Geometry-aware Common Spatial Patterns for Motor Imagery-based Brian-Computer Interfaces

Qin Jiang*, Yi Zhang, Xin Hu, Wei Wang and Geng-Yu Ge

Abstract—**Background: The common spatial patterns (CSP) is a widely used EEG feature extractor for motor imagery-based brain-computer interfaces, with the optimal spatial filter formulated as a generalized Rayleigh quotient. However, the traditional CSP uses Euclidean metric, which ignores the specific geometric structure of symmetry positive definite (SPD) matrices, resulting in issues such as swelling effect, non-complete space, and indefinite matrices. Methods: To address these limitations, this paper introduces three alternative approaches with considering the geometric properties of SPD matrices. The geometry-aware CSP with diagonalization (gaCSPd) replaces the Euclidean means in the joint diagonalization principle of CSP with Riemannian means. The geometry-aware CSP with maximum discriminative information between classes (gaCSPb) aims to find an optimal projection matrix on a Riemannian manifold while maximizing the Riemannian distance between classes. The geometry-aware CSP with maximum within-class variance (gaCSPw) seeks a low-dimensional submanifold with the maximum intra-class variance in the projected data. Results: Experiment results on two BCI competition datasets demonstrate the competitiveness against state-of-the-art methods and confirm the effectiveness of geometry-aware CSP as a feature extractor for motor imagery-based brain-computer interfaces.**

Index Terms—**Brain-Computer Interfaces; Motor Imagery; Common Spatial Patterns; Dimensionality Reduction; Riemannian Manifold.**

Manuscript received December 18, 2023; revised June 20, 2024.

This research was supported by Scientific and Technological Research Program of Chongqing Municipal Education Commission (No. KJQN202303123, No. KJQN202303111), the National Natural Science Foundation of China (No. 62106024), the China Postdoctoral Science Foundation (No. 2022M711458), the Chongqing Language Research Project (No. yyk22105).

Qin Jiang is a researcher of Chongqing College of Electronic Engineering, Chongqing, 401331 China (Corresponding author to provide Email: namy_jiang@hotmail.com, Phone:023-65926593, Fax:023-65627000)

Yi Zhang is a professor of School of Advanced Manufacturing Engineering, Chongqing University of Posts and Telecommunications, Chongqing, 400065 China (e-mail: zhangyi@cqupt.edu.cn).

Xin Hu is an associate professor of Yangtze Normal University, Chongqing, 408100 China (e-mail: huxin@yznu.edu.cn).

Wei Wang is a PhD candidate of Chongqing University of Posts and Telecommunications, Chongqing, 400065 China (e-mail: d190201021@stu.cqupt.edu.cn).

Geng-Yu Ge is a PhD candidate of Chongqing University of Posts and Telecommunications, Chongqing, 400065 China (e-mail: d190201004@stu.cqupt.edu.cn).

I. Introduction

Motor imagery-based brain-computer interface (MI-BCI) is a technology that allows individuals to interact with extern devices using their brain signals [1], providing a novel communication pathway for individuals with a severely disabilities. One of the most extensively studied MI-BCI is based on electroencephalography (EEG), which measures the electrical activity of neural oscillations associated with motor imagery. When performing a specific mental task, certain frequency bands in specific locations of brain exhibit relative power decrease or increase, known as event-related desynchronization or synchronization of neural oscillation [2]. By analyzing these changes, the motion intentions can be decoded from the EEG signals [3] and converted into recognizable computer commands for controlling a wheelchair [4] or a prosthetic limb[5],[6]. However, accurately and efficiently decoding sensorimotor rhythms poses a great challenge due to the sensitivity to noise and susceptibility to environmental and mental states of the subject.

Common Spatial Pattern (CSP) has shown to be one of the most efficient feature extractors for multi-channel EEG signals [7]. It effectively finds a set of spatial filters that maximize the variance of band-pass filtered EEG signals from one class while minimizing it from the other class [8],[9]. In essence, CSP is designed to maximize the projected variance ratio between the covariance matrices of two classes of EEG patterns [10], where the optimization problem is typically solved using as a generalized eigenvalue decomposition between the two inter-class covariance matrices. However, a standard CSP tends to be over fitted in a small training set [11], [12]. To address this issue, various regularized CSPs have been developed by incorporating a-prior information to the estimation of the inter-class covariance matrix [13]. The regularization matrix can take different forms, such as an identity matrix [13], a diagonal matrix [14], an l_1 or l_2 regularizer [15], or a generic covariance matrix constructed as a weighted sum of covariance matrices using data from other subjects [16]. Nevertheless, none of these methods consider the geometric properties of the covariance matrix itself.

Covariance matrices are used to capture the spatial dynamics of EEG signals, while interclass means encode the discriminative information [17]. Moreover, these covariance matrices can be viewed as points on a Riemannian manifold [18], which is a mathematical space with a curved structure.

By leveraging the intrinsic structure of this manifold, novel approaches for EEG data analysis can be developed. Barachant et al. [19],[20] introduced Riemannian geometry into the classification of motor imagery BCI and proposed the minimum distance to the Riemannian mean algorithm (MDRM) and the tangent space classification algorithm, whose experimental results outperformed complex and highly parametrized CSP classifiers. The MDRM also achieved competitive results in the classification of evoked potentials [21], [22] and event-related potential [23]. Xie et al. [24],[25] presented a series of dimensionality reduction algorithms for covariance matrices by combining Riemannian metric with traditional dimensionality reduction algorithms such as local linear embedding and isometric mapping. Horev et al. [26] formulated the principle component analysis (PCA) on Riemannian manifold, projecting covariance matrices of EEG signals into a lower-dimensional subspace to best preserve variance, analogous to PCA in the Euclidean space. Additionally, capitalizing on the geometric properties of symmetric positive definite (SPD) matrices, several domain adaptation techniques have been developed to make the time-series data across sessions or subjects comparable [27]. Zanini et al. [28] proposed a Riemannian alignment method that normalizes the covariance matrices of sessions or subjects with respect to the reference matrix in the resting state. Yair et al. [29] introduced a domain adaptation method using parallel transport on the SPD manifold, which projects the symmetric matrices onto a shared tangent space to alleviate domain shifts. Rodrigues et al. [30] presented a Riemannian Procrustes analysis, which estimates the statistical characteristics of datasets with the geometric means of the SPD matrices and matches statistical distributions through simple geometric transformations like translation, scaling, and rotation. Zhang et al. [31] introduced a manifold embedded knowledge transfer framework, which centers covariance matrices of subjects with respect to their Riemannian mean and performs domain adaptation by minimizing distribution divergence between domains while preserving geometric structure.

In light of the advantages of the geometric properties of Riemannian manifold, three geometry-aware CSP (gaCSP) frameworks are presented. The first framework, geometry-aware CSP with diagonalization (gaCSPd), modifies the joint diagonalization principle of CSP by utilizing Riemannian means instead of Euclidean means. The second framework, Geometry-aware CSP with maximum discriminative information between classes (gaCSPb), seeks an optimal projection matrix on Riemannian manifold while maximizing the interclass Riemannian distance to enhance the separability of different classes in the feature space. The third framework, Geometry-aware CSP with maximum within-class variance (gaCSPw), interprets CSP filtering as a dimensionality reduction technique on Riemannian manifold with the goal of maximally preserving discriminative information within each class.

The exiting CSP algorithm relies on Euclidean mean and PCA, which may not be suitable for analyzing SPD matrices due to a non-complete space [32]. The Euclidean mean may lead to distortions in regression or average, resulting in the swelling effect, where the determinant of the average is larger than any of the individual matrices [33]. To overcome this issue, the proposed geometry-aware CSP frameworks replace the Euclidean mean with the Riemannian mean. Additionally, the traditional CSP achieves maximum variance by applying PCA to the covariance matrices, while the Euclidean formulation of PCA fails to capture data variation in regions of high curvature [26], making it inefficient for analyzing SPD matrices. In contrast, geometry-aware CSP frameworks formulate the discriminative intra-class or inter-class information in terms of geodesics, leveraging the geometric interpretation of the optimal problem and the intrinsic structure of the data.

The main contributions of this paper are as follows:

1) Three different derivation forms in Riemannian manifold are presented based on the theory of the CSP algorithm.

2) A closed-form solution on the Stiefel manifold is proposed for optimizing the geometry-aware CSP methods.

The rest of this paper is arranged as follows. Section 2 provides a brief introduction to the Riemannian metric theory and the principle of the CSP algorithm. Section 3 explains the rationale behind the proposed methods and presents the formulations for solving the optimization problem. Section 4 describes the experiments conducted to evaluate the performance of the proposed methods. Section 5 explores the advantages of gaCSP, and Section 6 concludes this paper.

II. Background

Notions

In this paper, S_+^n denotes the space spanned by the $n\times n$ SPD matrices, and S^n denotes a space spanned by $n\times n$ symmetric matrices. I_n denotes an $n\times n$ identity matrix. $E_i \in \mathbb{R}^{n\times t}$ denotes a 'trial' that holds the EEG signal recorded with n electrodes and t time samples. C_i represents a covariance matrix in the Euclidean space, and X_i denotes a covariance descriptor in the Riemannian manifold. $GL(n)$ is a group of real invertible $n\times n$ matrices. $Exp_X(.)$ and $Log_X(.)$ denote the exponential and logarithmic maps at the reference point X, respectively. $\exp(.): S_+^n \to S^n$ is defined as $\exp(X) = Udiag(\exp(\lambda_1, \lambda_2, .. , \lambda_n))U^T$ and the matrix logarithmic operator $\log(.): S^n \to S_+^n$ is defined as $\log(X) = Udiag(\log(\lambda_1, \lambda_2, ..., \lambda_n))U^T$. $(.)^T$ denotes the transpose operator. $\|X\|_F = \sqrt{Tr(X^T X)}$ represents the Frobenius norm, where $Tr(.)$ is the sum of the diagonal elements.

A. Common Spatial Patterns

$E_i^{(c)} \in \mathbb{R}^{n\times t}$ denotes an EEG record of class $c \in \{1, 2\}$, and the covariance matrix of a signal trial is expressed as:

$$C_i^{(c)} = \frac{E_i^{(c)}.E_i^{(c)T}}{Tr(E_i^{(c)}.E_i^{(c)T})}, \quad (1)$$

CSP aims to find a decomposition in a set of dichotomous data by jointly diagonalizing the inter-class covariance matrices, resulting in modes that are common to both data sets while maximizing the distinguishability between them. The joint diagonalization steps are implemented as follows.

