CovMix: Covariance Mixing Regularization for Motor Imagery Decoding

Georgios Zoumpourlis
School of Electronic Engineering and Computer Science
Queen Mary University of London
g.zoumpourlis@qmul.ac.uk

Ioannis Patras
School of Electronic Engineering and Computer Science
Queen Mary University of London
i.patras@qmul.ac.uk

***Abstract*—In this paper we study the problem of motor imagery (MI) decoding using electroencephalography (EEG) signals. The spatial covariance matrix of EEG signals is a feature with many applications on brain-computer interfaces. Several previous works use EEG covariances directly as inputs to Riemannian classifiers for MI decoding, restricting the potential models that can be used for classification, to Riemannian geometry frameworks. Other works use covariances either as inputs to optimization objectives that derive spatial filters, or to perform alignment with respect to reference states. Such methods discard temporal information that is contained in EEG signals. We take a different approach, and utilize covariances as a means to concurrently align EEG signals and regularize a Convolutional Neural Network (CNN) that is trained on MI classification. Specifically, we randomly mix session-level and trial-level covariance matrices, traversing their geodesic on the Riemannian manifold, and perform EEG signal alignment using the mixed matrix. This is done during the training phase, effectively acting as regularization on the CNN model, as the signals are augmented using various transformation matrices to align them. We evaluate our method on the dataset of BCI Competition IV-2a, showing its superiority over traditional alignment.**

***Index Terms*—brain-computer interface, electroencephalography, motor imagery**

I. Introduction

Nowadays, the technological developments in the field of Brain-Computer Interfaces (BCI) are rendering the acquisition of signals that measure brain activity through Electroencephalography (EEG) increasingly accessible to researchers and consumers. Many problems are addressed through the analysis of EEG signals with machine learning techniques, such as BCI-controlled wheelchairs [12], automatic sleep staging [6] and predicting stroke patient rehabilitation outcome [18]. Among the most well-known EEG-based paradigms for BCIs, is that of motor imagery (MI). In the MI decoding problem, the goal is to analyze EEG signals to predict the motor movement type that a subject imagines (e.g. moving the left hand, right hand, feet or tongue). The imagination of motor movements is closely related to brain activity patterns of the sensorimotor cortex [8], indicating the need for methods with powerful spatial localization to detect them.

The work of Georgios Zoumpourlis was supported by QMUL Principal's Studentship.

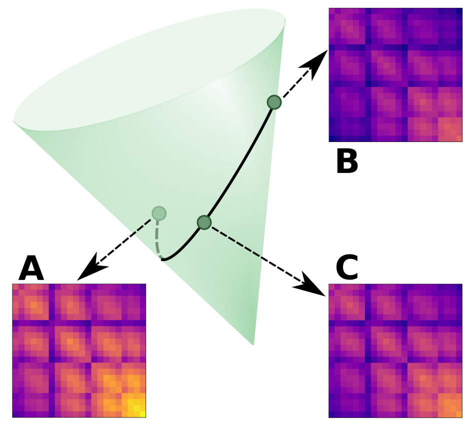

Fig. 1: Mixing two covariance matrices, by traversing their geodesic on the Riemannian manifold. The point corresponding to matrix $\mathbf{C}$, lies on the shortest path that connects $\mathbf{A}$ and $\mathbf{B}$.

Early works on MI decoding include spatial filtering techniques, such as Common Spatial Patterns (CSP) [5], [14], where the variances of the filtered signals are maximized/minimized over certain conditions (i.e. classes). CSP methods involve computing the average covariance matrix of the EEG signals for each class and jointly diagonalizing them. Handcrafted feature vectors are extracted from the spatially filtered signals, collapsing the temporal aspect by computing the variance in the dimension of time. Then, typical classifiers such as Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) [22] are used. However, the family of CSP methods presents poor cross-subject generalization and is restricted by the discarding of temporal information that occurs during feature extraction.

Instead of using EEG covariances to compute spatial filters, another line of research inspired by Riemannian geometry, directly uses them as features. Covariance matrices lie on the Riemannian manifold of symmetric positive definite (SPD) matrices, hence should be treated accordingly, refraining from methods designed for Euclidean space data. Applying Riemannian classifiers on covariance matrices, or projecting

978-1-6654-1337-4/22/$31.00 ©2022 IEEE

covariance matrices to the Euclidean space through tangent space mapping [3], are two common approaches. Despite their widespread use, Riemannian geometry frameworks are prone to outliers due to the non-stationarity of EEG signals [25] and do not allow temporal information extraction.

