery tasks (right hand, left hand, foot and tongue imagined movements). The recorded signals consisted of 22 EEG channels. The protocol of the experiment was given as follows. In the initial time $(0 - 2s)$, a short acoustic warning tone was presented. After two seconds $(2s)$, a cue in the form of an arrow pointing left, right, down or up appeared and remained on the screen from $2s$ to $3.25s$. This prompted the subjects to perform the motor imagery task until the
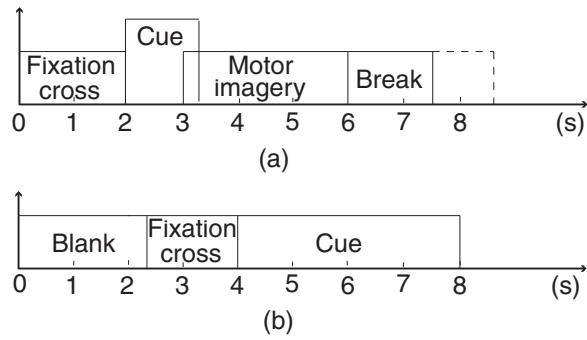
Figure 3. Timing scheme of the paradigm for motor imagery task. a) dataset II of competition IV; b) in-house datasets.

fixation cross disappears from the screen at $6s$. Lastly, there was a short break that lasted for 1.5s. The paradigm is illustrated in Fig. 3 (a). The time interval of processed data was restricted to the time segment between $3.75s$ and $5.75s$ during which the subject performed the mental tasks. For each subject and mental task, there were 72 training and 72 test trials. Thus, the overall number of training/test trials for each subject was 288/288. The EEG signals were sampled with a sampling rate 250Hz and filtered by an $8-30$Hz bandpass filter to analyze the μ and β rhythms.

2) Our in-house EEG data was recorded from 9 subjects (A01-A09) with 64 EEG channels. The protocol of the in-house experiment was given as follows. Two mental tasks, i.e., left/right hand imaged movements, were required to perform the in-house BCIs. In the initial interval $(0-2.25s)$, the screen remained blank. A cross appeared on the screen to attract the subject's visual fixation from $2.25s$ to $4s$. From $4s$ to $8s$, a left/right arrow cue was shown and the subject performed the required task. The paradigm is illustrated in Fig. 3 (b). The time interval for the processed data was restricted to the time segment between $5s$ and $7s$. For each subject and each task, there were 117 training and test trials. The overall number of training/test trials for each subject was 234/234. The EEG signals were sampled with a sampling rate of 250Hz and filtered by an $8 - 30$Hz bandpass filter.

Algorithms Evaluated: The BRLP was compared against four competing algorithms: LLE, Isomap, CSP and bilinear locality preserving (BLP). The ELM-TS-RE was compared against five competing classification algorithms: LDA-CSP, ELM-CSP, LDA-TS, ELM-TS and ELM-TS-E.

1) LLE: a classical dimensionality reduction algorithm based on locality preserving [19];
2) Isomap: a classical dimensionality reduction algorithm based on isometric mapping [36];
3) CSP: a benchmark feature extraction algorithm in motor imagery [6];
4) BLP: bilinear locality preserving on Riemannian graph, as shown in (11);
5) LDA-CSP: the CSP followed by LDA classifier [6];
6) ELM-CSP: the CSP followed by ELM classifier;

7) LDA-TS: LDA classifier performed on tangent space of Riemannian manifold [18];

8) ELM-TS: ELM classifier performed on tangent space of Riemannian manifold;

9) ELM-TS-E: ELM classifier performed on tangent space of embedding learned by BLP, without channel information;

Moreover, the proposed ELM-TS-RE algorithm also compared with the top 3 winner methods on dataset IIa of competition IV:

1) 1st winner: it used the filter bank CSP (FBCSP) for feature extraction and used the naive Bayes classifier for classification. FBCSP adaptively selected the feature from nine frequency sub-bands based on mutual information [37];

2) 2nd winner: it used standard CSP for feature extraction and Fisher's LDA for feature selection. The Bayesian classifier was used for classification. And the EEG signal was filtered by 8-30Hz bandpass filter [38];

3) 3rd winner: it used a recursive channel elimination method for channel selection and standard CSP for feature extraction. The ensemble SVM classifier was used for classification. And the EEG signal was filtered by 8-25Hz bandpass filter [38].

Parameters Setting: The number of nearest neighbors in LLE was set as 12 as suggested by [19]. The Isomap learned an embedding by $k = 6$ nearest neighbors [36]. The number of CSP spatial filter was set to be 8 as suggested in [39]. According to the parameter analysis in Experiment III, we set maximum iterative number $N_{ite} = 20$ and threshold $\tau = 10^{-3}$ as iterative stop criterion. In BRLP, we empirically set the number of selected filters $N_s = 6$, the number of nearest neighbor $k = 40$ and regularization parameter $\lambda = 0.15$. For the ELM-TS-RE algorithm, we set the number of hidden neuron $N_h = 2000$ and the balance factor $\theta = 0.001$ in ELM classifier. Moreover, the intrinsic dimensionality M in ELM-TS-RE was determined by 10-fold cross-validation, e.g., $\{10, 10, 10, 16, 10, 16, 10, 16, 16\}$ for the dataset IIa of BCI competition IV and $\{10, 10, 10, 10, 10, 14, 8, 8, 12\}$ for in-house datasets, respectively.