First, the covariance matrices of two groups are pooled to form two normalized symmetric matrices, denotes as $\sum_1$ and $\sum_2$, then an orthogonal matrix U are achieved by:

$$U(\sum_1 + \sum_2)U^T = \Lambda, \tag{2}$$

Equation (2) is further compressed or stretched along the principal axes to adjust the directions to an isotropic distribution, whose transformation matrix is $P = U^T\sqrt{\Lambda^{-1}}$. Thus (2) amounts to:

$$P\sum_1 P^T + P\sum_2 P^T = I, \tag{3}$$

Let $S_1 = P\sum_1 P^T$, $S_2 = P\sum_2 P^T$, where S_1 and S_2 are symmetric matrices. $S_1 = B\lambda_1 B^T$ and $S_2 = B\lambda_2 B^T$ are the diagonalizations of S_1 and S_2, respectively, with the constraint $\lambda_1 + \lambda_2 = I$. Then, (3) is expressed as:

$$B^T P(\sum_1 + \sum_2)P^T B = I, \tag{4}$$

Let $W = B^T P$, (4) is simplified as $W(\sum_1 + \sum_2)W^T = I$.

Equation (4) captures the essence of the CSP method, which states that the common transformation matrix (W) whitens the combined data from both classes, resulting in each individual having same principles axes and eigenvectors, with their corresponding eigenvalues adding up to one. As a consequence, the direction with the largest (smallest) eigenvalue of one of the groups is the eigenvector with the smallest (largest) eigenvalue of the other group. The solution for the transformation matrix can be summarized as a two-step process: joint whitening and individual diagonalization. Whitening matrix ensures the data from both classes have the same principal axes and eigenvectors, allowing for a common representation. Individual diagonalization then produces an orthogonal matrix that maximizes the separation between the two classes. The common spatial filter (ω) is obtained by multiplying the orthogonal matrix with the whitening matrix.

For a given EEG signal, the extracted feature is the logarithm of its variance after projection onto ω:

$$f_i = \log(\frac{var(\omega E_i)}{sum(var(\omega(E_i)))}), \tag{5}$$

where $var(.)$ and $sum(.)$ represent the operation of the variance and summing, respectively.

B. Riemannian Geometry

The solving procedure of the CSP algorithm reveals that the optimal solution involves the generalized eigen-decomposition of the inter-class means of covariance matrices. Since the covariance matrix is symmetric and positively definite, it can be identified as a point in a Riemannian manifold. Therefore, it is feasible to introduce Riemannian metric to analyze covariance matrices.

Riemannian mean: The Riemannian mean is defined as the point minimizing the metric dispersion:

$$M_R = \arg\min_{X^* \in S_+^n} \sum_{i=1}^{N} \delta_R{}^2(X^*, X_i), \tag{6}$$

where $\delta_R(X^*, X_i)$ denotes a distance suitable for the Riemannian manifold. Different from algebraic mean, Riemannian mean does not have a closed-form solution [34], but it can be solved using an iterative algorithm.

Affine-Invariant Riemannian Metric: The affine-invariant Riemannian metric (AIRM) defines the geodesic between two metrics:

$$\delta_{air}(X_i, X_j) = \left\|\log(X_i^{-1/2} X_j X_i^{-1/2})\right\|_F, \tag{7}$$

Log-Euclidean Metric: The Log-Euclidean Metric (LEM) defines the geodesic distance of two SPD matrices [35] by computing the distance between their corresponding tangent vectors at the identity matrix:

$$\delta_{lem}(X_i, X_j) = \left\|\log(X_i) - \log(X_j)\right\|_F, \tag{8}$$

Let $X = U\Sigma U^T$ be the eigendecomposition of the SPD matrix, where Σ is the diagonal matrix of the eigenvalue. The logarithmic map can be computed easily by $\log(X) = U\log(\Sigma)U^T$. Compared to the AIRM, the Log-Euclidean requires less computation. Additionally, the LEM uniquely defines the Riemannian mean by $M_R = \exp(\sum \log(X_i)/\mathrm{N})$, while the affine-invariant mean is obtained through an iterative process.

For any two covariance matrices from class 1 and class 2, assume that there exists an orthogonal transformation matrix W that satisfies $W^T(C_1^i + C_2^j)W = I$. Let $WC_1^i = \lambda_1 W$, then $WC_2^j = (I - \lambda_1)W$, the Euclidean distance between two projected matrices is given by the Frobenius norm:

$$\delta_E(WC_1^i W^T, WC_2^j W^T) = \sqrt{\sum_{i=1}^{p} 4(\lambda_1^i - 0.2)^2}, \tag{9}$$

The geodesic distance measured by Log-Euclidean metric is expressed as:

$$\delta_{lem}(WX_1^i W^T, WX_2^j W^T) = \sqrt{\sum_{i=1}^{p} \log^2(\frac{\lambda_1^i}{1-\lambda_1^i})}, \tag{10}$$

Based on (9) and (10), the functional relationship between distance and eigenvalues is given by $f_E(\lambda) = 2|\lambda - 0.5|$ for Euclidean space, and $f_R(\lambda) = \sqrt{\log^2(\lambda/1-\lambda)}$ for Riemannian metric, with $\lambda \in [0 \quad 1]$.

Fig. 1 depicts the curve of distance $f(.)$ changing with eigenvalue λ_i. It is observed that for any eigenvalue, the corresponding Riemannian distance is greater than the Euclidean distance. The largest and smallest eigenvalues correspond to the maximum distance. As the value deviates from 0.5, the gap between Euclidean distance and Riemannian distance increases rapidly. Conversely, as the eigenvalue approaches 0.5, the gap decreases sharply. Regardless of the specific eigenvalue, the Riemannian distance consistently outperforms the Euclidean distance, indicating that the Riemannian metric is more suitable for describing the difference between two matrices than the Euclidean distance. The following text explores the

feasibility of introducing Riemannian metrics into CSP to improve its feature description capabilities.

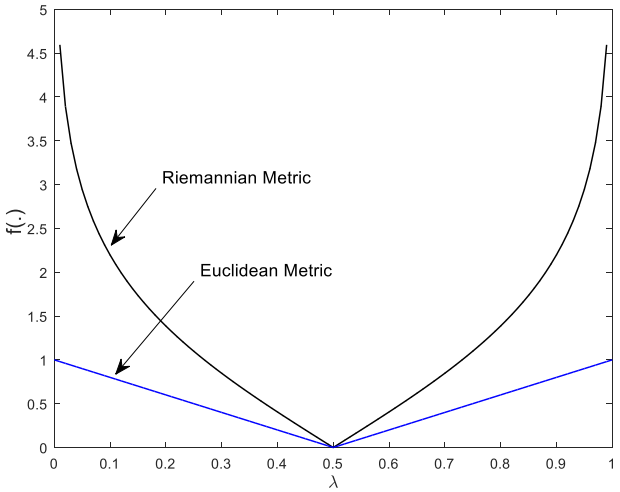


Fig. 1. The relationship between the divergence and the eigenvalues

III. Riemannian Geometry-based CSP

A. gaCSP with diagonalization (gaCSPd)

The gaCSPd algorithm extends the joint diagonalization principle used in the CSP algorithm to incorporate the Riemannian mean, which the key idea is to align each covariance matrix by approximating it as an identity. The gaCSPd algorithm reformulates a two-step solving process that is applicable to the Riemannian mean. The first step involves aligning the covariance matrices by approximating them as identity matrices. The second step involves applying the joint diagonalization technique to further refine the alignment and extract discriminative features.

Whitening: In gaCSPd, the interclass mean is whitened using the Riemannian mean, resulting in:

$$\bar{X}^{-1/2}(\bar{X}_1+\bar{X}_2)\bar{X}^{-1/2}=2I_n, \tag{11}$$

Let $\tilde{P}_1=\bar{X}^{-1/2}\bar{X}_1\bar{X}^{-1/2}$ and $\tilde{P}_2=\bar{X}^{-1/2}\bar{X}_2\bar{X}^{-1/2}$, it can be proven that $\tilde{P}_1$ and $\tilde{P}_2$ are symmetric positive definite matrices.

Diagonalization: Subsequently, a singular value decomposition is performed on $\tilde{P}_1$, yielding $\tilde{P}_1=U^T\Lambda_1 U$. Substituting this into (11) yields $\tilde{P}_2=2I_n-U^T\Lambda_1 U$, which is equivalent to $U^T(\Lambda_1+\Lambda_2)\,U=2I_n$. This relationship guarantees that the two transformation matrices $\tilde{P}_1$ and $\tilde{P}_2$ have the same principle axes and eigenvectors, and the dominant eigenvector produces the most discriminative features.

Finally, the geometrical transform matrix ($\omega\in\mathbb{R}^{n\times m}$) is constructed using the eigenvectors corresponding to the m/2 largest and smallest eigenvalues, where $\omega=\hat{U}^T\bar{X}^{-1/2}$ with $\hat{U}=U_{1:m/2,n-m/2:n}$.

Overall, although both gaCSPd and CSP are based on the principle of joint diagonalization decomposition, there are two significant differences: 1) gaCSPd uses a Riemannian mean instead of the algebraic mean (the differences between the two means will be detailed in the discussion section); 2) the whitening matrix of gaCSPd is composed of all eigenvectors and eigenvalues of the Riemannian mean.

B. gaCSP with maximum inter-class distance (gaCSPb)

When applied to EEG recordings, CSP filter compresses the covariance matrix of filtered signals into a low dimensional SPD matrix, which implies that the spatial filter transforms the SPD matrices into a more discriminative submanifold.

The generic dimensionality reduction on the Riemannian manifold is defined as $f: S_+^n \mapsto S_+^m$:

$$f(X)=\omega^T X\omega, \tag{12}$$

where X is a matrix in S_+^n ($0\prec X\in S_+^n$). To guarantee that the output of (12) is an SPD matrix in the submanifold S_+^m, the transform matrix $\omega\in\mathbb{R}^{n\times m}$ ($m<n$) must be full rank, i.e., $\omega^T\omega=I_m$.

The gaCSPb aims to find a projection matrix to reduce the dimension of the covariance matrix of raw EEG recording while preserving the distinguishability between classes. By employing the Riemannian distance as the criterion for interclass discrimination, the objective function is formulated as:

$$\begin{aligned}&\omega^*=\arg\max_{\omega\in\mathbb{R}^{n\times m}}\delta_R^2(\omega^T\bar{X}_1\omega,\omega^T\bar{X}_2\omega)\\&s.t.\quad \omega^T\omega=I_m\end{aligned}, \tag{13}$$

where $\bar{X}_c$ is the Riemannian mean from *c*-class.