Training Convolutional Neural Networks (CNN) as feature extractors towards MI classification, has been explored as an alternative direction, instead of computing handcrafted features from EEG signals. CNN models present strong performance both in intra-subject [19] and cross-subject [16] settings of MI decoding. Training CNNs on EEG time-series has facilitated learning more robust representations, allowing to filter the signals both temporally and spatially, in separate stages of the feature extraction process. In this context, EEG covariance matrices are used to perform data alignment [10], [26], transforming the EEG signals. As a result, the covariance matrix of the aligned EEG signals within each session (i.e. domain) is the identity matrix, which can be viewed as a domain generalization technique [15], suitable for cross-subject transfer learning.

Our work is inspired from [1] and [15], and focuses on network regularization. A common way to regularize neural networks is data augmentation, generating new data points either on the input space [27] or on the intermediate feature space [23]. Our goal is to train a CNN architecture on motor imagery classification from EEG time-series inputs, *combining* alignment and regularization during training. This is an unexplored research direction. Clearly, in such a case alignment needs to be done *before* feeding the EEG signals to the CNN, i.e. on the input space.

Building on the benefits of CNN-based learning and covariance-based alignment, we introduce a method called "CovMix", that mixes session-wise and trial-wise covariances to perform data augmentation on EEG time-series during training. We suggest an alignment process during training, where we align each trial choosing *randomly* an SPD matrix that lies on the geodesic between the session-wise and the trial-wise covariance matrices. This is in contrast to the standard practice of alignment, where the session-wise covariance matrix is used to transform all the EEG trials of each session, in both training and testing phases. Afterwards, the aligned trials are fed to a CNN model that is trained to classify them. Along the training process, the CNN model will receive each trial as input multiple times, yet aligned using different SPD matrices - although all of which will be lying on the same geodesic. Effectively, this regularizes the CNN model by feeding it with various transformations of each training sample. Inference is performed using the standard alignment, to keep the process being deterministic.

Our contributions are summarized as follows:

- We propose CovMix, a method that mixes session-wise and trial-wise covariance matrices to jointly perform EEG signal alignment and data augmentation during training.
- CovMix is performed before feeding the data to the classification network, thus can me used in any method that employs CNNs for motor imagery decoding .
- We evaluate networks trained on BCI Competition IV-2a dataset [20] with cross-subject settings, showing that adding CovMix acts as regularization to the classification network, yields stronger generalization results compared to the standard covariance-based alignment and other techniques.

The rest of the manuscript is organized as follows. In Section II we present an overview of the related work, while in Section III we illustrate our proposed approach. In Section IV we describe our evaluation setup and in Section V we present results of our method on cross-subject experiments and compare it with other techniques. Finally, in Section VI we conclude the paper.

II. Related work

In this Section, we present an overview of previous works for motor imagery decoding in cross-subject settings. Developing MI decoding models that are suitable for real-world scenarios, requires being able to generalize on new sessions of known subjects, as well as on entirely unseen subjects. Several factors vary across sessions and subjects, including equipment-related changes (e.g. EEG electrode positions/impedances), changes in the mental state of subjects and different head anatomies. Thus, several works in the recent literature [2], [28] treat each session of EEG recordings as a separate data domain, and cross-subject scenarios can be viewed a multi-source domain generalization problem [15].

Non-DL covariance-based works: Data scarcity is an important issue in the field of EEG-based learning, as EEG data collection is a non-trivial process. To address this problem, a method of data augmentation on the Riemannian space is introduced in [13], by interpolating on the log-Euclidean geodesic between trial-wise covariance matrices. Among the early covariance-based works that tackle MI decoding by attempting to learn from multiple subjects, are variants of the CSP algorithm. Lotte and Guan [17] propose incorporating information from multiple subjects to regularize the covariance matrices used in a CSP framework. However, for each target subject the method requires selecting a subset of relevant source subjects. Another work that builds on CSP, proposes weighting source subjects based on their similarity to the target subject [7].

Covariance-based alignment: When dealing with EEG data, a common way to achieve generalization is to perform covariance-based alignment. In the work of Zanini *et al.* [26], the trial-wise covariance matrices of each session are re-centered (i.e. aligned) with respect to a reference covariance matrix that corresponds to the resting state. Ideally, each re-centered covariance should reflect only the shift from the reference state, that is induced by a task (e.g. the imagination of a motor movement). Conforming to the nature of the data, a Minimum Distance to Mean (MDM) Riemannian classifier is employed. In [10], He and Wu examine several issues, including: i) applying alignment on the EEG signals instead of the covariance matrices ii) computing the reference covariance matrix using the Euclidean mean instead of the Riemannian