B. Results and Discussion

To assess the proposed methods, we designed three experiments in the following. Experiment I analyzed the implementation of BRLP algorithm. Experiment II evaluated the classification performance of the ELM-TS-RE algorithm. Experiment III discussed the convergence and parameter sensitivity for the proposed algorithms.

1) Results of Experiment I: In this experiment, in order to show how BRLP work, we first depicted the distribution of learned features by the BRLP, and then we showed the topographic maps of mapping matrix learned by the BRLP.

Fig. 4 shows the distribution of the two most discriminative features learned by BRLP, BLP, CSP, Isomap and LLE. The BRLP has larger between-class scatter and smaller within-class scatter against competing methods in left/right-hand motor imagery BCI. To obtain a more sophisticated

analysis, a discriminative index was further used to evaluate features in Fig. 4. The pointwise biserial correlation coefficient [40] was commonly used for discriminative index as

$$r = \frac{\sqrt{N_1 N_2}}{N_1 + N_2} \frac{mean\{f_i | l_i = 1\} - mean\{f_i | l_i = 2\}}{std\{f_i | l_i = 1, 2\}} \quad (15)$$

where N_1 and N_2 are the numbers of variables belong to the left-hand and right-hand classes, f_i and l_i are the value and class label of the i-th variable, respectively. r^2-value is equal to the square of r, and larger r^2-value indicates higher separability of features. In Fig. 5, the r^2-values of the two most discriminative features learned by BRLP are larger than the competing methods.
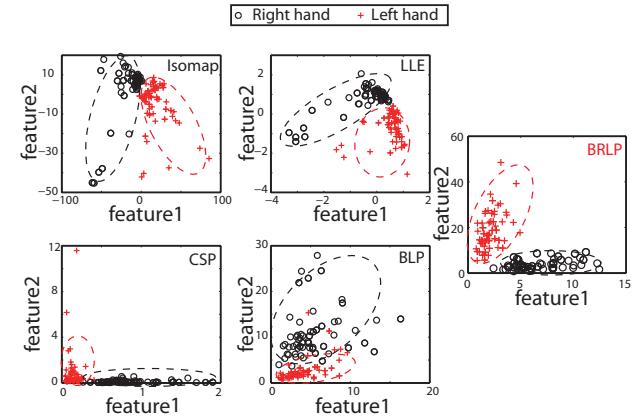
Figure 4. The distribution of two most discriminative features learned by Isomap, LLE, CSP, BLP and BRLP for the left/right hand motor imagery data of A01.
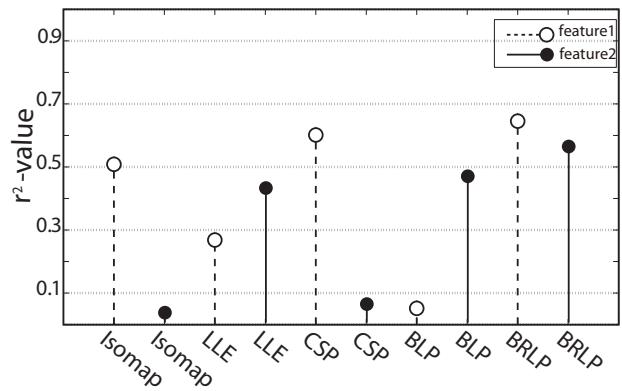
Figure 5. The r^2-values of the two most discriminative features learned by Isomap, LLE, CSP, BLP and BRLP, corresponding to Fig. 4.

Further, we also carried out an experiment to show the effect of test data on the performance of BRLP algorithm. In BRLP algorithm, the training and test data were used to construct a Riemannian graph. The local information of training and test data can be preserved in the learning of low-dimensional embedding. High performance was obtained by performing classification on the low-dimensional embedding with local information of training and test data. However, if the Riemannian graph was only constructed with the training data, the local information of the test data

will be ignored. It will easily result in low classification performance especially in the case of small training data. As shown in Fig. 6, whatever full or partial training data used, the features learned with training and test data have higher separability than the features only learned with training data.
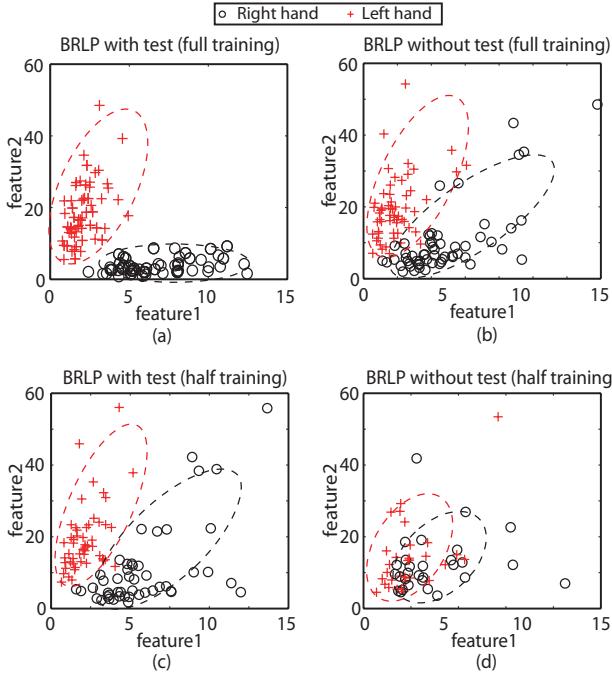
Figure 6. The distribution of two most discriminative features learned with and without test data by BRLP algorithm. a) BRLP with test data (full training data); b) BRLP without test data (full training data); c) BRLP with test data (half training data); d) BRLP without test data (half training data).