Compared to the spatial filter of CSP, the projection matrix of gaCSPb is an orthogonal matrix, which can more strictly eliminate redundant information, reducing dependence on irrelevant information.

C. gaCSP with maximum within-class Variance (gaCSPw)

The gaCSPw aims to find a low-dimensional SPD manifold where the projected data have the greatest variance within the class. To characterize the variance between the SPD matrices, the PCA is extended for matrix data that collectively have a maximum Riemannian distance to their Riemannian mean. Simultaneously, an effort should also be made to preserve the discriminative information provided by labels, thus a constraint condition is imposed to preserve the distinguishing information between classes.

The objective function of gaCSPw is given by:

$$\begin{aligned}&J(\omega)=\arg\max_{\omega\in\mathbb{R}^{n\times m}}\sum_{i=1}^{N_c}\delta_R^2(\omega^T X_c^i\omega,\omega^T\bar{X}_c\omega)\\&s.t.\quad \omega^T(\bar{X}_1+\bar{X}_2)\omega=I_m\end{aligned}, \tag{14}$$

where X_c^i represents the *i*-th covariance matrix and $\bar{X}_c$ is the Riemannian mean from *c*-class, and $\bar{X}$ denotes the Riemannian mean of the entire set.

Formula (14) is a distance-based optimization that preserves the maximum dispersion of the covariance matrices rather than the maximum variance of EEG records. However, there are two key issues in the optimization of (14). The first is that the function (14) does not accurately describe the distance of the low-dimensional matrices. As

the output of $\omega^T \bar{X}_c \omega$ is a compressed value of the *c*-class Riemannian mean, rather than the mean ($\bar{M}_c$) of the compressed matrix set $\{\omega^T X_c^i \omega\}$, i.e., $\omega^T \bar{X}_c \omega \neq \bar{M}_c$ Furthermore, according to the definition of equation (12), dimensionality reduction on Riemannian manifold requires the projection matrix to be a point on the Stiefel manifold. However, since not every point on the Stiefel manifold can satisfy the constraint of (14), the objective function (14) cannot be unified into an optimization problem on the Stiefel manifold.

To address the first issue, centroid alignment is performed using (11), which approximately transforms the raw covariance matrix to an identity matrix, resulting in $X_c^i \mapsto \tilde{X}_c^i = M X_c^i M$, where $M = \bar{X}^{-1/2} / \sqrt{2}$ represents the transformed matrix. As a result, the inter-class mean (denoted as $\tilde{X}_c$) of the transformed matrix $\left\{\tilde{X}_c^i\right\}$ is approximately a diagonal matrix, and its compressed mean is also a diagonal matrix, i.e., $\omega^T \tilde{X}_c \omega = \Lambda_m$. This makes it quite reasonable to approximate the mean of the compressed dataset ($\{\omega^T \tilde{X}_c^i \omega\}$) using the compressed mean ($\omega^T \tilde{X}_c \omega$). Another consequence of centroid alignment is to simplify the constraint of (14) into an orthogonal constraint ($\omega^T \omega = I_m$).

The above improvements transform the objective function (14) into:

$$J(\omega) = \arg\max_{\omega \in \mathbb{R}^{n \times m}} \sum_{i=1}^{N_c} \delta_R^2 (\omega^T \tilde{X}_c^i \omega, \omega^T \tilde{X}_c \omega), \quad s.t. \quad \omega^T \omega = I_m \tag{15}$$

The final optimal solution is expressed as $\tilde{\omega} = M * [\omega_1 : \omega_2]$, where ω_c represents the optimal projection corresponding to *c*-class solved by (15).

D. Optimization

Given that both (13) and (15) involve unitary constraints, the optimization problem can be transformed into an unconstrained solution on the Stiefel manifold [36]. In gaCSPb and gaCSPw, various metrics can be adopted to encode the distance between SPD matrices, such as LEM, AIRM, the Stein divergence, or the Jeffrey divergence. For ease of calculation, the presented optimization scheme selects LEM as the metric and transforms the optimization problem on the Stiefel manifold into an iterative generalized eigen-decomposition problem.

The LEM distance of (15) is expressed as:

$$\delta_R^2 (\omega^T \tilde{X}_c^i \omega, \omega^T \tilde{X}_c \omega) = \left\| \log(\omega^T \tilde{X}_c^i \omega) - \log(\omega^T \tilde{X}_c \omega) \right\|_F^2 , \tag{16}$$

According to the proof in reference [32], $\log(\omega^T X \omega)$ is approximated as $\omega^T \log(X) \omega$, thus equation (15) is further simplified as:

$$\omega^* = \arg\max_{\omega \in \mathbb{R}^{n \times m}} \sum_{c=1}^{2} Tr(\omega^T F(\omega) \omega) \quad s.t. \quad \omega^T \omega = I_m \tag{17}$$

$$F(\omega) = \sum (\log(\tilde{X}_c^i) - \log(\tilde{X}_c)) \omega . \omega^T (\log(\tilde{X}_c^i) - \log(\tilde{X}_c))$$

where $i \in N_c$, N_c denotes the number of *c*-class samples.

It is evident that $F(\omega)$ is explicitly determined by ω , resulting in a problem of (17) without a closed-form solution. Through the logarithmic map, SPD matrices is flattened into Euclidean space, allowing for optimal solution using an iterative eigen-decomposition.

At the *t*-th iteration, $F^{(t)}(\omega)$ is calculated by $\omega^{(t-1)}$:

$$F^{(t)}(\omega) = \sum_{i \in N_c} (\log(\tilde{X}_c^i) - \log(\tilde{X}_c)) \omega^{(t-1)} \omega^{(t-1)T} (\log(\tilde{X}_c^i) - \log(\tilde{X}_c)) \tag{18}$$

The optimal projection matrix $\omega^{(t)}$ obtained by the *t*-th iteration is expressed as:

$$\omega^{(t)} = \arg\max_{\omega \in \mathbb{R}^{n \times m}} \sum_{c=1}^{2} Tr(\omega^T F^{(t)}(\omega) \omega) \quad s.t. \quad \omega^T \omega = I_m , \tag{19}$$

Thus, the optimization of (15) boils down to a generalized eigenvector-eigenvalue decomposition problem. The transform matrix $\omega^{(t)}$ is given by the m largest eigenvectors of $F^{(t)}(\omega)$. The whole iterative procedure is shown in Algorithm 1, and the solving process of gaCSPb is similar to that of gaCSPw, except for replacing (18) involved in step 4 of algorithm 1 with:

$$F^{(t)}(\omega) = (\log(\bar{X}_1) - \log(\bar{X}_2)) \omega^{(t-1)} \omega^{(t-1)T} (\log(\bar{X}_1) - \log(\bar{X}_2)) \tag{20}$$

Algorithm 1 Iterative eigendecomposition solver for gaCSPw with the Log-Euclidean metric

Input: training set with labels $\{X_i, y_i\}_{i=1}^N$, dimension m , iterations N_{ite} , and threshold ε

Output: the optimal spatial filers $\tilde{\omega} \in \mathbb{R}^{n \times m}$

1. Initialize $\omega^{(0)}$ with an identity matrix $\omega^{(0)} \leftarrow R_{n \times m}$
2. Calculate the inter-class means $\bar{X}_1$ and $\bar{X}_2$ and $\bar{X}$
3. **for** t=1:1: [$N_{ite} - 1$]
4. Construct $F^{(t)}(\omega)$ using (19)
5. Implement eigendecomposition on $F^{(t)}(\omega)$
6. $\omega^{(t)} \leftarrow$ m largest eigenvectors of $F^{(t)}(\omega)$
7. **if** $\left\| \omega^{(t)} - \omega^{(t-1)} \right\|_F^2 \leq \varepsilon$
8. **Break;**
9. **end if**
10. **end for**
11. Eventual projection matrix $\tilde{\omega} = M * \omega$, where M denotes a whitening matrix.

E. Classification of gaCSP features

After processing with gaCSP, the covariance matrices of the EEG signals are embedded into a low-dimensional Riemannian manifold, where the EEG signal features is represented as point in this manifold. There are two conventional approaches for the classification of SPD matrices. One is the minimum distance to the Riemannian mean algorithm (MDRM) [37], which determines the type of sample by the shortest Riemannian distance between the tested matrix and the two interclass means, and assigns a label based on the nearest distance. The other approach is the linear classifier in tangent space [38], which maps the SPD matrices onto a common tangent plane and then opts to

linear classifiers to classify these vectorized matrices. This paper investigates the effectiveness of gaCSP using MDRM and SVM classifier in tangent space (TSVM). For MDRM, gaCSP is used to embed the trained covariance matrices into a low-dimensional manifold and to obtain the corresponding Riemannian means. For a given test instance, the label is assigned based on the nearest Riemannian mean. For TSVM, the covariance matrices are transformed with gaCSP, and the Riemannian mean of the low dimensional SPD matrix is calculated. Then all matrices are mapped to the tangent space at reference point of Riemannian mean, where a traditional classifier (such as LibSVM) is used. Fig. 2 shows a schematic diagram of MDRM and TSVM algorithms.

Additionally, gaCSP can also extract the variance-based features of EEG signals using (5), allowing the use of conventional linear classifiers for classification.

IV. Experiments and Results

A. Dataset Description

A series of experiments were performed on two public datasets to validate the effectiveness of the proposed gaCSP.

BCI Competition III Dataset IVa **(DatasetIVa)** contains two-class EEG signals, with a total of 280 cue-based trials recorded by 118 channels at a 100Hz sampling rate.

BCI Competition IV Dataset IIa **(DatasetIIa)** consists of the four-class cued motor imagery data recorded by 22 channels with a 250Hz sampling rate. Please refer to [39] for more information about these two datasets.

In our experiments, the two datasets were first filtered by a six-order 8-30Hz bandpass filter and then segmented. As recommended by the competition winner, the trials of DatasetIIa were segmented from 2.5s to 4.5s, and DatasetIVa was captured using a window of 3s after a cue onset 0.5s, resulting in trials of 22×500 and 118×300, respectively. The spatial filter pairs and the value m in S_+^m was determined by the energy indicator of the eigenvectors with η=0.95.

