9th International IEEE EMBS Conference on Neural Engineering
San Francisco, CA, USA, March 20 - 23 , 2019

Feature extraction from the Hermitian manifold for Brain-Computer Interfaces

Jiachen Xu[1,2,3], Vinay Jayaram[1], Bernhard Schölkopf[1], Moritz Grosse-Wentrup[1,3,4]
[1] Max Planck Institute for Intelligent Systems, Department of Empirical Inference, Tübingen, Germany
Email: {jxu, vjayaram, bs, moritzgw}@tue.mpg.de
[2] Technical University of Munich, Department of Electrical and Computer Engineering, Munich, Germany
[3] Ludwig Maximilian University of Munich, Department of Statistics, Munich, Germany
[4] University of Vienna, Faculty of computer science, Vienna, Austria

***Abstract*—Riemannian geometry-based methods have shown to be effective in many sorts of Brain-Computer Interface (BCI) applications, but are only capable of measuring the power of the measured signal. This paper proposes a set of novel features derived via the Hilbert transform and applies them to the generalized Riemannian manifold, the Hermitian manifold, to see whether the classification accuracy benefits from this treatment. To validate these features, we benchmark them with the Mother of All BCI Benchmarks framework, a recently introduced tool to make BCI methods research more reproducible. The results indicate that in some settings the analytic covariance matrix can improve BCI performance.**

I. INTRODUCTION

BCI research suffers from little and noisy data, making robust feature generation crucial for building good classifiers. However, most methods are either based on the band-power or covariance of the time domain signal, such as common spatial pattern (CSP) [1, 2], which may imply that only a small fraction of the information contained in the electroencephalography (EEG) signal has been efficiently utilized. Considering the intractable problem of data scarcity, the other feasible breakthrough is via extracting more information from the signal and constituting novel features.

To best deal with the non-stationarities of the EEG signal, neither spatial filtering nor a conventional power-based classification scheme is employed in our work. Instead, we adopt the Riemannian classification framework due to two reasons [3, 4]: First, Riemannian approaches intend to replace the conventional Euclidean metrics with Riemannian metrics, which are a more appropriate measure of the distance between the symmetric positive definite (SPD) covariance matrices computed from the filtered EEG signal. Next, this framework outperforms standard spatial-filtering based approaches significantly in a recent meta-analysis [5], which may imply that classifiers built on these metrics are more robust.

Since noise is always an impediment to power-based analyses, phase or frequency methods have been adopted in order to focus on non-power-related aspects of the EEG signal. In particular, phase locking value [6], phase synchrony [7], and the average shift in frequency during a segment of EEG data [8] have also shown to add relevant information to classification. Being able to take advantage of the power of the Riemannian approaches while also using the additional information present in phase-based features may further improve classification outcomes in BCI.

As the most direct method to decode the phase information from the time-domain signal, the Hilbert transform (HT) has been explored in BCI [8, 9]. The HT is chosen in our work for two additional reasons: First, the power information of the signal remains identical after the HT is applied, which may imply common power-based methods could still be valid for transformed signals. Second, through the HT it is possible to compute the analytic signal, which can extend the real analysis of EEG signal into the complex domain. Furthermore, the generalization of the Riemannian manifold is well-defined in the complex number domain, and its metrics make it possible to measure the distance between analytic signals, which carry both phase and amplitude information. We would anticipate that the additional information carried with the analytic signal and Hilbert transformed (HT-ed) features could be beneficial as indicated by the success of the phase locking value [6].

The main contribution of this paper is: we generalize the currently best-performing feature extraction framework for BCIs into the complex domain and generate a set of novel features from it. Afterward, these features are validated using a new meta-analysis tool that allows us to compare effectiveness across over 250 subjects from various open-access datasets.

The remainder of this paper is structured as below: In Section II the Riemannian framework and the HT are briefly introduced. Afterward, we proceed with the Hermitian metrics and analytic features. Subsequently, the Mother of All BCIs Benchmarks (MOABB) [5] and its settings are given. Some important results are presented in Section III and illustrated in Section IV. Ultimately, we discuss the prospect of the proposed features in Section V.

II. METHODS

A. Riemannian classification framework

1) Riemannian metric: To use Riemannian manifold-based techniques, we first compute the covariance matrices for the signal from each trial. Standard Euclidean distances between covariance matrices are not robust, but recent work has shown that various other metrics can more reliably capture

978-1-5386-7921-0/19/$31.00 ©2019 IEEE

the structure of a set of covariance matrices. In this paper, the adopted metric is the affine-invariant Riemannian metric (AIRM) [10] where the "affine-invariance" means that the distance between two SPD matrices remains invariant after any affine transformation [11]. While a robust measure of distance, geometric properties such as geodesics and intersection angles on the manifold are complicated to compute. Therefore, by employing the logarithm mapping of matrices [3], we project the covariances on the manifold to the vectors lying on the tangent space (TS) centered around a point in order to use standard linear classification methods. The centered point is usually chosen as the mean covariance matrix to minimize the mapping error. One major benefit of classifying the vectors on the TS is the implicit filtering on the data which could get rid of the influence of noise [3].