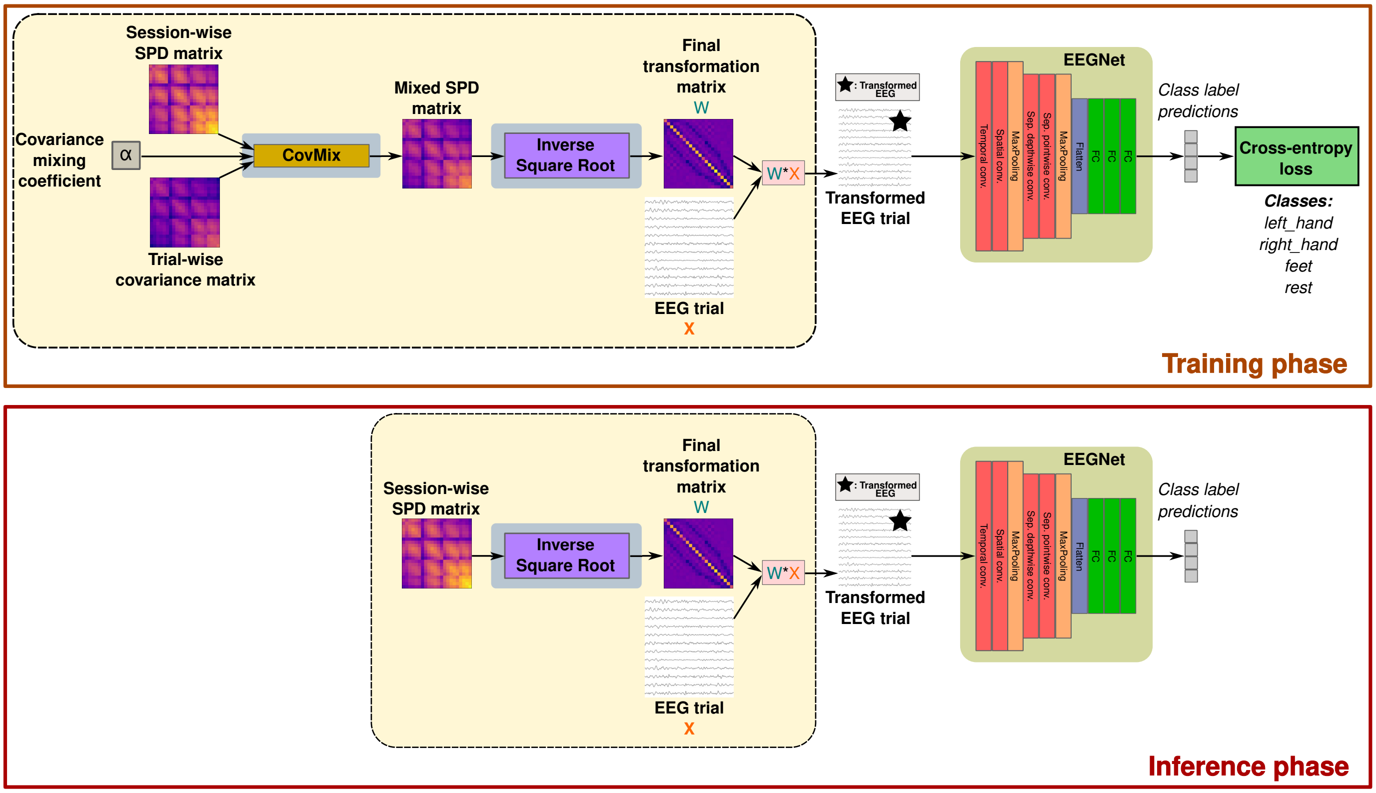

Fig. 2: Overview of our proposed method: Considering an EEG trial $\mathbf{X}$ that belongs to an EEG recording session $\mathbf{S}$ of the training set, we want to transform it using CovMix before feeding it to the classification model of EEGNet. To derive the transformation matrix $\mathbf{W}$, we need to compute the trial-wise covariance matrix of $\mathbf{X}$, and the session-wise SPD matrix (i.e. the Riemannian mean of all trial-wise covariances for session $\mathbf{S}$). CovMix mixes these two matrices following Eq. 4, using the scalar α that is sampled from a uniform distribution. Then, to perform the alignment, the signals are multiplied with the inverse square root of the mixed matrix, using Eq. 5, where the inverse square root of the session-wise SPD matrix is used). Finally, the transformed EEG signals are fed to EEGNet to be classified. In the inference phase we do not mix covariance statistics, and alignment is performed using Eq. 3.

mean and iii) computing the reference matrix using the imagery trials instead of the resting states. Their results show that superior results can be obtained by aligning the EEG signals and estimating the reference matrix from the imagery trials. Moreover, the restriction of employing Riemannian classifiers is overcomed.