Fig. 3 shows the cue timing schedule and event-related desynchronization (ERD) brain topography of DatasetIVa and DatasetIIa, respectively. It is observed that each task induces a unique ERD distribution, and the energy distribution exhibits significant individual differences. In terms of electrode-based feature representation, each electrode is considered as a dimension of the feature space, and the corresponding energy value represents the feature value. In this context, the CSP spatial filter is interpreted as a mechanism to reduce the influence of irrelevant electrodes and enhance the discriminative power of relevant electrodes by emphasizing the electrodes carrying more discriminative information and de-emphasizing irrelevant electrodes.

B. Results

1) Feature Distribution

This section visualized the feature distributions of gaCSP using the *t*-SNE toolkit. The gaCSP frameworks and CSP extracted variance-based features of the training samples by (5). Fig. 4 shows the transformed data distributions of 140 training samples of Subject 'AA', with different feature extractors. It is observed that gaCSPb and gaCSPw exhibit outstanding discriminability with smaller intra-class divergence and larger inter-class distance, highlighting the advantage of integrating CSP with Riemannian metrics. Additionally, although the distribution of gaCSPd is similar to CSP, gaCSPd exhibits better distinguishability between samples.

2) Sptial Filter

The projection matrix of gaCSP approximates the spatial filters of CSP, where each column corresponds to a principal direction and the corresponding value is the weight of the electrode. To verify the effectiveness of gaCSP in capturing task-related electrodes, topographical maps corresponding to the largest spatial filter were drawn under different task conditions ("right-handed foot", "left-right hand", and "foot-tongue"). Fig. 5 displays these maps, where color depth represents the correlation between the functional area and the specific task. Red indicates the degree of being activated, blue indicates inhibition, and green indicates that the given functional area is irrelevant to the task. It is observed that gaCSPd, although similar to CSP in shape, has fewer associated electrodes and higher weights on the task-related electrodes compared to CSP. gaCSPb and gaCSPw accurately locate the most active electrodes in CSP, resulting in more stable and accurate results compared to traditional CSP methods. This results illustrated that gaCSP are more resistant to noise and non-stationary signals than CSP.

3) Classification

This section delves into the classification performance of gaCSP on two publicly available datasets. DatasetIVa has been randomly divided into training and test data, with an equal number of samples for each class. DatasetIIa has been divided into two binary datasets, one for left and right tasks and another for foot and tongue tasks. gaCSP is capable of extracting both variance-based features and reduced SPD matrix features.

i) variance-based feature

The classification accuracies of CSP, gaCSPd, gaCSPw, and gaCSPb on DatasetIVa and DatasetIIa are summarized in Tables I and II, respectively. The SVM classifier was used for classification. The results indicate that, compared to CSP, average classification accuracies of the three gaCSP frameworks have been improved to varying degrees, with gaCSPb showing a more prominent improvement of 3.17%, 3.66% and 2.5%, respectively. A Wilcoxon signed-rank test (one-sided) was conducted on Table I and II to investigate the significance of accuracy differences over the CSP method (p<0.05). The results of Table III confirm the significant superiority of gaCSP in improving classification performance.

ii) SPD matrix feature

This section investigates the performance of SPD matrix feature generated by gaCSP. Table IV-VI report the results classified by MDRM and TSVM on DatasetIVa, DatasetIIa of Left-Right hand and DatasetIIa of Foot-Tongue, respectively. MDRM and TSVM are used as benchmark classifiers, with TSVM utilizing PCA to process tangent vectors to extract dimensions with 95% contribution rate. Overall, the results obtained with MDRM are superior to

those obtained with TSVM. Specifically, gaCSPb achieves average accuracies of 83.07%, 78.03%, and 74.51% on MDRM, compared to average accuracies of 79.37%, 75.97%, and 72.06% on TSVM. Similarly, gaCSPw achieves average accuracies of 82.8%, 77.02%, and 73.78% on MDRM, which are 4.4%, 3.98%, and 1.56% higher than those on TSVM. Additionally, gaCSPd achieves results of 80.41%, 73.67%, and 72.07% on MDRM, and 76.87%, 74.58%, and 71.86% on TSVM. The reason for these results may be that geometric features are able to better preserve discriminative information compared to flattened vector features. Table VII shows the *p*-values of different comparisons using Wilcoxon signed-rank test, which reveals that gaCSPb and gaCSPw have significantly higher performance than gaCSPd, but there is no significant difference between gaCSPb and gaCSPw.

4) Computational Efficiency

This section investigated the computational efficiency of each gaCSP algorithm. Table VIII reports the computational time evaluated under the environment of MATLAB R2016a on a laptop with 3.2GHz CPU (AMD Ryzen 7 5800H), 16GB RAM. It is worth noting that gaCSPd takes slightly more time than CSP with additional time spent on estimating Riemannian mean, and the time overrun increases with the increase of data dimensionality. In gaCSPb and gaCSPw, most of time is spent on estimating the Riemannian mean and calculating the Riemannian distance. It should be emphasized that although gaCSPb and gaCSPw require more computational time during the training phase, they do not require estimating Riemannian mean and distance for testing. Therefore, considering their significant classification results, gaCSPb and gaCSPw are still competitive.

TABLE I
CLASSIFICATION ACCURACIES ACHIEVED BY CSP, GACSPD, GACSPB AND GACSPW ON DATASET IVA, RESPECTIVELY.

	AA	AL	AV	AW	AY	Ave. ±Std.
CSP	70.57	97.86	65.45	83.43	**82.29**	79.92±11.28
gaCSPd	72.14	96.43	**70.15**	90.24	**82.29**	82.25±10.13
gaCSPw	77.86	**99.29**	67.71	87.14	80.62	82.52±10.22
gaCSPb	76.07	97.86	69.32	**90.79**	81.43	**83.09±9.81**

TABLE III
OVERVIEW OF WILCOXON SIGNED-RANK TEST P-VALUES (ONE-SIDED) FOR DIFFERENT COMPARISONS BASE ON TABLE I AND TABLE II.

Comparison	Dataset IVa	Dataset IIa (Left vs Right)	Dataset IIa (Foot vs. Tongue)
gaCSPd vs CSP	**0.0421**	0.0832	0.0671
gaCSPb vs CSP	**0.0031**	**0.0016**	**0.0301**
gaCSPw vs CSP	**0.0052**	**0.0019**	**0.0078**

TABLE VIII
THE TRAINING TIME (SECOND) OF CSP, GACSPD, GACSPB AND GACSPW.

	CSP	gaCSPd	gaCSPb	gaCSPw
DatasetIVa	0.328	1.259	11.402	72.435
DatasetIIa	0.098	0.149	0.283	5.011

TABLE II
CLASSIFICATION ACCURACIES ACHIEVED BY CSP, GACSPD, GACSPB AND GACSPW ON DATASET IIA, RESPECTIVELY.

	Left hand vs. Right hand				Foot vs. Tongue			
	CSP	gaCSPd	gaCSPw	gaCSPb	CSP	gaCSPd	gaCSPw	gaCSPb
A01	76.31	76.31	78.83	**80.91**	63.45	68.19	72.69	**74.81**
A02	60.56	62.34	**63.81**	62.34	82.71	84.44	**85.62**	83.65
A03	89.67	89.67	88.39	**90.48**	76.49	74.5	**79.65**	78.54
A04	65.78	61.86	65.61	**67.74**	61.09	61.09	**63.89**	61.34
A05	56.81	60.61	66.97	**70.51**	63.43	60.31	**65.54**	64.71
A06	60.37	58.64	**63.43**	62.03	60.92	**64.47**	60.31	63.27
A07	73.45	75.19	72.89	**77.46**	74.73	74.73	72.46	**75.19**
A08	86.17	86.17	86.63	**90.36**	87.71	85.28	87.19	**88.82**
A09	88.17	88.17	88.56	**90.81**	70.14	71.33	69.03	**72.85**
Ave.	73.03	73.22	75.01	**76.96**	71.19	71.59	72.93	**73.69**

TABLE VII
OVERVIEW OF WILCOXON SIGNED-RANK TEST P-VALUES (ONE-SIDED) FOR DIFFERENT COMPARISONS BASE ON TABLE IV-VI.

Comparison	MDRM			TSVM		
	DatasetIVa	DatasetIIa (Left-Right hand)	DatasetIIa (Foot-Tongue)	DatasetIVa	DatasetIIa (Left-Right hand)	DatasetIIa (Foot-Tongue)
gaCSPd vs gaCSPb	0.0021	0.0043	0.0033	0.0426	0.1722	0.2726
gaCSPd vs gaCSPw	0.0302	0.0019	0.0132	0.2041	0.0672	0.1985
gaCSPw vs gaCSPb	0.0572	0.4219	0.0382	0.0874	0.2283	0.3824

TABLE IV
CLASSIFICATION ACCURACIES ACHIEVED BY GACSPD, GACSPW AND GACSPB ON DATASET IVA, CLASSIFIED BY MDRM AND TSVM.

	MDRM	gaCSPd	gaCSPw	gaCSPb	TSVM	gaCSPd	gaCSPw	gaCSPb
AA	65.43	67.86	71.32	**73.28**	65.36	**74.64**	67.7	70.32
AL	96.43	97.86	**98.57**	**98.57**	75.36	90.71	97.14	**98.57**
AV	60.71	67.75	**70.21**	67.03	53.17	56.86	**65.29**	63.57
AW	73.5	83.57	87.96	**88.62**	68.57	82.86	81.84	**83.86**
AY	78.93	85	85.94	**87.86**	75.36	79.29	**80.53**	80.03
Ave.	75.00	80.41	82.80	**83.07**	67.56	76.87	78.50	**79.27**

TABLE V
CLASSIFICATION ACCURACIES ACHIEVED BY GACSPD, GACSPW AND GACSP WITH MDRM BEING THE CLASSIFIER ON DATASET IIA.

	Left hand vs. Right hand				Foot vs. Tongue			
	MDRM	gaCSPd	gaCSPw	gaCSPb	MDRM	gaCSPd	gaCSPw	gaCSPb
A01	74.72	71.94	78.89	**81.4**	60.94	67.09	74.09	**74.19**
A02	60.06	**63.61**	62.39	62.51	79.5	81.67	84.6	**85.92**
A03	88.53	85.83	90.43	**92.25**	74.19	73.81	73.72	**76.86**
A04	65.03	62.86	**67.33**	67.16	62.42	61.94	61.15	**63.45**
A05	55.83	68.06	75.4	**77.38**	66.25	62.92	65.7	**67.09**
A06	58.25	60.78	62.75	**64.04**	63.31	63.61	**64.94**	60.5
A07	73.81	73.42	**76.9**	75.43	74.73	76.11	**78.5**	77.97
A08	85.44	86.36	88.51	**89.59**	85.28	88.19	86.32	**88.66**
A09	86.92	90.14	90.59	**92.54**	72.11	73.33	74.97	**75.91**
Ave.	72.07	73.67	77.02	**78.03**	70.97	72.07	73.78	**74.51**

TABLE VI
CLASSIFICATION ACCURACIES ACHIEVED BY GACSPD, GACSPW AND GACSP WITH WITH TSVM BEING THE CLASSIFIER ON DATASET IIA.

