Amplitude-Phase Information Measurement on Riemannian Manifold for Motor Imagery-Based BCI

Shoulin Huang, Guoqing Cai, Tong Wang, and Ting Ma

***Abstract*—Phase synchronization phenomena are directly connected with the underlying neural mechanisms of certain cognitive processes. However, only the amplitude information is utilized in most electroencephalogram (EEG)-based brain–computer interfaces (BCIs). Few of the existing methods can simultaneously measure the amplitude and phase information required for classification. In this study, a novel common amplitude-phase measurement (CAPM) method is proposed. This method is capable of jointly measuring the phase and amplitude information of EEG signals on the Riemannian manifold. The proposed CAPM method comprises a two-step approach. First, a novel Riemannian graph embedding is proposed for dimensionality reduction while performing spatial-spectral filtering. The graph embedding is excellent in capturing the intrinsic features contained by the physiological signal. Second, to enhance robustness, a novel classifier is designed to incorporate the regularized linear regression in the computation of Riemannian distance. Experimental results on two BCI competition datasets demonstrate CAPM can yield high classification performance. The proposed CAPM method is a promising tool in analyzing EEG amplitude-phase characteristics and exhibits great potential in BCI applications.**

***Index Terms*—Brain–computer interface (BCI), amplitude-phase measure, Riemannian manifold, graph embedding, common spatial pattern (CSP).**

I. Introduction

SURFACE electroencephalogram (EEG) is widely applied in Brain–computer interfaces (BCIs) [1] because of its non-invasive and low cost. EEG signals modulated via motor imagery (MI) exhibit amplitude fluctuation over the sensorimotor cortex [1], [2]. This amplitude change can be detected using the popular common spatial pattern (CSP) [3], which is proven to be a highly efficient algorithm. Till date, a great number of CSP variants, such as composite CSP [4], divergence measure-based CSP [5], [6], probabilistic CSP [7] and spatial-spectral filtering-based CSP [8]–[10], have been proposed and exhibited high performance.

While amplitude information is the most often used feature in MI-based BCIs, the phase receives scarce attention. Many studies have indicated that phase synchronization reflects the underlying neural mechanisms of cognitive processes [11], [12]. Some studies have been proposed to use phase-related features for further improving the system performance [13], [14]. However, in these studies, the amplitude and phase are often measured separately. Every EEG signal can be described by its amplitude and instantaneous phase with the oscillation frequency [11]. As for both amplitude and phase contain important information related to neural activity, their joint measurement is supposed to help capture more intrinsic signal characteristics, thereby leading to a high BCI performance. To this purpose, advanced complex-valued signal processing methods are required for representations of amplitude and phase. Recently, Chakraborty *et al.* [15] proposed the application of CSP technique on phase and amplitude of EEG signals (APCSP), and further employed nonlinear principal component analysis (NLPCA) and auto-associative neural network (AANN) to handle the complex data model. Inspired by their study, we propose a novel common amplitude-phase measurement (CAPM) method, to decode complex-valued covariance matrices containing both amplitude and phase information of EEG signals. The covariance matrix of each trial can be treated as a point on the Riemannian manifold of symmetric and positive definite (SPD) matrices [16]. Thus, different from the APCSP method, our proposed method is fundamentally and globally based on Riemannian geometry, which can characterize the intrinsic geometrical relations among samples more accurately [16], [17].

The proposed CAPM method includes a two-step approach: a novel Riemannian graph embedding for dimensionality reduction and a regularized Riemannian distance for classification. To deal with the "curse of dimensionality," graph embedding [18] for Riemannian manifold in BCIs has recently received significant interest [19]–[21]. In the second step, the relation between a trial covariance and a class-related covariance is measured by their Riemannian distance [22] on the SPD manifold. The innovative CAPM method we proposed has contributions as listed: 1) A novel Riemannian geometry-based solution is proposed to simultaneously process amplitude and phase information of EEG signals. To our knowledge, the method is the first to provide a unified methodology for complex signal processing in BCIs. 2) The proposed Riemannian graph embedding is equivalent to a novel spatial-spectral filtering approach. Because EEG rhythmic activities are unequally distributed over the cortex and frequency

Manuscript received February 16, 2021; revised April 19, 2021; accepted May 24, 2021. Date of publication June 7, 2021; date of current version July 12, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC1312000, in part by the National Natural Science Foundation of China under Grant 61801145, in part by the Stability Support Program for Universities of Shenzhen under Grant GXWD20201230155427003-20200822115709001, and in part by the Shenzhen Science and Technology Program under Grants JCYJ20180306171800589 and KQTD20190929172545139. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Xun Chen. *(Corresponding author: Ting Ma.)*

Shoulin Huang, Guoqing Cai, and Tong Wang are with the School of Electronics and Information Engineering, Harbin Institute of Technology, Shenzhen 518055, China (e-mail: hslhitsz@outlook.com; cgq982023711@163.COM; tong.wang@umic.rwth-aachen.de).