Deep learning approaches: Banville *et al.* [1] adapt the concept of covariance-based spatial filtering, to a deep learning pipeline called "Dynamic Spatial Filtering" (DSF). In essence, before feeding an EEG trial to a task-specific CNN, the authors propose transforming the trial using deep learned spatial filters, to ignore bad channels and promote robustness to noise. The transformation is computed from the DSF module that receives the trial covariance as input. In [15], various techniques are investigated to improve generalization, including label smoothing [11], covariance-based Euclidean Alignment (EA) [10] and MixUp [27]. The technique of EA is helpful for cross-domain knowledge transfer, as it projects the data into a domain-invariant space. Their experiments show that while EA is beneficial, MixUp seems to have hardly any positive impact on the achieved accuracy, while in several cases it is detrimental to the learning process. This is in line with the findings of [9], where MixUp does not yield performance improvements. Summing up, deep learning methods can achieve strong performance on motor imagery tasks and enable training on multiple source domains.

III. Proposed Method

Let $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ be the set of n band-pass filtered EEG trials that a recording session $\mathbf{S}$ contains. Let $\mathbf{X}_i \in \mathbb{R}^{C \times T}$ be the i-th EEG trial of the session, where C is the number of EEG channels and T is the number of samples in the dimension of time. The covariance matrix $\mathbf{P}_i$ of trial $\mathbf{X}_i$ is calculated as $\mathbf{P}_i = \mathbf{X}_i \mathbf{X}_i^T$. Covariance matrices lie on the space of SPD matrices with dimension C, denoted as $\mathcal{P}(C)$, which is a Riemannian manifold. Considering two points (i.e. two SPD matrices) $\mathbf{P}_1$ and $\mathbf{P}_2$ on $\mathcal{P}(C)$, the Riemannian distance metric [4] is defined in Eq. 1:

$$\delta(\mathbf{P}_1, \mathbf{P}_2) = \left(\sum_{i=1}^{n} \log^2 \lambda_i \right)^{\frac{1}{2}} \quad (1)$$

where λ_i are the eigenvalues of $\mathbf{P}_1^{-1}\mathbf{P}_2$. The Riemannian mean $\overline{\mathbf{P}}$ of the set $\{\mathbf{P}_1, \ldots, \mathbf{P}_n\}$, is the SPD matrix that acts as the reference state for the entire session. There is no closed-form solution for the computation of $\overline{\mathbf{P}}$, thus it is solved as an optimization problem, i.e. minimizing the Riemannian distance between $\overline{\mathbf{P}}$ and all the trial-wise covariances, according to Eq. 2:

$$\overline{\mathbf{P}} = \underset{\mathbf{P} \in \mathcal{P}(C)}{\operatorname{argmin}} \sum_{i=1}^{n} \delta(\mathbf{P}_i, \mathbf{P}) \tag{2}$$

Riemannian Alignment: Typically, alignment [10], [24] on EEG signals is performed separately within each session, applying the same session-specific transformation to all trials. Considering the session-wise SPD matrix $\overline{\mathbf{P}}$ and the trial-wise signals $\mathbf{X}_i$, the aligned signals $\hat{\mathbf{X}}_i$ in this paper are computed as:

$$\hat{\mathbf{X}}_i = \overline{\mathbf{P}}^{-\frac{1}{2}} \mathbf{X}_i \tag{3}$$

CovMix Alignment: We propose an alternative method of alignment, that transforms each trial differently over multiple training steps. To provide diversity to the transformation matrix, we derive it by traversing the geodesic connecting the reference state (i.e. the session-wise SPD matrix $\overline{\mathbf{P}}$) and the trial-wise covariance matrix $\mathbf{P}_i$ (instead of deriving it directly from $\overline{\mathbf{P}}$). Having the reference state as starting point, ensures that the finally picked points are still relevant to the session-wise statistics. Picking points by traversing geodesics, ensures that our method respects the Riemannian nature of covariance matrix space.