	Left hand vs. Right hand				Foot vs. Tongue			
	TSVM	gaCSPd	gaCSPw	gaCSPb	TSVM	gaCSPd	gaCSPw	gaCSPb
A01	75.42	77.42	**81.4**	78.89	63.03	68.58	74.09	74.19
A02	56.89	**70.17**	62.51	62.39	78.14	78.83	81.6	**81.92**
A03	85.75	88.53	**89.25**	88.43	66.56	72.11	73.72	**76.86**
A04	53.44	65.03	67.16	**67.33**	43.19	**65.94**	61.15	58.45
A05	55.81	63.61	**77.38**	75.4	56.14	64.56	63.7	**65.09**
A06	51.19	58.94	**62.04**	60.75	64.69	**66.78**	64.94	60.5
A07	61.22	75.89	73.9	**76.43**	71.33	74.11	74.5	**76.97**
A08	81.28	83.36	**89.59**	88.51	85.14	83.06	**86.32**	82.66
A09	86.92	88.31	88.54	**90.59**	67.25	**72.81**	69.97	71.91
Ave.	67.55	74.58	**76.86**	76.52	66.16	71.86	**72.22**	72.06

V. DISCUSSION

The experimental results demonstrates that gaCSP algorithm is generally superior to CSP. Specifically, the features extracted by gaCSP are significantly more discriminating than those of CSP, as shown in Fig. 4. Furthermore, gaCSP frameworks consistently outperform CSP in terms of accuracy across most subjects, as demonstrated by the results in Tables I-VI. In addition, gaCSP frameworks exhibit stronger anti-interference ability and better filter robustness, as shown in Fig. 5. In order to further understand the reasons for these remarkable results, it is necessary to analyze the principles behind the algorithms.

A. Riemannian Mean vs. Euclidean Mean

Both gaCSPd and CSP utilize the joint diagonalization principle of inter-class means of covariance matrices and share a similar solving producers. However, in gaCSPd, the algebraic mean used in CSP is replaced with the Riemannian mean, which is a key different between gaCSPd and CSP. Fig. 6 visually illustrates the difference between Riemannian mean and arithmetic mean. To facilitate visualization, only EEG series on C3 and C4 electrodes were considered, and 50 trials were selected from the training set of subject 'AW'. Among these trials, 44 samples were grouped together, while the remaining 6 samples were located far away from the cluster center.

The covariance matrix C_i of the i-th trial is defined as:

$$C_i = \begin{pmatrix} Var(C_{3i}) & Cov(C_{3i}, C_{4i}) \\ Cov(C_{3i}, C_{4i}) & Var(C_{4i}) \end{pmatrix}, \quad (21)$$

where C_{3i} and C_{4i} are the time series of the i-th trial.

As illustrated in Fig. 6, for the concentrated SPD matrices (denoted by '+'), the discrepancy between Riemannian mean (represented by '○') and algebraic mean (represented by '□') are tiny, and both perform well in elaborating the central tendency for the variance. However, when outliers (denoted by '∗') are considered, the geometric mean (marked by a filled circle '●') deviates less from the center compared to the arithmetic mean (represented by a filled square'■'), indicating that Riemannian metric is more robust to outliers. In other words, the Riemannian mean is less influenced by the presence of outliers and provides a more reliable estimate of the central tendency in the presence of extreme data points. This quality is profit to alleviate the influence of noise and artifacts, providing a more robust and anti-interference mean for analyzing nonstationary EEG signals.

To further demonstrate the ability of Riemannian mean against noise interference, an eigendecomposition was performed on the inter-class Riemannian mean. Fig. 7 depicts the spatial patterns corresponding to the four largest eigenvalues from gaCSPb and CSP. Each pattern represents the source activities of signals acquired at different electrodes. The first two spatial patterns of gaCSPb are almost consistent with those of CSP, indicating that both methods can capture the most prominent variance direction. In contrast, in the latter two patterns, gaCSPb exhibits less interference compared to CSP, indicating that Riemannian mean has better anti-interference ability, resulting in cleaner and more informative spatial patterns.

B. gaCSP feature vs. CSP feature

Fig. 8 summarizes the difference in average accuracy between manifold-based features and variance-based features. It is observed that in MDRM classifier scenario, SPD matrix features extracted by gaCSP have better performance than variance features extracted by CSP. The following analyzes the differences between CSP features and geometric features in principle.

Let ω be the matrix holding the first m/2 and last m/2 eigenvectors of $\Sigma_1^{-1}\Sigma_2$, such that $\omega^T\Sigma_1\omega = \Lambda_1$ and $\omega^T\Sigma_2\omega = \Lambda_2$, where Λ_1 and Λ_2 are diagonal matrices. Given two arbitrary covariance matrices, after filtering with ω, results in $\omega^T C_1^i \omega = D_1^i$ and $\omega^T C_2^j \omega = D_2^j$, with diagonal elements denoted as $d_1^{i,p}$ and $d_2^{j,p}$, respectively. The diagonal elements are the variances of the EEG trials filtered by ω, and their logarithm values are the variance features obtained by (5). Therefore, the Euclidean distance between two variance-based features can be expressed as $\sum_p^m (\log d_1^{i,p} - \log d_2^{j,p})^2$. Since the Riemannian distance ($\delta_R^2(D_1^i, D_2^j)$) between D_1^i and D_2^j is a function of the eigenvalues of $(D_1^i)^{-1}D_2^j$ as per (8), the distance of geometric features are associated with the extreme eigenvalues of the CSP features. However, matrices D_1^i and D_2^j are not diagonal, thus their diagonal elements are not their eigenvalues, resulting in:

$$\sum_p^m (\log d_1^{i,p} - \log d_2^{j,p})^2 \le \delta_R^2(D_1^i, D_2^j), \quad (22)$$

The equality only holds when D_1^i and D_2^j are diagonal matrices. The inequality suggests that the manifold features of gaCSP exhibit a greater inter-class distance compared to the variance features of CSP, which could potentially explain why the classification results of gaCSP on the MDRM classifier are better than those of the variance features. Additionally, gaCSP manifold features demonstrate enhanced robustness in the presence of noise or non-stationary factors.

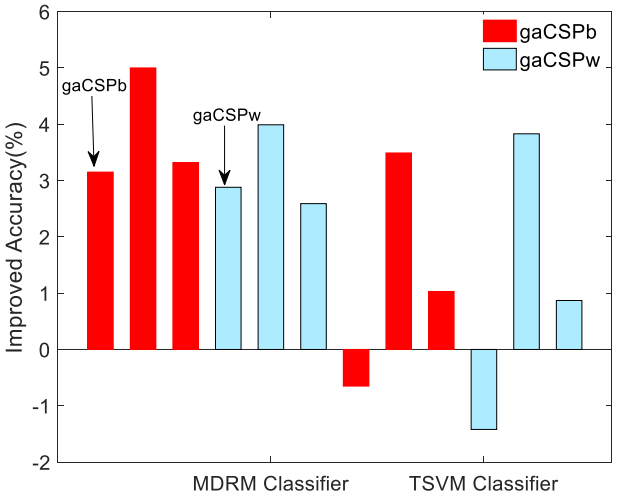


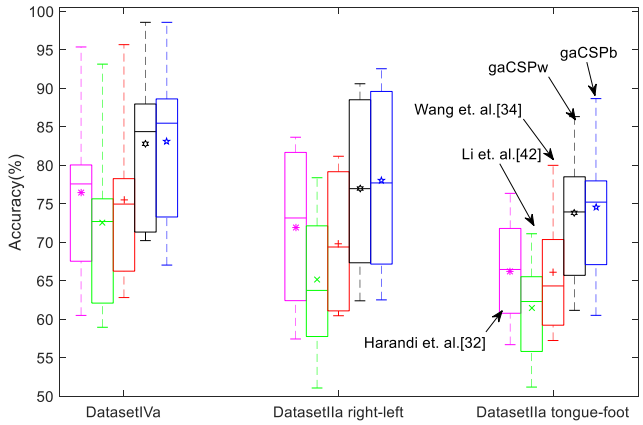
Fig. 8. Comparison of accuracy between gaCSP-based manifold

C. gaCSP vs. other Geometric Dimension Reduction Methods

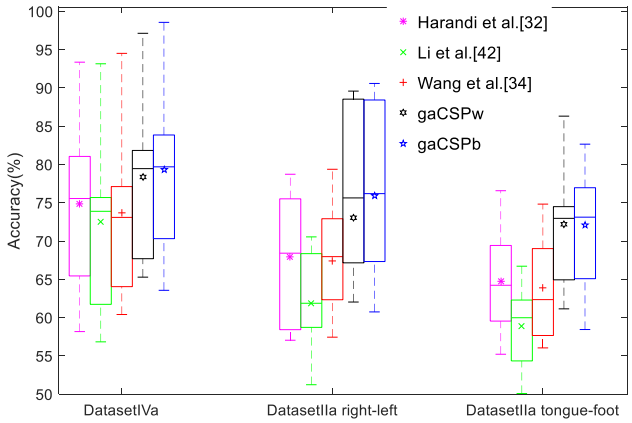
This section compares the performance of the proposed manifold dimensionality reduction frameworks, gaCSPb and gaCSPw, with other manifold dimensionality reduction algorithms mentioned in the literature. Harandi et al. [32] introduced a graph mapping dimensionality reduction on the SPD manifold with the aim to preserve the manifold structure of the original data in the submanifold, where the affinity matrix encodes the similarities of adjacent data. Li et al. [42] proposed a locality preserving projection on Grassmann manifold, aiming to preserve the local structure of data without considering labels, where the Grassmann manifold is spanned by a q-dimensional orthonormal matrix of the SPD manifold. Wang et al. [34] introduced the local

linear embedding into the Grassmann manifold to maintain local discriminant information in an adequate lower-dimensional feature space, where the involved affinity matrix encodes the information of the inter-class dispersion and the intra-class compactness. In comparison to these approaches, our proposed gaCSPb and gaCSPw frameworks consider global discriminant information using maximum inter-class distance rule and intra-class separability using maximum intra-class variance.