The mapping matrix of BRLP can be regarded as spatial filters for EEG signal. Fig. 7 shows the topographic maps of spatial filters learned by BRLP, BLP and CSP for the left/right hand motor imagery BCI. It is clear that large spatial filter coefficients of BRLP concentrate around the electrodes C_3 and C_4, which cover the area dedicated to the right-hand and left-hand imagery movements. In addition, comparing the spatial filters learned by BRLP and BLP in Fig. 7, we can observe that BRLP has higher spatial filter coefficients than BLP in C_3 and C_4, because the regularization term in BRLP utilizes the channel prior information from CSP filters.

2) Results of Experiment II: In this section, we further tested the classification performance of the proposed ELM-TS-RE algorithm on the dataset IIa of BCI competition IV and in-house datasets with 10-fold cross-validation procedure, since the cross-validation can evaluate the model in the training sets and provide insights on how the model will generalize an independent test sets. Moreover, the cross-validation was also used to determine the intrinsic dimensionality M in BRLP. In the 10-fold cross-validation procedure, the training set was partitioned into 10 subsets with equal size. In each run, 9 subsets were used for learning while a remaining subset was used for validation.

Fig. 8 (a) shows the results of cross-validation on the dataset IIa of BCI competition IV. Since dataset IIa was collected from a four-class experiment, the one-versus-one strategy was used to extend the binary classification of ELM-TS-RE to perform the classification. The Kappa coefficient was adopted to evaluate the classification performance, since the Kappa coefficient takes into account the misclassification of multi-class problem [41]. Fig. 8 (b) shows the results of cross-validation on in-house datasets. Since it is a two-class problem, for simplicity, the classification accuracy was used as performance measure for in-house datasets. As shown in Fig. 8 (a) and (b), the proposed ELM-TS-RE has higher performance than the five competing classification algorithms, e.g., ELM-TS-E, ELM-TS, LDA-TS, ELM-CSP and LDA-CSP. Especially for the comparison between ELM-TS-RE and ELM-TS-E, because the ELM-TS-RE utilized the channel prior information leads to higher performance than ELM-TS-E, we can infer that high performance of proposed ELM-TS-RE might be attributable in part to the ability of embedding learned by BRLP to exploit prior knowledge of channel.

An interesting result in Fig. 8 (a) is that the standard deviation values are relatively large compared with the difference between the mean values. In fact, the key reason of high variance is the significant difference between the subjects, where some subjects (S01, S03, S07, S08 and S09) can perform better motor imagery task and the others (S02, S04, S05 and S06) are poor in motor imagery. On the other hand, because all in-house subjects are well-training for motor-imagery, the ratio of standard deviation versus the difference between the mean values in Fig. 8 (b) are not as high as those in Fig. 8 (a). In addition, we also provided significance analysis for cross-validation results in Fig. 8. From the sign test results in Table I, it is clear that the differences between ELM-TS-RE and other competing methods are statistically significant.

Lastly, we compared the proposed ELM-TS-RE with the winner of the dataset IIa of BCI competition. The performances of the top 3 winners (1st, 2nd and 3rd) were included in the comparison. As shown in Table II, the ELM-TS-RE consistently outperformed the 2nd and 3rd winner methods for all the 9 subjects, when a bandpass filter (8-30Hz or 8-25Hz) was used for processing. A sign test revealed that the performance of the ELM-TS-RE method was significantly higher than 2nd winner method ($p < 0.005$) and 3rd winner method ($p < 0.005$). However, compared with the 1st winner method, the proposed ELM-TS-RE did not have higher performance for all subjects. There was no significant difference between the ELM-TS-RE and the 1st winner method ($p = 0.765$). The reason was that the different filters were used in processing for the 1st winner method. The 1st winner method used filter bank and different frequency information was selected for different subject. The proposed ELM-TS-RE only used a fixed 8-30Hz bandpass filter for all subjects.