Ting Ma is with the School of Electronics and Information Engineering, Harbin Institute of Technology, Shenzhen 518055, China, and also with the Advanced Innovation Center for Human Brain Protection, Capital Medical University, Beijing 100070, China, and also with the Peng Cheng Laboratory, Shenzhen, Guangdong 518066, China (e-mail: tmahit@outlook.com).

Digital Object Identifier 10.1109/LSP.2021.3087099

band, the combined optimization of spatial and spectral filters can efficiently capture the discriminative features [8], [9]. 3) A novel classifier incorporating the regularized linear regression in the computation of Riemannian distance is proposed to enhance the robustness.

II. Methodology

A. Phase-Related Trial Representation

We use wavelet transform to obtain the natural amplitude-phase representation of EEG signals firstly. Let $\mathbf{X} = [\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_C(t)]^{\mathrm{T}} \in \mathbb{R}^{C\times T}$ be a single trial EEG signal, where C and T are the numbers of channels and sampled points respectively, and $\mathbf{x}_c(t)$ is the cth channel signal. The amplitude-phase contents of $\mathbf{x}_c(t)$ at frequency f and time t are given by the convolution of $\mathbf{x}_c(t)$ with a complex Morlet wavelet [23]:

$$\hat{\mathbf{x}}_c(t,f) = \mathbf{x}_c(t) * \boldsymbol{\psi}(t,f) = \mathbf{a}_c(f,t)e^{i\boldsymbol{\theta}_c(f,t)} \tag{1}$$

where $\boldsymbol{\psi}(t,f)$ is the wavelet function defined by $\boldsymbol{\psi}(t,f) = \frac{1}{\sqrt{\pi f_b}}\exp(i2\pi ft)\exp(-\frac{t^2}{f_b})$, with f_b being the bandwidth parameter. In this study, f_b is set to 1 as suggested in [17]. If there is K observed frequency bins, a $K \times T$ complex-valued matrix $\mathbf{X}_c$ is formed with the row vectors $\hat{\mathbf{x}}_c(t,f_k)$ ($k = 1,2,\ldots,K$). Then, an amplitude-phase representation is given by $\hat{\mathbf{X}} = [\mathbf{X}_1^{\mathrm{T}}, \mathbf{X}_2^{\mathrm{T}}, \ldots, \mathbf{X}_C^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{KC\times T}$ for a trial $\mathbf{X}$. Similarly, an amplitude representation $\mathbf{A} \in \mathbb{R}^{KC\times T}$ or a pure phase representation $\boldsymbol{\Theta} \in \mathbb{R}^{KC\times T}$ can be obtained from $\boldsymbol{a}_c(f_k,t)$ or $\boldsymbol{\theta}_c(f_k,t)$, where

$$\begin{cases} \boldsymbol{a}_c(f_k,t) = \|\hat{\mathbf{x}}_c(t,f_k)\| \\ \boldsymbol{\theta}_c(f_k,t) = \arctan\frac{Im[\hat{\mathbf{x}}_c(t,f_k)]}{Re[\hat{\mathbf{x}}_c(t,f_k)]} \end{cases}, \; k = 1\ldots K, c = 1\ldots C$$

where $\|\cdot\|$ denotes L_2 Frobenius norm, and $Re[\mathbf{x}]$ or $Im[\mathbf{x}]$ returns the real or imaginary part of $\mathbf{x}$.

B. Riemannian Graph Embedding

We now describe a novel graph embedding algorithm to project a high-dimensional manifold into a more discriminative, low-dimensional representation. We assume the training set $\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{C}}_{-1} \cup \boldsymbol{\mathcal{C}}_{+1}$, where $\boldsymbol{\mathcal{C}}_z$ is the subset belonging to a class $z \in \{-1,+1\}$. Given N as the total trial number, we obtain the nth phase-related trial as $\mathbf{Y}_n \in \mathbb{C}^{L\times T}$, where L is equal to KC. The covariance matrix of $\mathbf{Y}_n$ can be given by $\mathbf{P}_n = \frac{1}{T-1}\mathbf{Y}_n\mathbf{Y}_n^{\mathrm{H}}$, where H denotes the operator of conjugate transpose. Please note that the covariance matrix can be a data point in the SPD manifold.