We mix (i.e. interpolate) the session-wise SPD matrix $\overline{\mathbf{P}}$ with that of the i-th trial $\mathbf{P}_i$, obtaining the mixed SPD matrix $\mathbf{P}_{\text{mix}}$. An illustration of performing interpolation on the Riemannian manifold is provided in Fig. 1. We use a scalar $\alpha, 0 \leq \alpha \leq 1$, which we call covariance mixing coefficient, to control the distance between $\mathbf{P}_{\text{mix}}$ and $\overline{\mathbf{P}}$. More specifically, we sample α from a uniform distribution $\mathcal{U} \sim [0, 1]$, and compute the weighted Riemannian average [4] between matrices $\overline{\mathbf{P}}$ and $\mathbf{P}_i$ as shown in Eq. 4, so that $\delta(\mathbf{P}_{\text{mix}}, \overline{\mathbf{P}}) = \alpha \cdot \delta(\mathbf{P}_i, \overline{\mathbf{P}})$.

$$\mathbf{P}_{\text{mix}} = \overline{\mathbf{P}}^{\frac{1}{2}} \left(\overline{\mathbf{P}}^{-\frac{1}{2}} \mathbf{P}_i \overline{\mathbf{P}}^{-\frac{1}{2}} \right)^{\alpha} \overline{\mathbf{P}}^{\frac{1}{2}} \tag{4}$$

We can control the regularization induced to the classification network by CovMix, by restricting the range of values that are sampled for α. To do so, we use the hyperparameter α_{max} to set the maximum value of α, and sample from the distribution $\mathcal{U} \sim [0, \alpha_{\text{max}}]$. CovMix is applied *only* during the training phase, similarly to data augmentation methods, aligning the signals as follows:

$$\hat{\mathbf{X}}_i = \mathbf{P}_{\text{mix}}^{-\frac{1}{2}} \mathbf{X}_i \tag{5}$$

During the inference phase, we apply the Riemannian alignment and transform the signals using Eq. 3.

Let us note that our transformation matrix is not an arbitrary deep learned matrix, but an SPD matrix obtained by traversing on particular geodesics of the Riemannian manifold, that connect session-wise and trial-wise covariances. This ensures that the transformation matrix is *close* to the reference matrix for each domain, to facilitate training on multiple source domains as in [15]. Moreover, our transformation matrix does not aim to suppress noise on EEG trials, as [1] does, but to generate diverse augmentations of trials. Having done the covariance mixing in the Riemannian space, we perform data augmentation in the Euclidean space of EEG time-series. This allows us to use classifiers such as CNN models, unlike the method of [13] that generates augmented data points that require Riemannian classifiers.

CNN architecture: We opt to use a modified version of EEGNet [16] as our CNN architecture for motor imagery classification. Specifically, we remove the batch normalization layer at the temporal convolution stage, and use 3 fully connected (FC) layers at the classification head.

The pipeline of performing CovMix, along with the components of the CNN architecture, are shown in Fig. 2.

IV. Evaluation Setup

Dataset: We apply our method on the motor imagery decoding problem, studying the dataset of BCI Competition IV Dataset 2a (IV-2a) [20]. The dataset contains EEG recordings of 9 participants, collected over two different days for each subject (i.e. there are two sessions per participant), having 25 electrodes (22 EEG and 3 electrooculographic channels) and a sampling frequency of 250Hz. The classes of the dataset correspond to 4 different imaginary movement types that the subjects performed, namely left hand, right hand, feet and tongue. Each session contains 72 trials of each class. Our EEG preprocessing pipeline has the following steps: 1) bringing the EEG signals into the measurement unit of uV (microvolts) 2) keeping only the channels of 22 EEG electrodes, discarding the EOG electrodes 3) notch filtering to remove the 50Hz component 4) bandpass filtering in the range 4-38 Hz and 5) resampling signals to 100Hz. We crop the temporal window $[0.0, 4.0]$ seconds of each trial, where $t = 0$ is the event onset. The size of each input sample is $C \times T$, where C=22 (number of EEG channels) and T=400 (number of time samples). Evaluation is performed in a Leave-One-Subject-Out (LOSO) manner, using both sessions for all subjects.

Comparison to other methods: We compare CovMix with three other methods, namely Riemannian Alignment (RA) [10], MixUp [27] and MixStyle [29]. We implement RA as in [24], using the Riemannian mean of covariances and transforming the EEG signals instead of re-centering the covariances. For MixUp, considering two data samples $\mathbf{x}_i$, $\mathbf{x}_j$ and their labels encoded as one-hot vectors $\mathbf{y}_i$, $\mathbf{y}_j$ we create the augmented samples as $\mathbf{x}' = \lambda \mathbf{x}_i + (1 - \lambda)\mathbf{x}_j$ and $\mathbf{y}' = \lambda \mathbf{y}_i + (1 - \lambda)\mathbf{y}_j$, where $\lambda \sim \text{Beta}(2.0, 2.0)$. We also evaluate the method of MixStyle, which is a state-of-the-art domain generalization technique that can be plugged in between CNN layers. For MixStyle, let $\mathbf{x}$ be a batch of samples, and $\tilde{\mathbf{x}}$ be the randomly shuffled version of $\mathbf{x}$ across the batch dimension. First, the feature statistics $\gamma_{\text{mix}} = \lambda\sigma(\mathbf{x}) + (1 - \lambda)\sigma(\mathbf{x})$ and $\beta_{\text{mix}} = \lambda\mu(\mathbf{x}) + (1 - \lambda)\mu(\mathbf{x})$ are computed, using the operators

TABLE I: Evaluation of several methods on the motor imagery classification problem, using the dataset of BCI Competition IV-2a. The dataset contains 9 participants, and columns S01-S09 correspond to the accuracy of LOSO evaluation on one participant each time. All the number reported in this table are the average of 3 runs.