For fair comparison, we proposed a filter bank ELM-TS-RE algorithm, which had same filter process as the 1st winner method. The graphical illustration of filter bank
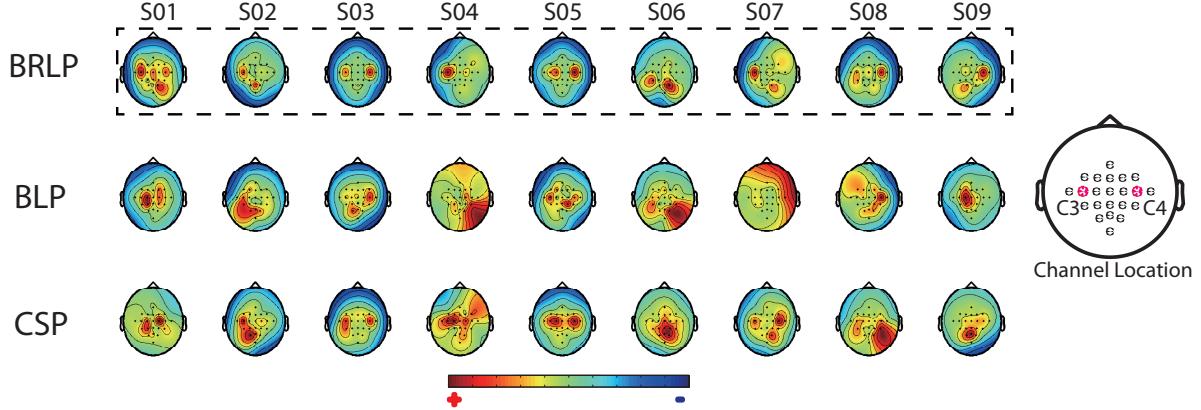
Figure 7. Topographic maps of the spatial filters learned by BRLP, BLP and CSP methods for the left/right hand motor imagery data from the 9 subjects of dataset IIa of BCI competition IV.
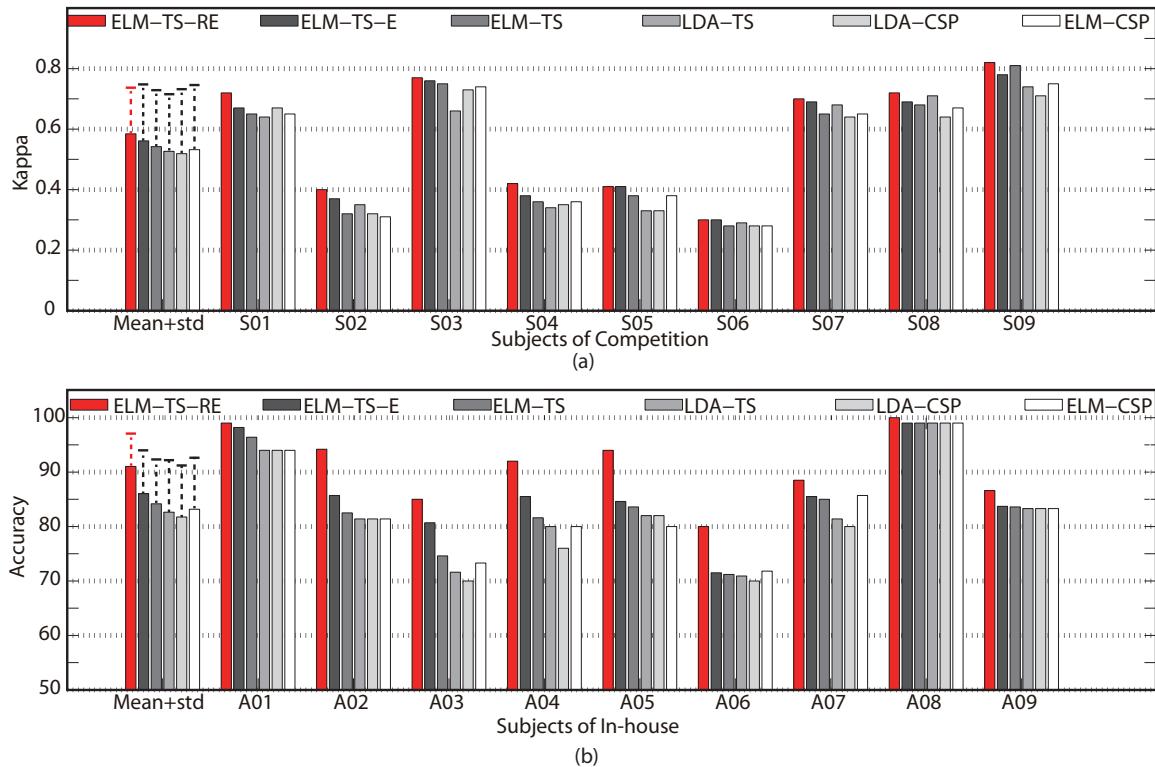
Figure 8. Comparison of classification performance of all studied algorithms on the motor imagery dataset via 10-fold cross-validation. a) dataset IIa of BCI competition IV; b) in-house datasets.

Table I
A SIGN TEST RESULTS FOR THE PROPOSED METHOD VERSUS COMPETING METHODS ON THE CROSS-VALIDATION RESULTS OF FIG. 8.

	Competition Dataset (Fig. 8 (a))	In-house dataset (Fig. 8 (b))
	p-value	p-value
ELM-TS-RE vs. ELM-TS-E	*	**
ELM-TS-RE vs. ELM-TS	†	**
ELM-TS-RE vs. LDA-TS	†	**
ELM-TS-RE vs. ELM-CSP	†	**
ELM-TS-RE vs. LDA-CSP	†	**

Note: $\sim$ nonsignificant, * $p \leq 0.05$, ** $p \leq 0.01$, † $p \leq 0.005$

Table II
COMPARISON OF THE KAPPA VALUES BETWEEN PROPOSED ALGORITHM AND WINNER RESULTS ON DATASET IIa OF BCI COMPETITION IV FOR
PREDICTION OF TEST DATA.