We first construct a graph $\boldsymbol{\mathcal{G}}$ with N vertices. Let $\mathbf{S} \in \mathbb{R}^{N\times N}$ be the adjacency matrix, whose element s_{nr} denotes the weight of the edge between two vertices $\mathbf{Y}_n$ and $\mathbf{Y}_r$. To preserve the local structure in training set, we define s_{nr} as:

$$s_{nr} = \begin{cases} \exp(-\frac{\delta_{nr}}{\sigma}), & \text{if } \mathbf{Y}_n \in \boldsymbol{\mathcal{C}}_z(\mathbf{Y}_r), \\ & \text{and } \mathbf{Y}_r \in \boldsymbol{\mathcal{C}}_z(\mathbf{Y}_n) \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

In (2), $\boldsymbol{\mathcal{C}}_z(\mathbf{Y}_n)$ is the set of p nearest neighbors of $\mathbf{Y}_n$ which share the same label z with $\mathbf{Y}_n$, and parameter σ is a positive constant. Given $\mathbf{P}_n$ and $\mathbf{P}_r$ respectively as the covariance matrices of $\mathbf{Y}_n$ and $\mathbf{Y}_r$, δ_{nr} denotes the Riemannian distance [22] between $\mathbf{P}_n$ and $\mathbf{P}_r$ as

$$\delta_{nr} = \delta_R(\mathbf{P}_n,\mathbf{P}_r) = \|\log({\mathbf{P}_n}^{-1}\mathbf{P}_r)\| = \left(\sum_{l=1}^{L}\log^2\beta_l\right)^{\frac{1}{2}} \tag{3}$$

where β_l, $l = 1,2,\ldots,L$ are the real eigenvalues of ${\mathbf{P}_n}^{-1}\mathbf{P}_r$. The graph $\boldsymbol{\mathcal{G}}$ is finally obtained to preserve the local structure of same class data on Riemannian manifold.

Next we aim to find a transformation $\mathbf{W} \in \mathbb{C}^{L\times M}$ ($M < L$) such that $\mathbf{Y}_n$ can be mapped into a low-dimensional data $\hat{\mathbf{Y}}_n \in \mathbb{C}^{M\times T}$ by $\hat{\mathbf{Y}}_n = \mathbf{W}^{\mathrm{H}}\mathbf{Y}_n$. The graph embedding is expected to utilize both the local structure and discriminative information preserved in training set. Thus the objective function is defined as follows:

$$\max_{\mathbf{W}} \frac{\mathbf{Q} - \sum_{n,r}\|\hat{\mathbf{Y}}_n - \hat{\mathbf{Y}}_r\|^2 s_{nr}}{\sum_n \|\hat{\mathbf{Y}}_n\|^2} \tag{4}$$

where $\mathbf{Q}$ is defined as

$$\mathbf{Q} = \sum_{n\in\boldsymbol{\mathcal{C}}_{+1}}\|\hat{\mathbf{Y}}_n\|^2 - \sum_{n\in\boldsymbol{\mathcal{C}}_{-1}}\|\hat{\mathbf{Y}}_n\|^2 \tag{5}$$

The matrix $\mathbf{Q}$ aims to ensure discrimination, and the denominator in (4) is used to preserve the global structure [24]. Besides, the objective function also attempts to ensure that if two points $\mathbf{Y}_n$ and $\mathbf{Y}_r$ are "close" and share same class label then their map data $\hat{\mathbf{Y}}_n$ and $\hat{\mathbf{Y}}_r$ are "close" as well. Since $\|\mathbf{B}\|^2 = tr(\mathbf{B}\mathbf{B}^{\mathrm{H}})$ and $\hat{\mathbf{Y}}_n = \mathbf{W}^{\mathrm{H}}\mathbf{Y}_n$, we see that

$$\sum_{n,r}\|\hat{\mathbf{Y}}_n - \hat{\mathbf{Y}}_r\|^2 s_{nr} = \sum_{n,r} tr((\hat{\mathbf{Y}}_n - \hat{\mathbf{Y}}_r)(\hat{\mathbf{Y}}_n - \hat{\mathbf{Y}}_r)^{\mathrm{H}})s_{nr}$$

$$= tr\left(\mathbf{W}^{\mathrm{H}}\sum_{n,r}(\mathbf{Y}_n - \mathbf{Y}_r)(\mathbf{Y}_n - \mathbf{Y}_r)^{\mathrm{H}}s_{nr}\mathbf{W}\right)$$