Fig.9 summarizes the performance of gaCSPb and gaCSPw with other dimensionality reduction algorithms. It is observed that both gaCSPb and gaCSPw consistently outperform other methods in all scenarios, with a significant difference (paired *t*-test $p<0.01$). Furthermore, it is noted that [42] performs poorly in all scenarios, possibly due to its unsupervised mechanism. On the other hand, [32] and [34] exhibit similar performance, and a paired *t*-test shows no significant difference. This similarity in performance may be attributed to the fact that they adopt similar objective functions and constraints, despite operating on different manifolds. It is worth mentioning that [32] reduces dimensionality by preserving the local neighborhood of samples on the Riemannian manifold, while gaCSPb and gaCSPw maximize inter-class dispersion from a global perspective.



(a) MDRM



(b) TSVM

Fig. 9 Classification accuracies achieved by several manifold dimensionality reduction algorithms with (a) MDRM, and (b) TSVM

D. Limitations

Based on the experimental results, it appears that the gaCSP methods enhance feature extraction performance by utilizing Riemannian metrics. However, there are still some limitations to be addressed. Firstly, the gaCSP algorithm essentially is a supervised feature extractor that relies on labeled data to estimate the inter-class mean. Although gaCSP exhibits better robustness than CSP, it does not completely eliminate the influence of noise or small training settings. Secondly, gaCSP generally requires more time for training compared to CSP, and gaCSPw takes even longer due to factors such as data dimensions, initial conditions, and optimization methods. The convergence time is also significantly affected by the chosen initializations. Strategies such as using spatial filters obtained by CSP as initial values can help reduce training time while still improving the performance of gaCSP. Lastly, previous studies have confirmed that the performance of CSP is influenced by the frequency band and time window of the EEG signal [40],[41] and the integration of CSP with frequency or time optimization schemes has significantly improved its performance. Therefore, our future work will focus on combining gaCSP with frequency-band and time-window selection schemes. To sum up, gaCSP can serve as an alternative to the CSP algorithm in scenarios where training time is not a significant concern.

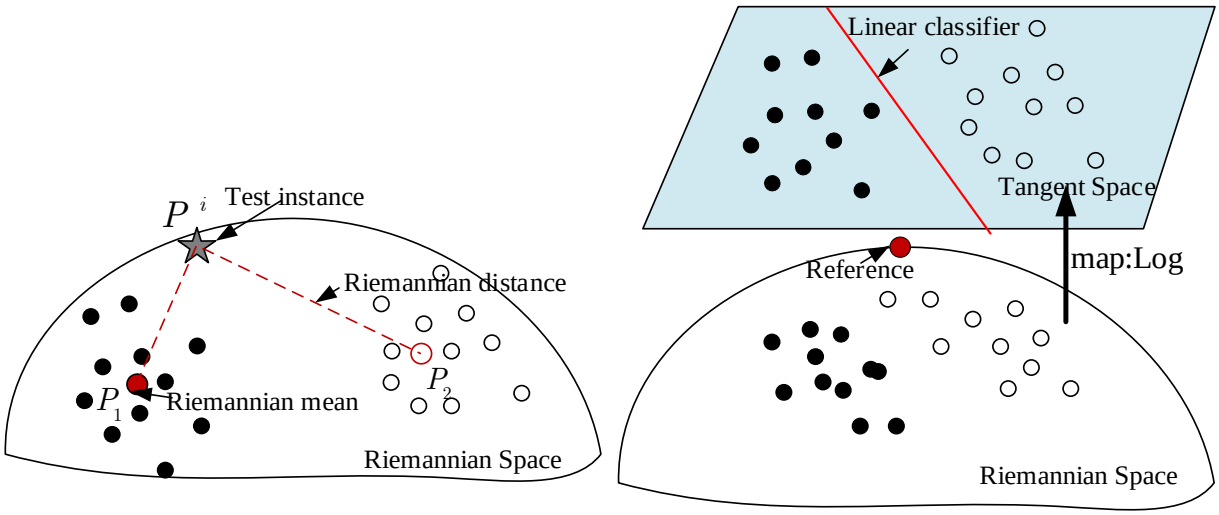
VI. Conclusions

This paper presents three geometry-aware CSP frameworks to enhance the spatial patterns and features extracted from EEG signals. gaCSPd jointly diagonalizes the inter-class Riemannian means by utilizing the robustness of the geometric mean to outliers, gaCSPb seeks an optimal projection matrix on the Riemannian manifold while maximizing the inter-class Riemannian distance, and gaCSPw extends spatial filtering to principal component analysis in Riemannian manifold, preserving discriminative information within classes while maximizing inter-class discrimination. Furthermore, the optimization problems of gaCSPb and gaCSPw are solved by an iterative generalized eigen-decomposition on the Stiefel manifold. To evaluate the effectiveness of gaCSP, a series of experiments have been conducted on two public EEG datasets, and gaCSP has been compared with several other competing methods. Experimental results indicate that gaCSP is a promising candidate for improving MI-based BCI performance. Specifically, 1) the projection matrix of gaCSP has a better anti-interference ability and better robustness. 2) The features of gaCSP, whether variance-based features or low-dimensional SPD matrix features, are more discriminative. 3) gaCSP is a better alternative to CSP.

Although the spatial patterns and features of gaCSP are significantly improved compared with CSP, gaCSP is also limited by the inherent limitations of CSP, such as overfitting on small training sets and being affected by signal frequency band and time window selection. To address these limitations, future work will focus on combining gaCSP with frequency band optimization, time window optimization, or prevalent regularization methods to explore the potential of gaCSP in reducing the calibration time and addressing the challenges of cross-subject.

ACKNOWLEDGMENT

The authors would like to thank the Barrier-free Service Robot Laboratory of Chongqing University of Posts and Telecommunications for its equipment and technical support.



(a) minimum distance to the Riemannian mean **(b)** linear classifier in tangent space

Fig. 2. The schematic diagram for MDRM and TSVM classifiers.

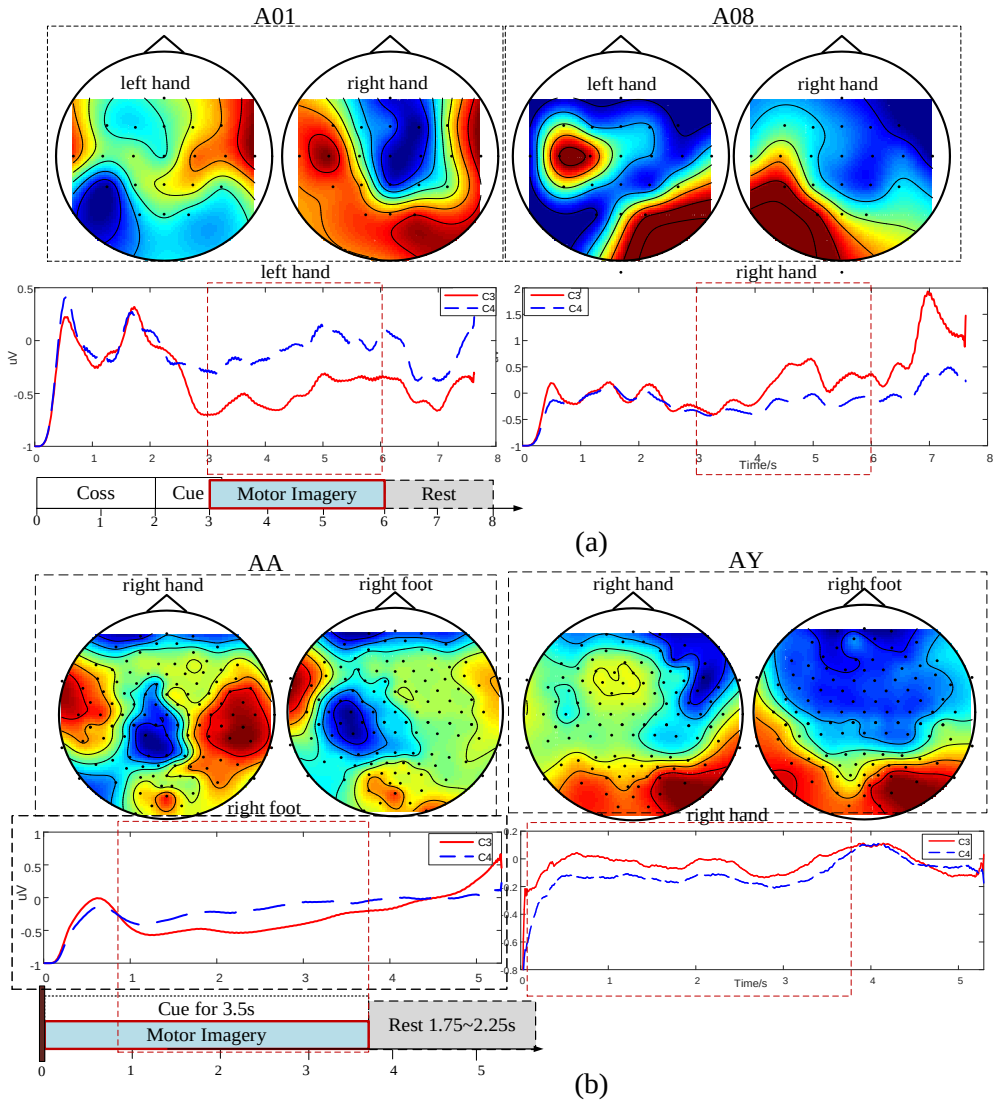


Fig. 3. ERD distributions and the timing schedules of the datasets, **(a)** ERD distribution and timing schedule of DatasetIVa;**(b)** ERD distribution and timing schedule of DatasetIIa.