Method	S01	S02	S03	S04	S05	S06	S07	S08	S09	Mean Acc. (%)
Baseline	68.98	38.02	71.47	45.02	36.34	38.54	45.19	67.99	69.50	53.45
Riemannian Alignment	75.80	35.36	81.25	49.13	45.08	44.09	49.24	70.08	71.35	57.93
MixUp	75.63	34.14	79.11	48.67	46.70	46.81	52.54	74.24	69.79	58.63
MixStyle	76.09	42.36	80.20	55.84	45.71	47.97	58.39	67.18	73.09	60.76
CovMix (default), $\alpha_{\max} = 1.0$	76.21	41.72	82.46	54.40	44.73	47.16	57.17	73.20	71.35	60.93
CovMix (best), $\alpha_{\max} = 0.7$	78.12	43.86	83.68	58.33	45.54	47.51	60.53	75.34	72.91	**62.87**

of $\mu(\cdot)$ (mean value) and $\sigma(\cdot)$ (standard deviation) along the temporal dimension, with $\lambda \sim \mathrm{Beta}(0.1, 0.1)$. Then, MixStyle is performed with a probability of 50% on training batches, computing $\mathbf{x}' = \gamma_{\mathrm{mix}} \frac{\mathbf{x}-\mu(\mathbf{x})}{\sigma(\mathbf{x})} + \beta_{\mathrm{mix}}$ and detaching the operators of $\mu(\cdot)$ and $\sigma(\cdot)$ from the gradient computation. Upon attempting to plug MixStyle in several convolutional stages of EEGNet, we find that using it after the third convolutional layer is the most effective choice, and report results with this setting. We also report results of a baseline EEGNet model trained without any signal aligning (mentioned as "Baseline"), to serve as a reference for evaluating the benefits of alignment.

V. Experimental results

Training hyperparameters: Batch size is set to 64 and training is conducted for 120 epochs. The cross-entropy loss is used as the criterion for MI classification, where the targets are 4 classes. Stochastic Gradient Descent is selected as the optimizer (momentum=0.9, weight decay=0.01). We set the dropout probability of EEGNet to 0.1.

Results: Table I shows the results of all methods on the IV-2a dataset. Compared to the baseline model that is trained without any alignment on the EEG signals, the Riemannian Alignment (RA) method provides an accuracy boost of $+4.48\%$ (from 53.45% to 57.93%). Training EEGNet with CovMix using the default setting ($\alpha_{\max} = 1.0$), further increases the performance by $+3\%$. The regularization method of MixUp improves the accuracy only by $+0.70\%$ compared to RA, while the domain generalization approach of MixStyle gives a more considerable increase of $+2.83\%$ over RA. Overall, CovMix yields the highest accuracy of 60.93% outperforming all other methods.

Considering the regularization effect of MixUp, we find it to be insufficient. One drawback of applying MixUp in EEG signals, is the existence of large inter-subject differences on the channel-wise statistics. Thus, directly mixing signals from different domains on the input space (i.e. the time-series) may be detrimental for multi-source training. An indirect solution to this issue, is to scale the signal values of each trial in the range $[-1, +1]$ as in [15], before applying MixUp. Nevertheless this scaling is not consistent within each session, as it depends on trial statistics. In contrast to MixUp that transforms batch samples on the input space, MixStyle is done on the space of intermediate layer features. Thus, it is less prone to cross-domain differences on the input space. Yet for MixStyle to be effective, it needs to be plugged in to early layers of CNNs, where features mostly reflect domain-related information (while late layers are expected to be increasingly related to class label information). Differently from MixUp and MixStyle, CovMix does not involve mixing information stemming from different domains. The results show that regularization can be effectively induced in an intra-domain manner.