Method	Mean Kappa	subject								
		S01	S02	S03	S04	S05	S06	S07	S08	S09
Filter Bank ELM-TS-RE	**0.628**	**0.77**	**0.43**	**0.78**	**0.52**	**0.51**	**0.32**	**0.78**	**0.78**	**0.77**
ELM-TS-RE	**0.584**	**0.76**	**0.36**	**0.76**	**0.48**	**0.35**	**0.31**	**0.75**	**0.74**	**0.75**
1st	0.570	0.68	0.42	0.75	0.48	0.40	0.27	0.77	0.75	0.61
2nd	0.520	0.69	0.34	0.71	0.44	0.16	0.21	0.66	0.73	0.69
3rd	0.310	0.38	0.18	0.48	0.33	0.07	0.14	0.29	0.49	0.44

Table III
A SIGN TEST FOR THE RESULTS OF TABLE II.

	ELM-TS-RE	Filter Bank ELM-TS-RE
1st	$\sim$	†
2nd	†	†
3rd	†	†

Note: $\sim$ nonsignificant, * $p \leq 0.05$, ** $p \leq 0.01$, † $p \leq 0.005$

ELM-TS-RE was shown in Fig. 9. Compared with the ELM-TS-RE, the filter bank ELM-TS-RE added two steps in processing of EEG data: a filter bank comprising of multiple bandpass filters, and feature selection via mutual information based best individual feature (MIBIF) algorithm [37]. From the comparison results in Table II, it is clear that the filter bank ELM-TS-RE consistently outperformed the 1st winner method for all the 9 subjects after using same filter bank process. A sign test indicated that the performance improvement of the filter bank ELM-TS-RE was statistically significant ($p < 0.005$). However, the filter bank ELM-TS-RE was more computationally expensive compared to ELM-TS-RE, as it performed multiple BRLP algorithms corresponding to multiple frequency sub-bands. We showed the significant test results for the proposed methods versus winner methods in Table III. It can be seen that the proposed methods could obtain significantly higher performance than winner methods after using same filter process.
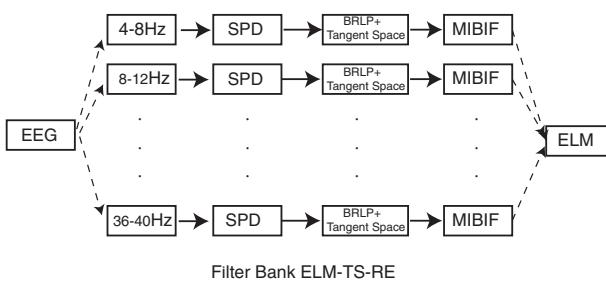
Figure 9. The graphical illustration of filter bank ELM-TS-RE.

3) Results of Experiment III: In this experiment, we studied the convergence and parameter setting problems of the proposed methods.

Fig. 10 shows the convergence of mapping matrix $\mathbf{V}_t$ learned from BRLP. We plotted nine convergence curves
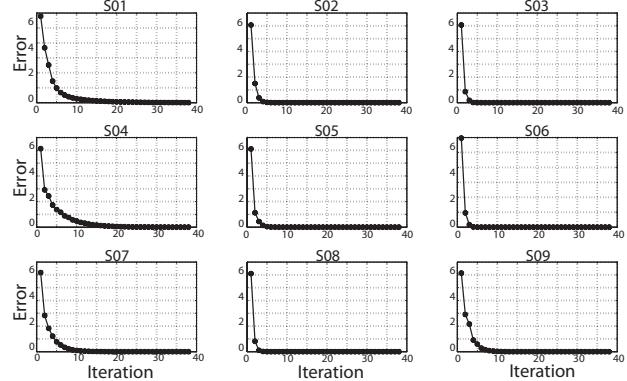
Figure 10. Convergence curves of mapping matrix obtained by executing the BRLP on dataset IIa of BCI competition IV within 40 iterations.

corresponding to nine subjects of dataset IIa of BCI competition IV. The error of convergence is defined as

$$Error = \| \mathbf{V}_t - \mathbf{V}_{t-1} \|_F^2 . \quad (16)$$

With increasing of the number of iteration steps, all the nine executions of BRLP algorithm converged within 20 iteration steps. To achieve a trade-off between performance and efficiency, the maximum iterative number N_{ite} was empirically set as 20 and threshold τ was empirically set as 10^{-3}.

Sensitivity analysis of the parameters, like the number of selected filters N_s and regularization parameter λ for BRLP, the number of hidden neuron N_h and balance factor θ for ELM-TS-RE, were also carried out in this section.