Thus, (4) can be rewritten as

$$\max_{\mathbf{W}} \frac{tr(\mathbf{W}^{\mathrm{H}}(\mathbf{R}_{+1} - \mathbf{R}_{-1} - \mathbf{V})\mathbf{W})}{tr(\mathbf{W}^{\mathrm{H}}(\mathbf{R}_{+1} + \mathbf{R}_{-1})\mathbf{W})} \tag{6}$$

where $\mathbf{R}_z$, $z \in \{-1,+1\}$ and $\mathbf{V}$ are respectively defined as

$$\mathbf{R}_z = \sum_{n\in\boldsymbol{\mathcal{C}}_z}\mathbf{Y}_n\mathbf{Y}_n^{\mathrm{H}} = (T-1)\sum_{n\in\boldsymbol{\mathcal{C}}_z}\mathbf{P}_n \tag{7}$$

$$\mathbf{V} = \sum_{n,r}(\mathbf{Y}_n - \mathbf{Y}_r)(\mathbf{Y}_n - \mathbf{Y}_r)^{\mathrm{H}}s_{nr} \tag{8}$$

The solution to the problem (6) can be given by the eigenvectors corresponding to the largest M eigenvalues of the generalized problem $(\mathbf{R}_{+1} - \mathbf{R}_{-1} - \mathbf{V})\mathbf{w} = \lambda(\mathbf{R}_{+1} + \mathbf{R}_{-1})\mathbf{w}$. It should be noted that our graph embedding achieves the combined optimization of spatial and spectral filters. Specifically, the mth column $\mathbf{w}_m \in \mathbb{C}^L$ of $\mathbf{W}$ denotes a spatial-spectral filter consisted by a spatial filter $\boldsymbol{\gamma} = [\gamma_1,\gamma_2,\ldots,\gamma_C] \in \mathbb{R}^C$ and a set of spectral filters $\hat{\mathbf{w}}_c \in \mathbb{C}^K (c = 1\ldots C)$, as follows

$$\mathbf{w}_m = [\gamma_1\hat{\mathbf{w}}_1^{\mathrm{T}}, \gamma_2\hat{\mathbf{w}}_2^{\mathrm{T}}, \ldots, \gamma_C\hat{\mathbf{w}}_C^{\mathrm{T}}]^{\mathrm{T}} \tag{9}$$

$$\hat{\mathbf{w}}_c = \frac{1}{\gamma_c}[w_{m,(c-1)K+1}, w_{m,(c-1)K+2}, \ldots, w_{m,cK}]^{\mathrm{T}} \tag{10}$$

Algorithm 1: The Proposed CAPM in Training Stage.

Input: Given the training set $\mathcal{C}$ with trials $\{\mathbf{X}^{(n)} \in \mathbb{R}^{C\times T}, n = 1\ldots N$ and corresponding class labels $\mathbf{z} \in \mathbb{R}^N$;
Output: Projection matrix $\mathbf{W} \in \mathbb{C}^{L\times M}$; classification parameter $\mathbf{b} \in \mathbb{R}^{M\times 1}$;
1: Calculate N phase-related representations $\mathbf{Y}_n \in \mathbb{R}^{L\times T}$;
2: Calculate covariance matrices $\mathbf{P}_n, n = 1\ldots N$;
3: Construct the graph $\mathcal{G}$ according to (2);
4: Obtain the optimal mapping matrix $\mathbf{W}$ by solving (6);
5: Calculate covariance matrices of the mapping data by $\hat{\mathbf{P}}_n = \mathbf{W}^{\mathrm{H}}\mathbf{P}_n\mathbf{W}, n = 1\ldots N$;
6: Calculate mean covariance matrices $\hat{\mathbf{P}}_{-1}$ and $\hat{\mathbf{P}}_{+1}$;
7: Obtain the matrix $\mathbf{D}$ by computing $d_{n,m}$ according to (13);
8: Train classification parameter $\mathbf{b}$ by solving (14);

Algorithm 2: The Proposed CAPM in Evalution Stage.

Intput: Unlabeled trial $\mathbf{X}^{(test)} \in \mathbb{R}^{C\times T}$; projection matrix $\mathbf{W}$; classification parameter $\mathbf{b}$;
Output: The predicted class label z_{test};
1: Calculate its phase-related representation $\mathbf{Y}_{test}$;
2: Calculate covariance matrix $\mathbf{P}_{test}$ of $\mathbf{Y}_{test}$;
3: Calculate projection covariance by $\hat{\mathbf{P}}_{test} = \mathbf{W}^{\mathrm{H}}\mathbf{P}_{test}\mathbf{W}$;
4: Obtain a vector $\mathbf{d}_{test}$ according to (13);
5: $z_{test} = \mathrm{sign}(\mathbf{d}_{test}\mathbf{b})$;