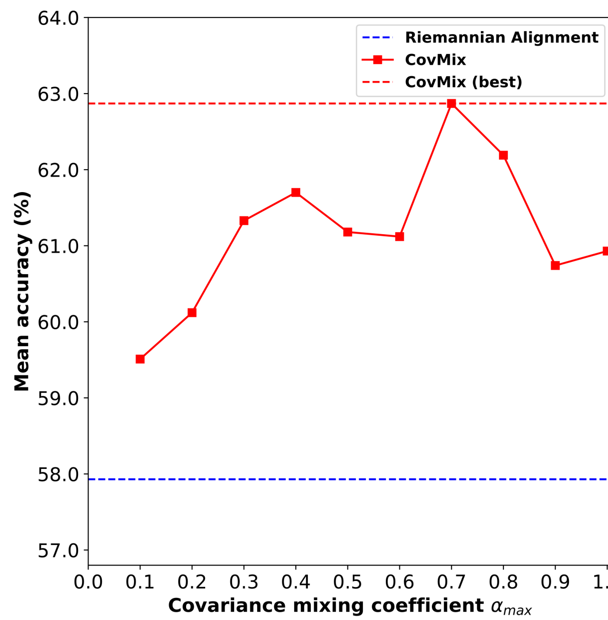

Fig. 3: The amount of regularization induced to the network by CovMix, is controlled through the hyperparameter $\alpha_{\max}$. The achieved performance obtained using CovMix, improves as we increase $\alpha_{\max}$ from 0.1 up to 0.7.

Ablation study: To examine the impact of hyperparameter $\alpha_{\max}$ on the performance of CovMix, we do an ablation study and run experiments setting $\alpha_{\max}$ from 0.1 to 1.0 with a step of 0.1. Smaller values of $\alpha_{\max}$ induce smaller regularization to the network. The results of our sweep are shown in Fig. 3. We observe that the performance of CovMix does not fall below that of RA, for any value of $\alpha_{\max}$. Tuning $\alpha_{\max}$ leads to even better performance compared to the default setting of $\alpha_{\max} = 1.0$, reaching a maximum accuracy of 62.87% when $\alpha_{\max} = 0.7$. The test subjects benefit differently from the values of $\alpha_{\max}$. In six out of nine subjects (specifically subjects 1, 3,

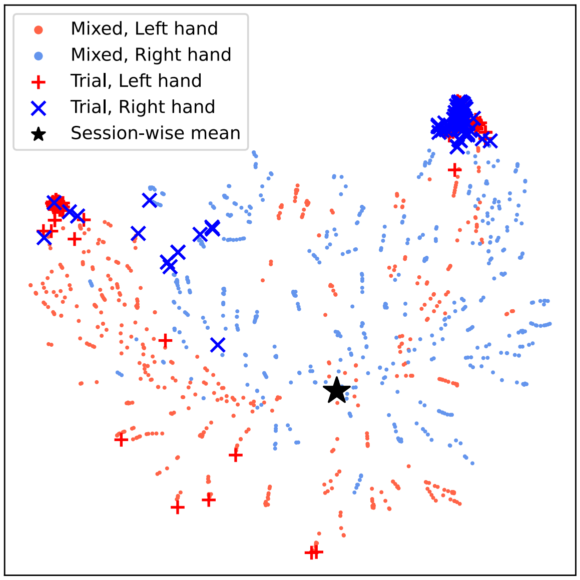

Fig. 4: Visualization of t-SNE embeddings from the trial-wise covariance matrices and the mixed SPD matrices that were obtained by performing random interpolations with CovMix. We use EEG signals from the second session of subject 9. Notice also the Riemannian barycenter of *all* trials (plotted with marker "★").

5, 7, 8 and 9), we achieve the highest test accuracies when setting $\alpha_{\max}$ between 0.6 and 0.8. However, for the rest three subjects (2, 4 and 6) the highest test accuracies occur when $\alpha_{\max}$ is in the range of 0.1 to 0.3.

Visualization of augmented SPD matrices: In Fig. 4, we provide a t-SNE [21] visualization of the covariances corresponding to trials, and the SPD matrices generated using CovMix with randomly sampled values of α. We can see that the points corresponding to interpolated SPD matrices, mainly occupy the space between the barycenter of the entire session (i.e. all trials from all classes), and the points of trial-wise covariance matrices.

VI. Conclusions

In this paper we discuss the problem of EEG-based MI decoding in transfer learning scenarios. Alternatively to methods that extract handcrafted features from EEG signal time-series, we use a CNN model as feature extractor. Through our proposed method, we concurrently perform alignment on the EEG signals and regularization on the CNN, applying different signal transformations during the training phase. We use a Riemannian framework to derive the transformation matrices, mixing trial-level and session-level covariance statistics. We conduct experiments on the BCI IV-2a dataset for MI classification, showing that CovMix performs superiorly against the traditional Riemannian Alignment, the regularization method of MixUp, and the domain generalization method of MixStyle. Our results indicate the potential of leveraging covariance-based alignment as a means towards regularization of deep neural networks.