Firstly, we studied the effects of the parameters λ and N_s on the performance of BRLP. Fig. 11 shows the r^2-values of the most discriminative features learned by BRLP under the different λ and N_s. If λ is too large, the effects of locality preserving term in (12) will be weaken and the optimization (12) will focus more on the regularization
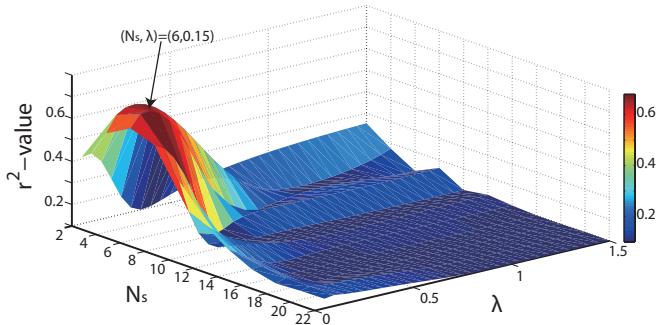
Figure 11. The r^2-values of the most discriminative features learned by BRLP under the different λ and N_s for subject S03.

term. In this case, the BRLP will extract features with low separability because of lacking local characteristics of Riemannian manifold. With the decreasing value of λ, the higher r^2-value is obtained while the number of selected filters N_s is correctly set. However, if the λ is too small, the optimization (12) will ignore the prior information of EEG channels contained in regularization term. It is clear that the correct setting of λ and N_s play an important role in application of BRLP.

We also provided analysis on how N_s affect the channel information in BRLP. Fig. 12 shows the average absolute channel weights $\frac{1}{N_s}\sum_{i=1}^{N_s}\left|\frac{\mathbf{w}_i}{\|\mathbf{w}_i\|}\right|$ versus N_s for the left/right hand motor imagery BCI. With a small value of N_s (like $N_s = 2, 4$), high weights on C_3 and C_4, which play a key role in the right-hand and left-hand imagery movements, could be observed. Unfortunately, we could find large weight on other inessential channel like C_{P3}. In large values of N_s ($N_s = 20, 22$) are set, although the weights on C_3 and C_4 channel will be smaller, the effect of inessential channel C_{P3} is reduced. Consequently, the values of N_s should be determined by achieving a trade-off between essential and inessential channels. As shown in the Fig. 11, we can obtain highest r^2-value with $(N_s, \lambda) = (6, 0.15)$.

Secondly, we analyzed the effects of the number of hidden neuron N_h and balance factor θ on the performance of ELM-TS-RE. Fig. 13 shows the accuracy of ELM-TS-RE versus N_h with different θ. The ELM network will suffer from over-fitting problem in the case of large number of hidden neurons. Appropriate θ should be selected to balance the loss function and L_2 regularization term in the ELM. As shown in Fig. 13, with the decreasing value of θ, the relative effect of L_2 regularization term in ELM increases and the over-fitting problem is alleviated. We can obtain highest performance of ELM on both training and testing set with $\theta = 0.001$. Furthermore, if θ is too small ($\theta = 0.0003$), the performance of ELM becomes lower because of under-fitting problem.

V. CONCLUSIONS

Some EEG channels are more important than others in the processing of motor imagery BCI. In this paper, we proposed a bilinear locality preserving mapping method
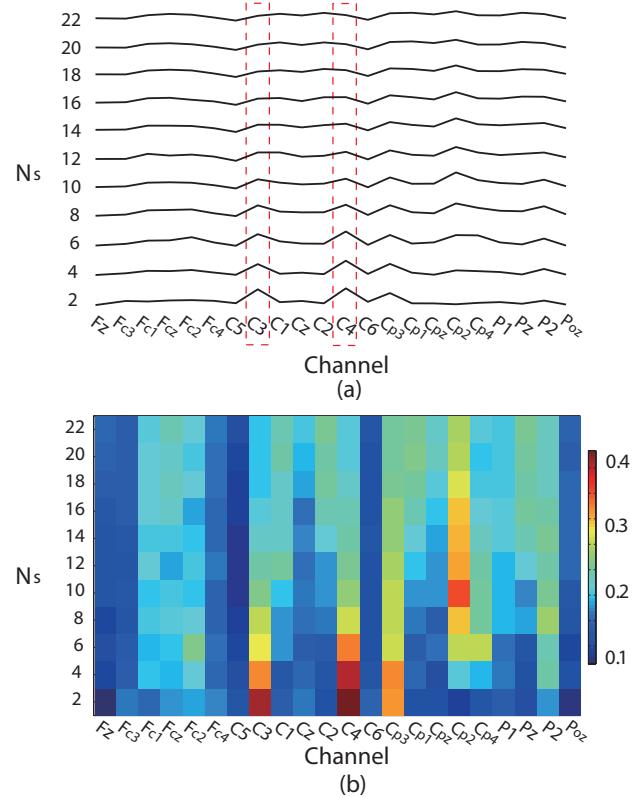
Figure 12. The effect of N_s on the channel weight learned from the left/right hand motor imagery data of subject S03. a) the wave of channel weight under different N_s; b) visualization of channel weight.
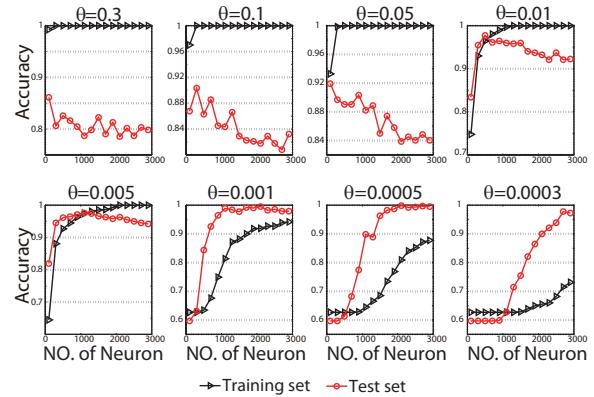
Figure 13. Classification accuracy of ELM-TS-RE versus number of neuron (N_h) under the different θ.

on the Riemannian graph with a channel prior information from CSP filters to learn a low-dimensional embedding. Furthermore, we also proposed a classification algorithm by executing ELM classifier on the tangent space of learned embedding. The experimental results on the dataset IIa of BCI competition IV and in-house datasets show the high performance of the proposed methods.