where $w_{m,\hat{k}}$ is the $\hat{k}$th element of $\mathbf{w}_m$, and $\hat{\mathbf{w}}_c$ denotes the spectral weights for the cth channel at the K observed frequency bins. The decomposition into the spatial filter and spectral filters is not unique. Similar to the common spatio-spectral pattern (CSSP) algorithm [8], an intuitive solution of $\boldsymbol{\gamma}$ can be given by

$$\gamma_c = \mathrm{sign}(w_{m,(c-1)K+1}) \left(\sum_{\hat{k}=1}^{K} w_{m,(c-1)K+\hat{k}}^2 \right)^{0.5} \tag{11}$$

where sign(x) returns -1 if $x < 0$, otherwise it returns 1.

C. Regularized Riemannian Distance for Classification

For a learned low-dimensional data $\hat{\mathbf{Y}}_n$, its covariance matrix $\hat{\mathbf{P}}_n$ is given by $\mathbf{W}^{\mathrm{H}}\mathbf{P}_n\mathbf{W}$. The class label z_n corresponding to $\hat{\mathbf{Y}}_n$ can be predicted by the minimum distance to riemannian mean (MDRM) [22] algorithm:

$$z_n = \mathrm{sign}[\delta_R(\hat{\mathbf{P}}_{-1}, \hat{\mathbf{P}}_n) - \delta_R(\hat{\mathbf{P}}_{+1}, \hat{\mathbf{P}}_n)] \tag{12}$$

where $\hat{\mathbf{P}}_z$ is the arithmetic mean covariance matrices of EEG signals belonging to class z. According to (3), we see that:

$$z_n = \mathrm{sign}\left[\left(\sum_{m=1}^{M} \log^2 \eta_m \right)^{0.5} - \left(\sum_{m=1}^{M} \log^2 \rho_m \right)^{0.5} \right]$$

where η_m and ρ_m are the mth real eigenvalues of $\hat{\mathbf{P}}_{-1}^{-1}\hat{\mathbf{P}}_n$ and $\hat{\mathbf{P}}_{+1}^{-1}\hat{\mathbf{P}}_n$, respectively. Let denote $\mathbf{d}_n \in \mathbb{R}^{1\times M}$ a vector with the mth element $d_{n,m}$ as

$$d_{n,m} = \log^2 \eta_m - \log^2 \rho_m,\ m = 1, 2, \ldots, M \tag{13}$$

We can find that (12) is equivalent to $\mathrm{sign}(\sum_{m=1}^{M} d_{n,m})$. Moreover, if $\log^2 \eta_m$ is close to $\log^2 \rho_m$, the element $d_{n,m}$ is non-significant in determining the label value of z_n. Thus, we assign each $d_{n,m}$ a weight and define a new classification rule as $\mathrm{sign}(\phi(\mathbf{d}_n))$, where $\phi(\mathbf{d}_n) = \mathbf{d}_n\mathbf{b}$, $\mathbf{b} \in \mathbb{R}^{M\times 1}$. Consequently, the designed objective is converted to optimize the parameter $\mathbf{b}$. In the training stage, we organize N inputs $\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_N$ by a matrix $\mathbf{D} \in \mathbb{R}^{N\times M}$ and let $\mathbf{z} \in \mathbb{R}^{N\times 1}$ be the corresponding label vector. Then we can learn the parameter $\mathbf{b}$ by minimizing the classification error. Moreover, we expect to make $\mathbf{b}$ have some amount of zero components for robust learning. Thus, we formulate the optimization objective function as a popular regularized linear regression model

$$\min_{\mathbf{b}} \frac{1}{2}\|\mathbf{z} - \mathbf{D}\mathbf{b}\|^2 + \frac{\lambda(1-\alpha)}{2}\|\mathbf{b}\|^2 + \lambda\alpha\|\mathbf{b}\|_1 \tag{14}$$

where λ and α are the tuning parameters, and $\|\mathbf{b}\|_1$ is the L_1-norm. Due to the nature of the L_1 penalty, some components of $\mathbf{b}$ will be shrunk to exact zero if $\lambda\alpha$ is large enough. The L_2 penalty is used for smoothness. Obviously, (14) can be viewed as the elastic net optimization problem [25] to solve. The pseudo-code of the proposed CAPM method is given in **Algorithms 1**, **2**.