References

[1] H. Banville, S. U. Wood, C. Aimone, D.-A. Engemann, and A. Gramfort. Robust learning from corrupted eeg with dynamic spatial filtering. *arXiv preprint arXiv:2105.12916*, 2021.

[2] G. Bao, N. Zhuang, L. Tong, B. Yan, J. Shu, L. Wang, Y. Zeng, and Z. Shen. Two-level domain adaptation neural network for eeg-based emotion recognition. *Frontiers in Human Neuroscience*, 14, 2020.

[3] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten. Multiclass brain–computer interface classification by riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, 2011.

[4] R. Bhatia. *Positive definite matrices*. Princeton University Press, 2009.

[5] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller. Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal processing magazine*, 25(1):41–56, 2007.

[6] S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, and A. Gramfort. A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(4):758–769, 2018.

[7] M. Cheng, Z. Lu, and H. Wang. Regularized common spatial patterns with subject-to-subject transfer of eeg signals. *Cognitive neurodynamics*, 11(2):173–181, 2017.

[8] M. Hamedi, S.-H. Salleh, and A. M. Noor. Electroencephalographic motor imagery brain connectivity analysis for bci: a review. *Neural computation*, 28(6):999–1041, 2016.

[9] D.-K. Han and J.-H. Jeong. Domain generalization for session-independent brain-computer interface. In *2021 9th International Winter Conference on Brain-Computer Interface (BCI)*, pages 1–5. IEEE, 2021.

[10] H. He and D. Wu. Transfer learning for brain–computer interfaces: A euclidean space data alignment approach. *IEEE Transactions on Biomedical Engineering*, 67(2):399–410, 2019.

[11] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li. Bag of tricks for image classification with convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 558–567, 2019.

[12] D. Huang, K. Qian, D.-Y. Fei, W. Jia, X. Chen, and O. Bai. Electroencephalography (eeg)-based brain–computer interface (bci): A 2-d virtual wheelchair control based on event-related desynchronization/synchronization and state control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(3):379–388, 2012.

[13] E. Kalunga, S. Chevallier, and Q. Barthélemy. Data augmentation in riemannian space for brain-computer interfaces. In *STAMLINS*, 2015.

[14] Z. J. Koles. The quantitative extraction and topographic mapping of the abnormal components in the clinical eeg. *Electroencephalography and clinical Neurophysiology*, 79(6):440–447, 1991.

[15] D. Kostas and F. Rudzicz. Thinker invariance: enabling deep neural networks for bci across more people. *Journal of Neural Engineering*, 17(5):056008, 2020.

[16] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.

[17] F. Lotte and C. Guan. Learning from other subjects helps reducing brain-computer interface calibration time. In *2010 IEEE International conference on acoustics, speech and signal processing*, pages 614–617. IEEE, 2010.

[18] R. Mane, E. Chew, K. S. Phua, K. K. Ang, N. Robinson, A. Vinod, and C. Guan. Prognostic and monitory eeg-biomarkers for bci upper-limb stroke rehabilitation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(8):1654–1664, 2019.

[19] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, 38(11):5391–5420, 2017.

[20] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. Mueller-Putz, et al. Review of the bci competition iv. *Frontiers in neuroscience*, 6:55, 2012.

[21] L. Van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[22] V. Vapnik. *The nature of statistical learning theory*. Springer science & business media, 1999.

[23] Y. Wang, G. Huang, S. Song, X. Pan, Y. Xia, and C. Wu. Regularizing deep networks with semantic data augmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[24] L. Xu, M. Xu, Y. Ke, X. An, S. Liu, and D. Ming. Cross-dataset variability problem in eeg decoding with deep learning. *Frontiers in human neuroscience*, 14:103, 2020.
[25] F. Yger, M. Berar, and F. Lotte. Riemannian approaches in brain-computer interfaces: a review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10):1753–1762, 2016.
[26] P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthoumieu. Transfer learning: A riemannian geometry framework with applications to brain–computer interfaces. *IEEE Transactions on Biomedical Engineering*, 65(5):1107–1116, 2017.
[27] H. Zhang, M. Cissé, Y. N. Dauphin, and D. Lopez-Paz. Mixup: Beyond empirical risk minimization. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, 2018.
[28] L.-M. Zhao, X. Yan, and B.-L. Lu. Plug-and-play domain adaptation for cross-subject eeg-based emotion recognition. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. sn, 2021.
[29] K. Zhou, Y. Yang, Y. Qiao, and T. Xiang. Domain generalization with mixstyle. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*, 2021.