ACKNOWLEDGMENTS

REFERENCES

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

[2] K. P. Thomas, C. Guan, C. T. Lau, A. P. Vinod, and K. K. Ang, "A new discriminative common spatial pattern method for motor imagery brain computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 11, pp. 2730–2733, Nov 2009.

[3] X. Liao, D. Yao, D. Wu, and C. Li, "Combining spatial filters for the classification of single-trial EEG in a finger movement task," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 821–831, May 2007.

[4] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement," *Electroenc. Clin. Neurophys.*, vol. 103, no. 6, pp. 642–651, Dec. 1997.

[5] R. Zhang, D. Yao, P. Valdés-Sosa, F. Li, P. L. T. Zhang, T. Ma, Y. Li, and P. Xu, "Efficient resting-state EEG network facilitates motor imagery performance," *J Neural Eng*, vol. 12, no. 6, p. 066024, Dec. 2015.

[6] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Proc. Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2008.

[7] P. Xu, X. Xiong, Q. Xue, P. Li, R. Zhang, Z. Wang, P. A. Valdes-Sosa, Y. Wang, and D. Yao, "Differentiating between psychogenic nonepileptic seizures and epilepsy based on common spatial pattern of weighted EEG resting networks," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1747–1755, Jun. 2014.

[8] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selelction in motor imagery based brain-computer interface," in *Proc. 27th Annu. Int. Conf. Eng. Med. Biol. Soc.*, Jan. 2005, pp. 5392–5395.

[9] H. Lu, H. L. Eng, C. Guan, K. N.Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial pattern with aggregation for EEG classification in small-sample setting," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2936–2946, Dec. 2010.

[10] Z. Gu, Z. Yu, Z. Shen, and Y. Li, "An online semi-supervised brain-computer interface," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 9, pp. 2614–2623, Sept. 2013.

[11] O. Chapelle, B. Schölkopf, A. Zien *et al.*, *Semi-supervised learning*. Cambridge, MA: MIT press Cambridge, 2006, ch. 1, pp. 1–13.

[12] Y. Li, C. Guan, H. Li, and Z. Chin, "A self-training semi-supervised SVM algorithm and its application in an EEG-based brain computer interface speller system," *Pattern Recogn. Lett.*, vol. 29, no. 9, pp. 1285–1294, Jul. 2008.

[13] J. Long, Y. Li, and Z. Yu, "A semi-supervised support vector machine approach for parameter setting in motor imagery-based brain computer interfaces," *Cogn. Neurodyn.*, vol. 4, no. 3, pp. 207–216, Sep. 2010.

[14] Y. Li and C. Guan, "An extended em algorithm for joint feature extraction and classification in brain-computer interfaces," *Neural Comput.*, vol. 18, no. 11, pp. 2730–2761, Sep. 2006.

[15] H. Wang and D. Xu, "Comprehensive common spatial patterns with temporal structure information of EEG data: Minimizing nontask related EEG component," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2496–2505, Sept. 2012.

[16] B. Cheng, J. Yang, S. Yan, Y. Fu, and T. S. Huang, "Learning with L1-graph for image analysis," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 858–866, Dec. 2010.

[17] S. Yan and H. Wang, "Semi-supervised learning by sparse representation." in *SIAM Int. Conf. on Data Mining (SDM)*, Nevada, USA, Apr. 2009, pp. 792–801.

[18] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain–computer interface classification by Riemannian geometry," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 4, pp. 920–928, Apr. 2012.

[19] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[20] X. He and P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing Systems 16*, S. Thrun, L. Saul, and B. Schölkopf, Eds. MIT Press, 2004, pp. 153–160.

[21] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.

[22] K. Yu, K. Shen, S. Shao, W. C. Ng, and X. Li, "Bilinear common spatial pattern for single-trial ERP-based rapid serial visual presentation triage," *J Neural Eng*, vol. 9, no. 4, p. 046013, Aug. 2012.

[23] Y. Zhang, G. Zhou, Q. Zhao, J. Jin, X. Wang, and A. Cichocki, "Spatial-temporal discriminant analysis for ERP-based brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 2, pp. 233–243, Mar. 2013.

[24] W. Wu, X. Gao, B. Hong, and S. Gao, "Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL)," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 6, pp. 1733–1743, Jun. 2008.

[25] J. Meng, G. Liu, G. Huang, and X. Zhu, "Automated selecting subset of channels based on CSP in motor imagery brain-computer interface system," in *Proc. IEEE Int. Conf. Rob. Biomimetics*, 2009, pp. 2290–2294.

[26] 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, Feb. 2011.

[27] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on Riemannian manifolds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1713–1727, Oct. 2008.