III. Experiments and Results

A. Experimental Setup

1) EEG Data Sets: We evaluated the proposed CAPM method on two datasets. (i) Data set IVa of BCI Competition III [26]: the EEG data were recorded from five subjects with a sampling rate of 100 Hz. The EEG segment in each trial was extracted from 0.5–3.5 s after the cue. Ten electrodes were used: FC3, C5, C3, C1, CP3, FC4, C2, C4, C6, CP4. These electrodes are widely applied in MI-based BCIs. This dataset is mainly used to check whether our method can effectively capture the discriminative information. (ii) Dataset 2b of BCI Competition IV [27]: the objective is to confirm whether our method can surpass the high-performance methods in terms of classification performance. The EEG data were sampled with 250 Hz from nine subjects at electrodes C3, Cz, and C4. For a fair comparison, we adopted the same time segments as the best winner [28] in the competition.

The EEG signals were filtered by a 50th-order finite impulse response (FIR) linear phase filter in band 7–30 Hz, which covers the main oscillation frequency bands of MI [2]. This filter can avoid phase distortion of the signals after filtering.

2) Comparative Algorithms: The proposed CAPM method was performed on the amplitude, phase, and amplitude-phase representations, respectively. We refer to them as CAPM-A, CAPM-P, and CAPM-AP. Our method was compared with the CSP-L, CSP-F, APCSP [15], APCSP-M and APCSP-F algorithms. The CSP-L denotes CSP combined with a linear discriminant analysis (LDA) classifier. The APCSP-M is a modified version of APCSP, where the MDRM classifier replaced NLPCA and AANN in the original APCSP. The only difference between CSP-F/APCSP-F and CSP-L/APCSP-M is that the former was applied on the data filtered with a passband obtained by cross-validation (CV). Besides, we provide the results of the top three algorithms on Data sets 2b of BCI Competition IV [29] for comparisons.

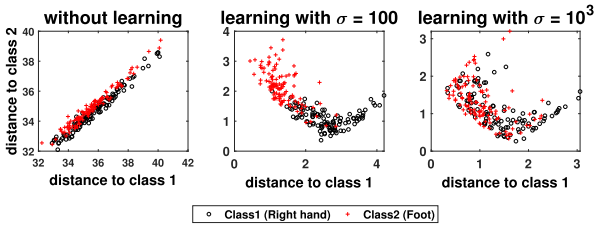

Fig. 1. The distribution of Riemannian distance on test set of subject aw.

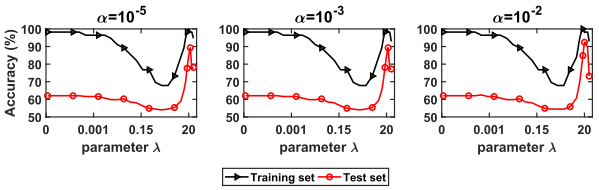

Fig. 2. Classification accuracies of subject aw obtained by CAPM-AP under different λ and α.

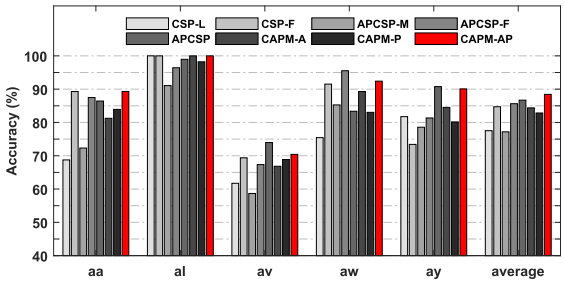

Fig. 3. Comparison of Classification Accuracies on BCI III Dataset IVa.

3) Parameter Setting: The number of CSP spatial filter was set to be 4 according to [30]. For CSP-F and APCSP-F, the broadband 7–30 Hz was divided into 22 sub-bands, i.e., 7–9 Hz, 8–10 Hz,..., 28–30 Hz. The best sub-band for each subject was tuned to obtain the best accuracy in the training set by 5×5 CV. In CAPM, the observed frequencies were ranged from 7 Hz to 30 Hz with a step of 2 Hz. For the graph embedding, we empirically set the parameters $p = 5$ and $\sigma = 100$. The dimensionality M was chosen from $\{4, 6, 8, 10, 12, 14, 16\}$. For the regularized parameters, α was set to 0.01, and λ was chosen from 10^{-4}, 10^{-3},..., 10, 20, 50, 10^2. Both parameters M and λ were determined by 5×5 CV on the training set.

B. Results and Discussion

To show the graph embedding effect, we compared the distribution of Riemannian distance under different conditions. Fig. 1 shows that our graph embedding method exhibits the high separability in Riemannian distance when σ is a suitable value. We also analyzed how parameters α and λ affect the classification results. In Fig. 2, we illustrate the accuracies of CAPM-AP under different α and λ. When λ is small ($\lambda < 0.001$), the high training accuracies but relatively low test accuracies are depicted. This case indicates the over-fitting problem. With the increase of λ and an appropriate α, the effect of L_1 penalty increases, and the over-fitting problem is alleviated. We can obtain the highest performance on both training set and test set with $\lambda = 20$ and $\alpha = 0.01$.