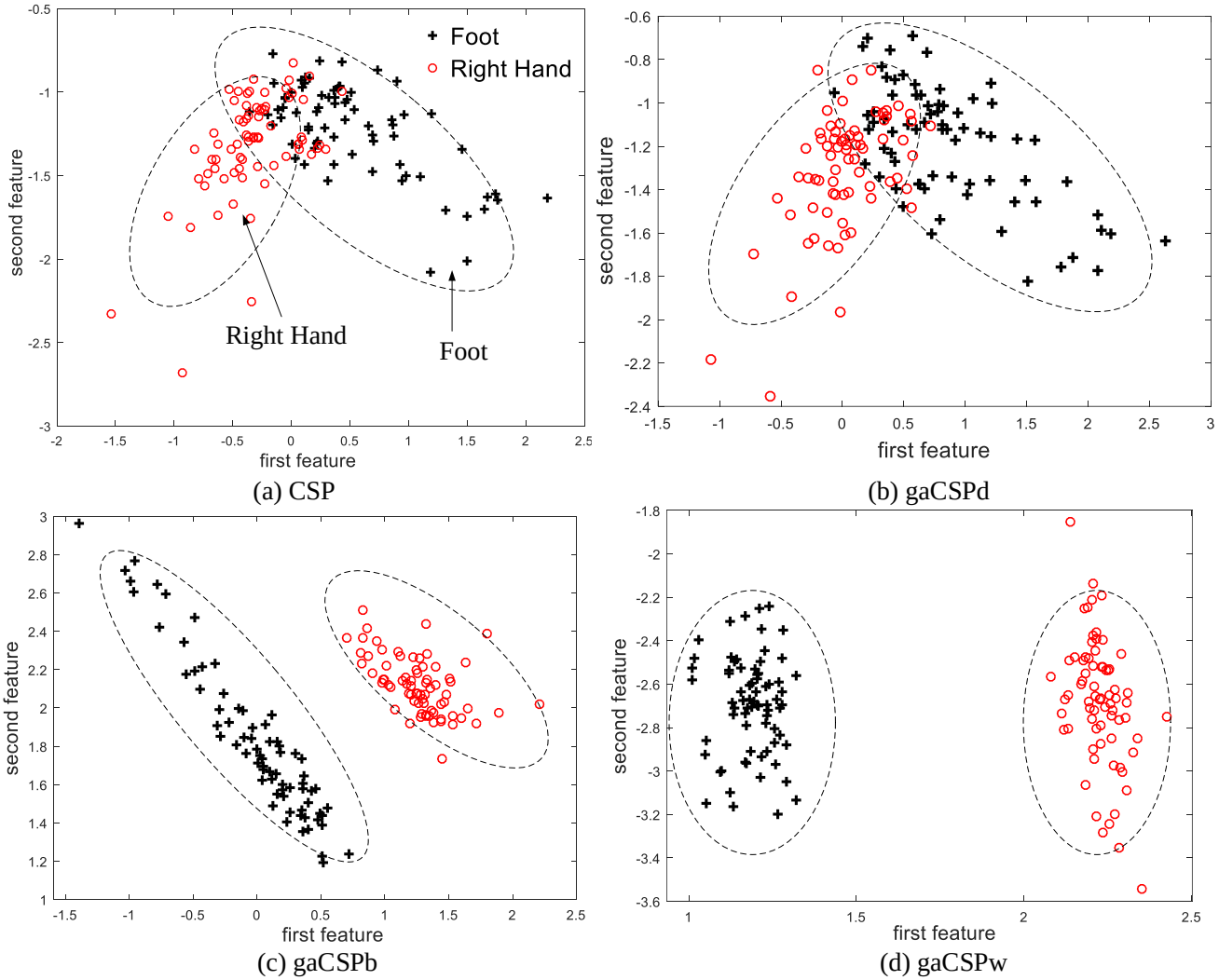


Fig. 4. *t*-SNE visualization of of the discriminative features learned by different CSP variants on DatasetIVa- subject 'AA'.

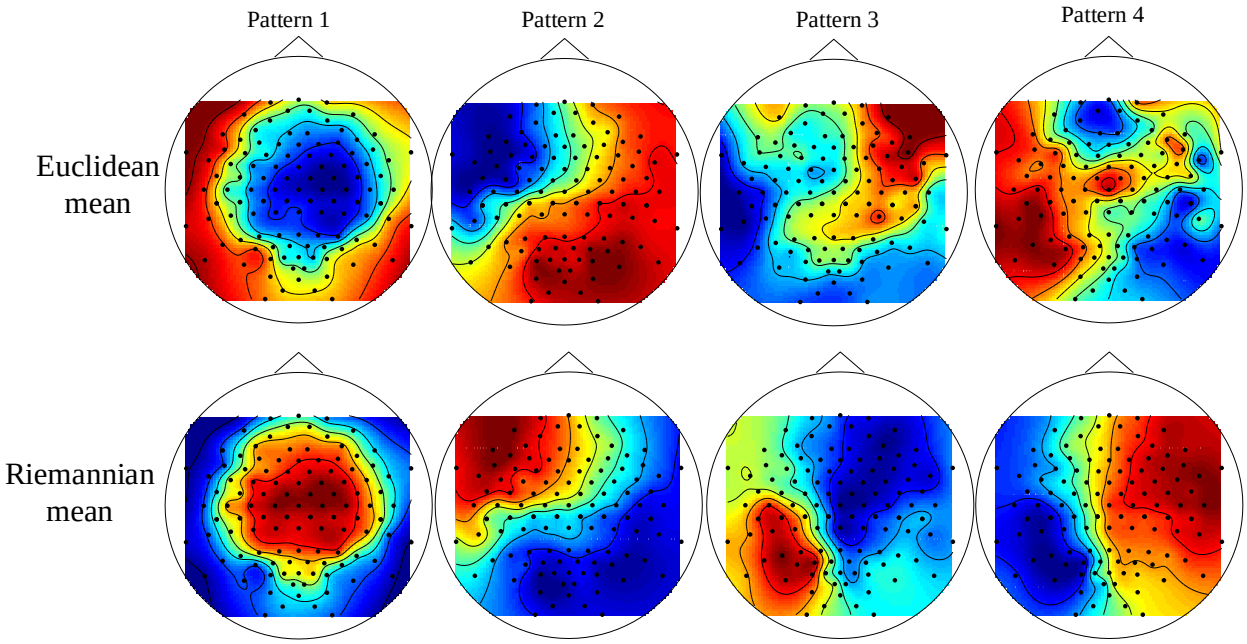


Fig. 7. The largest four spatial patterns learned by the CSP and gaCSP for the right hand/foot motor imagery data.

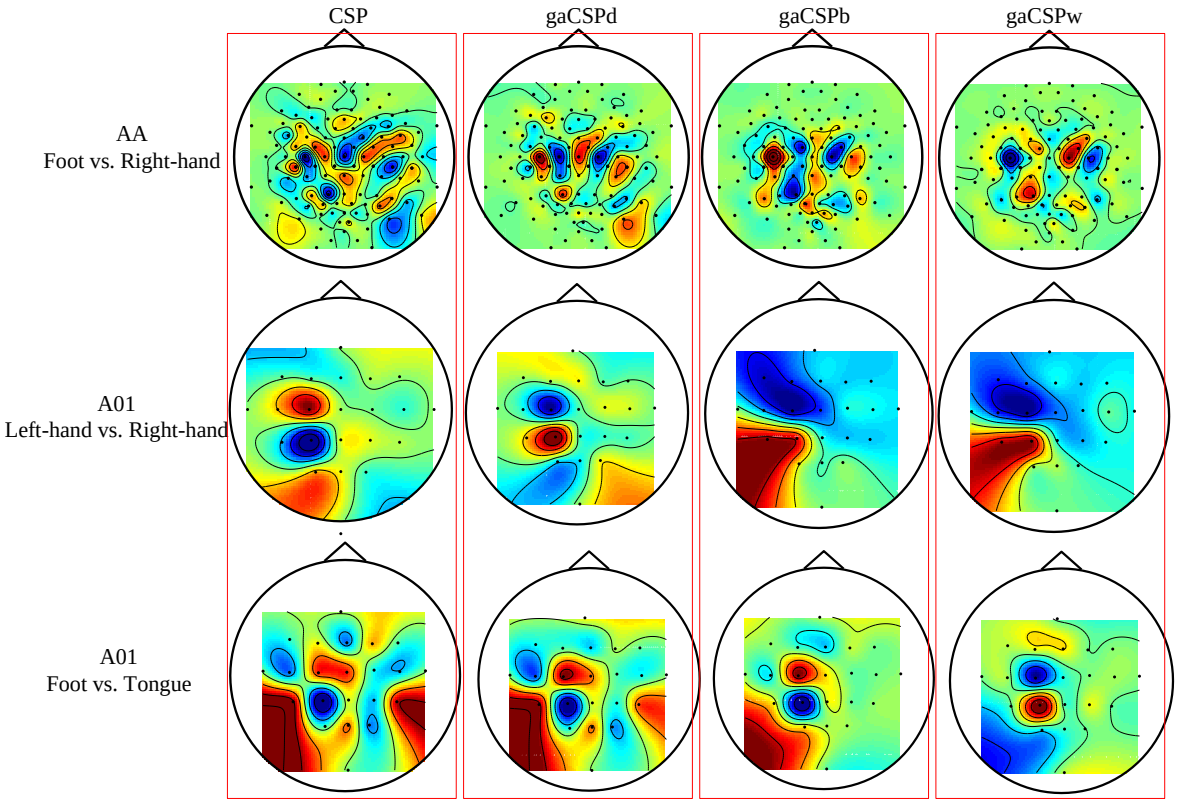


Fig. 5. Topographic maps corresponding to the largest spatial filter of the CSP and gaCSP methods.

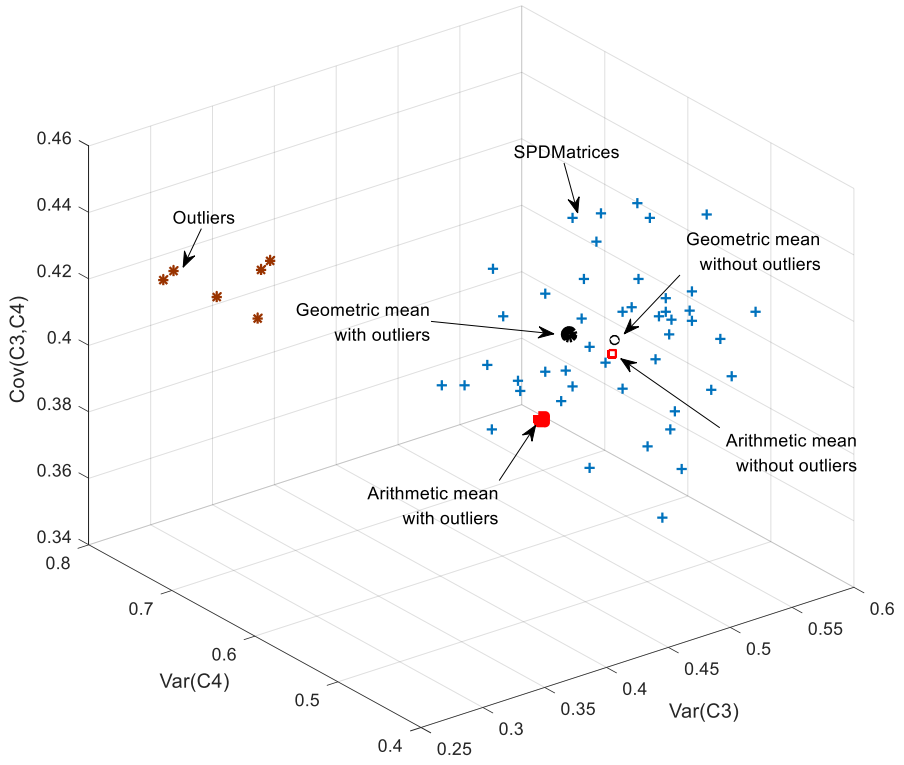


Fig. 6. The arithmetic and geometric means with (without) outliers for subject-AW.