[28] X. Xie, Z. L. Yu, H. Lu, Z. Gu, and Y. Li, "Motor imagery classification based on bilinear sub-manifold learning of symmetric positive-definite matrices," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 504–516, Jun. 2017.

[29] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, Dec. 2006.

[30] W. Förstner and B. Moonen, "A metric for covariance matrices," in *Geodesy-The Challenge of the 3rd Millennium*, E. W. Grafarend, F. W. Krumm, and V. S. Schwarze, Eds. Berlin:Springer-Verlag, 2003, pp. 299–309.

[31] M. Moakher, "A differential geometric approach to the geometric mean of symmetric positive-definite matrices," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 3, pp. 735–747, Jul. 2005.

[32] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.

[33] X. He, D. Cai, and P. Niyogi, "Tensor subspace analysis," in *Advances in Neural Information Processing Systems 18*, Y. Weiss, B. Schölkopf, and J. Platt, Eds. MIT Press, 2006, pp. 499–506.

[34] Y. Liu, Y. Liu, and K. C. Chan, "Multilinear isometric embedding for visual pattern analysis," in *IEEE 12th Int. Conf. Computer Vision Workshops (ICCV)*, Kyoto, Sept. 2009, pp. 212–218.

[35] G.-B. Huang, L. Chen, and C.-K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Trans. neural networ.*, vol. 17, no. 4, pp. 879–892, Jul. 2006.

[36] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[37] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *IEEE International Joint Conference on Neural Networks, 2008.* IEEE, 2008, pp. 2390–2397.

[38] M. Tangermann, M. K. Robert, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. Miller, G. Mller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the bci competition iv," *Front Neurosci*, no. Jul, 8 2012.

[39] M. Grosse-Wentrup and M. Buss, "Multiclass common spatial patterns and information theoretic feature extraction," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1991–2000, Aug. 2008.

[40] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K. R. Müller, "Single-trial analysis and classification of ERP components - A tutorial," *NeuroImage*, vol. 56, no. 2, pp. 814 – 825, May 2011.

[41] G. Townsend, B. Graimann, and G. Pfurtscheller, "A comparison of common spatial patterns with complex band power features in

a four-class BCI experiment," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 4, pp. 642–651, Apr. 2006.

Xiaofeng Xie received the B.S. degree in automation from South China University of Technology, Guangzhou, China, in 2011. He is currently working toward the Ph.D. degree in pattern recognition and intelligent systems at the South China University of Technology, Guangzhou, China. His research interests include the pattern recognition and signal processing.

Zhu Liang YU received his BSEE in 1995 and MSEE in 1998, both in electronic engineering from the Nanjing University of Aeronautics and Astronautics, China. He received his Ph.D. in 2006 from Nanyang Technological University, Singapore. He joined Center for Signal Processing, Nanyang Technological University from 2000 as a research engineer, then as a Group Leader from 2001. In 2008, he joined the College of Automation Science and Engineering, South China University of Technology and was promoted to be a full professor in 2011. His research interests include signal processing, pattern recognition, machine learning and their applications in communications, biomedical engineering, etc.

Zhenghui GU received the Ph.D. degree from Nanyang Technological University in 2003. From 2002 to 2008, she was with Institute for Infocomm Research, Singapore. She joined the College of Automation Science and Engineering, South China University of Technology, in 2009 as an associate professor. She was promoted to be a full professor in 2015. Her research interests include the fields of signal processing and pattern recognition.

Jun Zhang received the Bachelors and Masters degrees in computer science from the Xiangtan University, Xiangtan, China, in 2002 and 2005, respectively, and the doctoral degree in pattern recognition and intelligence system from the South China University of Technological, Guangzhou, China, in 2012. He was a postdoctoral research fellow in electronic engineering with University of South California, Los Angeles, USA, from 2015 to 2016. He is currently an associate professor with the School of Information Engineering, Guangdong University of Technology, Guangzhou, China, where he has been the Head of the Electrical Engineering Department since March 2016. His current research interests include the fields of compressive sensing, machine learning and biomedical signal processing.

Ling Cen Dr. Ling Cen received her Ph.D. degree in Electrical and Computer Engineering, from the National University of Singapore (NUS). She is currently a professor in college of Information Engineering, Guangdong University of Technology, China. Her research interests are in the areas of optimization, pattern recognition, machine learning, and their applications in science and engineering. She has published over 60 articles in refereed international journals and conferences, including 4 book chapters, and filed 3 international patent applications.

Yuanqing Li was born in Hunan Province, China, in 1966. He received the B.S. degree in applied mathematics from Wuhan University, Wuhan, China, in 1988, the M.S. degree in applied mathematics from South China Normal University, Guangzhou, China, in 1994, and the Ph.D. degree in control theory and applications from South China University of Technology, Guangzhou, China, in 1997. Since 1997, he has been with South China University of Technology, where he became a full professor in 2004. In 2002-04, he worked at the Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Saitama, Japan, as a researcher. In 2004-08, he worked at the Laboratory for Neural Signal Processing, Institute for Inforcomm Research, Singapore, as a research scientist.

His research interests include, blind signal processing, sparse representation, machine learning, brainCcomputer interface, EEG and fMRI data analysis. He is the author or coauthor of more than 60 scientific papers in journals and conference proceedings.