Fig. 3 and Table I show the classification results on the test set for the two datasets, respectively. The Wilcoxon signed rank test was adopted to further show statistical significance of differences between the paired methods (the proposed CAPM-AP method vs. the other methods). As shown in Fig. 3, the CAPM-AP method outperforms other methods in mean accuracy and the improvements are statistically significant except APCSP-F ($p = 0.312$) and APCSP ($p = 0.156$). It is worth noting that the algorithms utilizing amplitude and phase contents often can obtain better performance than the algorithms only utilizing amplitude or phase. On the other hand, CSP-F/APCSP-F outperforms CSP-L/APCSP-M by 7.19%/8.45% in mean accuracy. This confirms the importance of frequency information in portraying the rhythmic activities. However, user-set or predefined passband is not always 'optimal' [9]. The proposed CAPM-AP method can find the class-discriminative band by the optimization of spectral filters. This may be one reason that CAPM-AP performs better than APCSP (with a 1.75% improvement). The top three methods in Table I also consistently optimize frequency band before extracting CSP-based features [28], [29]. However, the proposed CAPM-AP method outperforms them over 6 subjects. In mean accuracy, CAPM-AP outperforms the best winner method by 1.3%. Although this improvement is slight, the statistical difference is close to being significant ($p = 0.072$).

To further interpret the superiority of the proposed CAPM-AP method, Fig. 4 shows the resulting spectrum after applying the most discriminative filter learned by each method. Both CSP-F and APCSP-F detect the passband around 11–14 Hz. Without a time-consuming procedure of passband selection, CAPM-AP also captures this useful band. Additionally, a large peak around 7–10 Hz can be found in the spectra obtained by APCSP-F and CAPM-AP. The classification results on subject aw indicate that this discriminative band assists in enhancing the performance. We conclude that simultaneous measurement of amplitude and phase can reveal more discriminative information. The underlying mechanism reflected by EEG amplitude-phase characteristics would be a meaningful issue for exploration. Besides, Riemannian geometry-based methods have shown promising results for transfer learning [17] in BCIs. Further studies can optimize the CAPM method and investigate its application in transfer learning.

TABLE I
PERFORMANCE COMPARISON IN ACCURACIES(%) FOR PROPOSED CAPM AND WINNER METHODS ON DATASET 2B OF BCI COMPETITION IV

Methods	S1	S2	S3	S4	S5	S6	S7	S8	S9	Mean
1st	70.0	60.5	**61.0**	97.5	**93.0**	80.5	78.0	92.5	87.0	80.0
2nd	71.0	60.5	57.0	97.0	85.5	81.0	80.5	92.0	**89.0**	79.3
3rd	59.5	56.0	56.0	88.5	78.5	74.5	69.0	92.5	80.5	72.8
CAPM-AP	**72.5**	**65.5**	56.5	**98.0**	92.0	**82.0**	**83.0**	**94.0**	88.0	**81.3**

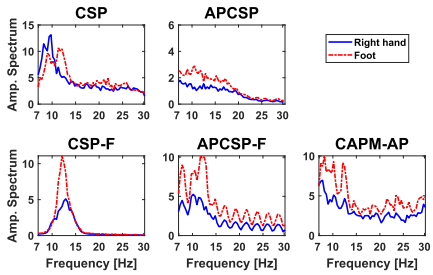

Fig. 4. The averaged amplitude spectra over trials for subject aw after applying the most discriminative filter on the training set.

REFERENCES

[1] I. Lazarou, S. Nikolopoulos, P. C. Petrantonakis, I. Kompatsiaris, and M. Tsolaki, "EEG-based brain-computer interfaces for communication and rehabilitation of people with motor impairment: A novel approach of the 21st century," *Frontiers Hum. Neurosci.*, vol. 12, pp. 1–18, Jan. 2018, Art. no. 14.

[2] G. Pfurtscheller and F. H. L. da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neuriophysiol.*, vol. 110, no. 11, pp. 1842–1857, 1999.

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

[4] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," *IEEE Signal Process. Lett.*, vol. 16, no. 8, pp. 683–686, Aug. 2009.

[5] W. Samek, M. Kawanabe, and K. Mller, "Divergence-based framework for common spatial patterns algorithms," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 50–72, 2014.

[6] S. Kumar, T. K. Reddy, V. Arora, and L. Behera, "Formulating divergence framework for multiclass motor imagery EEG brain computer interface," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2020, pp. 1344–1348.