REFERENCES

[1] Dipti Pawar, and Sudhir Dhage, "Feature Extraction Methods for Electroencephalography based Brain-Computer Interface: A Review," *IAENG International Journal of Computer Science*, vol. 47, no.3, pp501-515, 2020.

[2] Léa Pillette, Roc A, N'Kaoua B, et al, "Experimenters Influence on Mental-Imagery based Brain-Computer Interface User Training," *International Journal of Human-Computer Studies*, vol. 149, no.1, 102603, 2021.

[3] Chaudhary U, Mrachacz-Kersting N, Birbaumer N," Neuropsychological and neurophysiological aspects of brain-computer-interface (BCI) control in paralysis," *The Journal of physiology*, vol. 599, no.9, pp2351-2359, 2021.

[4] Zhang R, Li Y, Yan Y, et al. " Control of a Wheelchair in an Indoor Environment Based on a Brain–Computer Interface and Automated Navigation," *IEEE Transactions on Neural Systems & Rehabilitation*

Engineering A Publication of the IEEE Engineering in Medicine & Biology Society, vol.24, no.1, pp128-139, 2016.

[5] Flesher S N, Downey J E, Weiss J M, et al. "A brain-computer interface that evokes tactile sensations improves robotic arm control," *Science*, vol.372, no. 6544, pp 831-836, 2021.

[6] Yuriy M, Murat K, Erkan O, et al. "Developing a 3- to 6-state EEG-based brain-computer interface for a virtual robotic manipulator control," *IEEE Transactions on Biomedical Engineering*, vol.66, no.4, pp977-987, 2018.

[7] Chaisaen R, Autthasan P, Mingchinda N, et al. "Decoding EEG Rhythms During Action Observation, Motor Imagery, and Execution for Standing and Sitting," *IEEE Sensors Journal*, vol.20: 13776-13786, 2020.

[8] Mishuhina V and Jiang X. "Feature Weighting and Regularization of Common Spatial Patterns in EEG-Based Motor Imagery BCI," *IEEE Signal Processing Letters*, vol.25, no.6, pp783-787, 2018.

[9] Talukdar U, Hazarika S M, Gan J Q. "Adaptation of Common Spatial Patterns based on mental fatigue for motor-imagery BCI, " *Biomedical Signal Processing and Control*, vol.58, 101829, 2020.

[10] Wang B, Wong C M, Kang Z, et al. "Common Spatial Pattern Reformulated for Regularizations in Brain-Computer Interfaces, " *IEEE Transactions on Cybernetics*, vol.51, no.10, pp5008-5020, 2020.

[11] He H, Wu D. (2017, Oct.)"Transfer Learning Enhanced Common Spatial Pattern Filtering for Brain Computer Interfaces (BCIs): Overview and a New Approach, "(Online). Avilable: https://arxiv.org/pdf/1808.05853.pdf.

[12] Wu D, Jiang X, Peng R. "Transfer learning for motor imagery-based brain-computer interfaces: A tutorial," *Neural Network*, vol.153, pp 235-253, 2022.

[13] Lotte F, Guan C. "Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms," *IEEE Transactions on Biomedical Engineering*, vol.58, no.2, pp355-362, 2011.

[14] Samek W, Kawanabe M, Müller K. "Divergence-Based Framework for Common Spatial Patterns Algorithms," *IEEE Reviews in Biomedical Engineering*, vol.7, pp50-72, 2014.

[15] Arvaneh M, Guan C, Ang K.K, et al. "Optimizing the channel selection and classification accuracy in EEG-based BCI," *IEEE Transactions on Biomedical Engineering*, vol.58, no.6, pp1865-1873, 2011.

[16] Mohammad N C, Homa K A, Mohammad R D. "Ensemble Regularized Common Spatio-Spectral Pattern (Ensemble RCSSP) Model for Motor Imagery-based EEG Signal Classification," *Computers in Biology and Medicine*, vol.135, 104546, 2021.

[17] Navarro-Sune X, Hudson A L, De Vico Fallani F, et al. "Riemannian Geometry Applied to Detection of Respiratory States from EEG Signals: The Basis for a Brain–Ventilator Interface," *IEEE Transactions on Biomedical Engineering*, vol.64, no.5, pp1138-1148, 2017.

[18] Yger F, Berar M, Lotte F. "Riemannian Approaches in Brain-Computer Interfaces: A Review," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol.25, no.10, pp1753-1762, 2017.

[19] Barachant A, Bonnet S, Congedo M, et al. "Multiclass Brain–Computer Interface Classification by Riemannian Geometry," *IEEE Transactions on Biomedical Engineering*, vol.59, no.4, pp 920-928, 2012.

[20] Barachant A, Bonnet S, Congedo M, et al. "Riemannian geometry applied to BCI classification, " Lecture Notes in 2010 International Conference on Latent Variable Analysis and Signal Separation, Sep. 27-30, 2010, St. Malo, France, pp629–636.

[21] Chevallier S, Kalunga E K, Barthélemy Q, Monacelli E. "Review of Riemannian Distances and Divergences, Applied to SSVEP-based BCI, " *Neuroinformatics*, vol.19, no.1, pp93-106, 2021.

[22] Kalunga E K, Chevallier S, Barthélemy Q, et al. "Online SSVEP-based BCI using Riemannian Geometry, "*Neurocomputing*, vol.191, pp 55-58, 2016.

[23] Korczowski L, Congedo M, Jutten C. "Single-trial classification of multi-user P300-based Brain-Computer Interface using Riemannian geometry, " 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015, pp1769-1772.

[24] Xie X, Yu Z L, Gu Z, et al. "Bilinear Regularized Locality Preserving Learning on Riemannian Graph for Motor Imagery BCI, " *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol.26, no.3, pp 698-708, 2018.

[25] Xie X, Yu Z L, Gu Z, et al. "Classification of symmetric positive definite matrices based on bilinear isometric Riemannian embedding, " *Pattern Recognition: The Journal of the Pattern Recognition Society*, vol.87, pp 94-105, 2019.

[26] Horev I, Yger F, Sugiyama M. "Geometry-Aware Principal Component Analysis for Symmetric Positive Definite Matrices," *Machine Learning*, vol.106, pp493–522, 2017.

[27] Wu D, Xu Y, Lu B-L. "Transfer Learning for EEG-Based Brain–Computer Interfaces: A Review of Progress Made Since 2016, "*IEEE Transactions on Cognitive and Developmental Systems*, vol.14, no.1, pp4-19, 2022.

[28] Zanini P, Congedo M, Jutten C, et al. "Transfer Learning: A Riemannian Geometry Framework with Applications to Brain–Computer Interfaces, " *IEEE Transactions on Biomedical Engineering*, vol.65, no.5, pp1107-1116, 2018.

[29] Yair O, Ben-Chen M, Talmon R. "Parallel transport on the cone manifold of SPD matrices for domain adaptation, " *IEEE Transactions on Signal Processing*, vol.67, no.7, pp 1797-1811, 2019.

[30] Rodrigues P L C, Jutten C, Congedo M. "Riemannian procrustes analysis: transfer learning for brain–computer interfaces, " *IEEE Transactions on Biomedical Engineering*, vol.66, no.8, pp2390-2401, 2018.

[31] Zhang W, Wu D. "Manifold embedded knowledge transfer for brain-computer interfaces, " *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol.28, no.5, pp1117-1127, 2020.

[32] Harandi M, Salzmann M, Hartley R. "Dimensionality Reduction on SPD Manifolds: The Emergence of Geometry-Aware Methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.40, no.1, pp48-62, 2018.

[33] Arsigny V, Fillard P, Pennec X, et al. "Geometric Means in a Novel Vector Space Structure on Symmetric Positive-Definite Matrices," *SIAM Journal on Matrix Analysis & Application*, vol.29, no.1,pp328-347, 2011.

[34] R. Wang, X. -J. Wu, Z. Liu and J. Kittler, "Geometry-Aware Graph Embedding Projection Metric Learning for Image Set Classification," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 3, pp. 957-970, Sept. 2022.

[35] Park S W, Kwon J. "Riemannian submanifold framework for log-Euclidean metric learning on symmetric positive definite manifolds, "*Expert Systems with Application*, vol.202, 117270, 2022.

[36] Liu T, Shi Z, Liu Y. "Joint Normalization and Dimensionality Reduction on Grassmannian: A Generalized Perspective,"*IEEE Signal Processing Letters*, vol.25, no.6, pp 858-862, 2018.

[37] Congedo M, Barachant A, Bhatia R. "Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review," *Brain-Computer Interfaces*, vol.4, no.3, pp155-174, 2017.

[38] Shariat A, Zarei A, Karvigh S A, et al. "Automatic detection of epileptic seizures using Riemannian geometry from scalp EEG recordings, " *Medical & Biological Engineering & Computing*, vol.59, no.7, pp1431-1445, 2021.

[39] Jiang Q, Zhang Y, Zheng K. "Motor Imagery Classification via Kernel-Based Domain Adaptation on an SPD Manifold," *Brain Sciences*, vol.12, pp 659-681, 2022.

[40] Jiao Y, Zhang Y, Chen X, er al. "Sparse Group Representation Model for Motor Imagery EEG Classification. *IEEE Journal of Biomedical and Health Informatics*," vol.23, no.2, pp631-641, 2019.

[41] Zhang Y, Zhou G, Jin J, er al. "Temporally Constrained Sparse Group Spatial Patterns for Motor Imagery BCI," *IEEE Transactions on Cy*bernetics, vol.49, no.9, pp3322-3332, 2019.

[42] Li, Yangyang, and Ruqian Lu. "Locality preserving projection on SPD matrix Lie group: algorithm and analysis," *Science China Information Sciences*, vol.61, pp 1-15, 2018.