[7] W. Wu, Z. Chen, X. R. Gao, Y. Q. Li, E. N. Brown, and S. K. Gao, "Probabilistic common spatial patterns for multichannel EEG analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 639–653, Mar. 2015.

[8] S. Lemm, B. Blankertz, G. Curio, and K. R. Muller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 9, pp. 1541–1548, Sep. 2005.

[9] H. Higashi and T. Tanaka, "Simultaneous design of FIR filter banks and spatial patterns for EEG signal classification," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 4, pp. 1100–1110, Apr. 2013.

[10] T. K. Reddy, V. Arora, L. Behera, Y. K. Wang, and C. T. Lin, "Multi-class fuzzy time-delay common spatio-spectral patterns with fuzzy information theoretic optimization for EEG-based regression problems in braincomputer interface (BCI)," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 10, pp. 1943–1951, Oct. 2019.

[11] P. Sauseng and W. Klimesch, "What does phase information of oscillatory brain activity tell us about cognitive processes?" *Neurosci. Biobehav. Rev.*, vol. 32, no. 5, pp. 1001–1013, Jul. 2008.

[12] R. F. Helfrich *et al.*, "Spectral fingerprints of large-scale cortical dynamics during ambiguous motion perception," *Hum. Brain Mapping*, vol. 37, no. 11, pp. 4099–4111, 2016.

[13] L. I. Jie and L. Zhang, "Phase interval value analysis for the motor imagery task in BCI," *J. Circuits*, *Syst. Comput.*, vol. 18, no. 8, pp. 1441–1452, 2009.

[14] X. Li, H. Fan, H. X. Wang, and L. P. Wang, "Common spatial patterns combined with phase synchronization information for classification of EEG signals," *Biomed. Signal Process. Control*, vol. 52, pp. 248–256, 2019.

[15] B. Chakraborty, L. Ghosh, and A. Konar, "Designing phase-sensitive common spatial pattern filter to improve brain-computer interfacing," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 7, pp. 2064–2072, Jul. 2020.

[16] F. Yger, M. Berar, and F. Lotte, "Riemannian approaches in brain-computer interfaces: A review," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 25, no. 10, pp. 1753–1762, Oct. 2017.

[17] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, 2018, Art. no. 031005.

[18] S. Yan, D. Xu, B. Zhang, H. J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.

[19] P. Rodrigues, F. Bouchard, M. Congedo, and C. Jutten, "Dimensionality reduction for BCI classification using Riemannian geometry," in *Proc. Graz BCI*, Sep. 2017.

[20] X. F. Xie, Z. L. Yu, Z. H. Gu, J. Zhang, L. Cen, and Y. Q. Li, "Bilinear regularized locality preserving learning on Riemannian graph for motor imagery BCI," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 26, no. 3, pp. 698–708, Mar. 2018.

[21] F. P. Kalaganis, N. A. Laskaris, E. Chatzilari, S. Nikolopoulos, and I. Kompatsiaris, "A Riemannian geometry approach to reduced and discriminative covariance estimation in brain computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 1, pp. 245–255, Jan. 2020.

[22] 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.

[23] M. Le Van Quyen *et al.*, "Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony," *J. Neurosci. Methods*, vol. 111, no. 2, pp. 83–98, 2001.

[24] X. He, D. Cai, and P. Niyogi, "Tensor subspace analysis," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 18, 2005, pp. 499–506.

[25] Jerome, T. Friedman, R. Hastie, and Tibshirani, "Regularization paths for generalized linear models via coordinate descent." *J. Stat. Softw.*, vol. 33, no. 1, pp. 1–22, 2010.

[26] "Description of BCI Competition III Data Set IVa," Accessed: Feb. 16, 2021. [Online]. Available: http://www.bbci.de/competition/iii/desc_IVa.html

[27] "Description of BCI Competition IV Data Set IIb," Accessed: Feb. 16, 2021. [Online]. Available: http://bbci.de/ competition/iv/desc_2b.pdf

[28] A. K. Keng, C. Z. Yang, W. Chuanchu, G. Cuntai, and Z. Haihong, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers Neurosci.*, vol. 6, pp. 1–9, 2012, Art. no. 39.

[29] "Results of BCI Competition IV Data Set IIb," Accessed: Feb. 16, 2021. [Online]. Available: http://www.bbci.de/competition/iv/results/

[30] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Common spatial pattern revisited by Riemannian geometry," in *Proc. IEEE Int. Workshop Multimedia Signal Process.*, 2010, pp. 472–476.