B. Hilbert transform and analytic signal

The HT is routinely employed to estimate phase [12] or frequency [8, 9] features via the analytic signal. Because the instantaneous phase information can be extracted by comparing the information encoded within the real portion with the imaginary portion of analytic signal, i.e., $x(t)$ and $\mathcal{H}\left[(x(t))\right]$ respectively. In our paper, besides the phase-related features, we are also interested in the information contained by the imaginary portion of the analytic signal, i.e., the HT-ed signal $\mathcal{H}\left[(x(t))\right]$. In the time domain, it can be easily seen from its definition that the HT produces a weighted sum of the neighborhood of a point, where the weighting is signed and inversely proportional to distance.

Except for computing phase-related features, another advantage we take from the HT and the analytic signal is the ability to generalize real analysis into the complex analysis, which is seldom considered in the BCI field but rather common in the signal processing fields.

C. Classification on the Hermitian manifold

1) Hermitian metric: A Riemannian manifold is defined as a real manifold, which means that a complex-valued point cannot be measured properly with Riemannian metrics. Hence, we choose the Hermitian metrics which are a complex analogue of the Riemannian metrics, which in the particular case of the AIRM means we utilize the Hermitian inner product instead of the fully real version.

2) Analytic features: In the manifold approach, the features to be classified are the tangent vectors (TVs) mapped from the covariance matrices on the manifold, which are computed from three kinds of signal matrices: the time domain signal $\mathbf{X}(t)$, the HT-ed signal $\mathcal{H}\left[\mathbf{X}(t)\right]$, and the analytic signal $\tilde{\mathbf{X}}(t) = \mathbf{X}(t) + i\mathcal{H}\left[\mathbf{X}(t)\right]$.

After computing the corresponding covariance matrices, they are mapped onto the TS and then vectorized into the TVs. However, the TVs from the analytic signal are complex, which presents a problem to standard classification techniques. Therefore, the classification of analytic TVs in the complex domain will not be inspected in our work. Instead, the real part and imaginary part of the complex TVs are examined independently. Moreover, in order to check the effectiveness of additional information encoded with the analytic signal, we concatenate the real and imaginary part of the complex TVs as a new type of analytic features. Therefore, five types of features are tested in our work as summarized below:

- **RE**: TVs from the covariance matrices of time domain signal - $\mathbb{E}\left\{\mathbf{X}(t)\mathbf{X}^T(t)\right\}$
- **IM**: TVs from the covariance matrices of HT-ed signal - $\mathbb{E}\left\{\mathcal{H}\left[\mathbf{X}(t)\right]\mathcal{H}^T\left[\mathbf{X}(t)\right]\right\}$
- **Cov. Mat. real**: TVs from the real part of analytic covariance matrices - $\mathbb{E}\left\{\mathbf{X}(t)\mathbf{X}^T(t) + \mathcal{H}\left[\mathbf{X}(t)\right]\mathcal{H}^T\left[\mathbf{X}(t)\right]\right\}$
- **Cov. Mat. imag**: TVs from the imaginary part of analytic covariance matrices-$\mathbb{E}\left\{\mathcal{H}\left[\mathbf{X}(t)\right]\mathbf{X}^T(t) - \mathbf{X}(t)\mathcal{H}^T\left[\mathbf{X}(t)\right]\right\}$
- **Cov. Mat. re+im**: Concatenation of Cov. Mat. real and Cov. Mat. imag

D. MOABB Framework

We choose the MOABB framework for our analysis for two main reasons: First, by applying our proposed method on multiple datasets, we could acquire more trustworthy results. Second, our work will be more reproducible thanks to the open datasets and the framework provided by MOABB.

1) Datasets: In the following analysis, three typical paradigms of motor imagery (MI) signal are chosen: left hand, right hand, and feet MI. For each pair of comparisons, eight datasets are chosen to keep balance. For more details on the selected MOABB datasets, e.g., the number of channels and subjects, please refer to Table I.

TABLE I: Brief Summary of MOABB Datasets

Dataset Name	Motor Imagery		#Channels	#Subjects	#Sessions
	Left vs Right Hand	Feet vs Right Hand			
Alexandre Motor Imagery		✓	16	8	1
BNCI 2014-001	✓	✓	22	9	2
BNCI 2014-002		✓	15	14	1
BNCI 2014-004	✓		3	9	5
BNCI 2015-001		✓	13	12	2/3
BNCI 2015-004		✓	30	9	2
Cho et al. 2017	✓		64	49	1
Munich Motor Imagery	✓		128	10	1
Physionet Motor Imagery	✓	✓	64	109	1
Shin et al. 2017	✓		25	29	3
Weibo et al. 2014	✓	✓	60	10	1
Zhou et al. 2016	✓	✓	14	4	3

2) Setting of framework: We first limit our analysis to MI. Next, the classifications are all done within-session such that the non-stationarity is reduced to a minimum. As we are working with MI, our analysis is restricted to α and β band (8Hz $\sim$ 32Hz) which is less contaminated by noise and more neurophysiologically interpretable. In all datasets, all channels are selected except for the electrooculography (EOG) channel.

E. Pipeline

1) Features: The HT is first performed on the filtered signal along the time axis to compute the analytic signal. Next, in order to ensure all covariance matrices are full-rank, the estimation of the covariance matrices is regularized via the oracle approximating shrinkage estimator. Subsequently, the covariance matrices are projected to the TVs which comprises two steps: First, the geometric mean is computed for all covariance matrices based on AIRM. Afterward, the covariance matrices on the manifold are mapped to the TS based on the mean point and then converted into TVs.

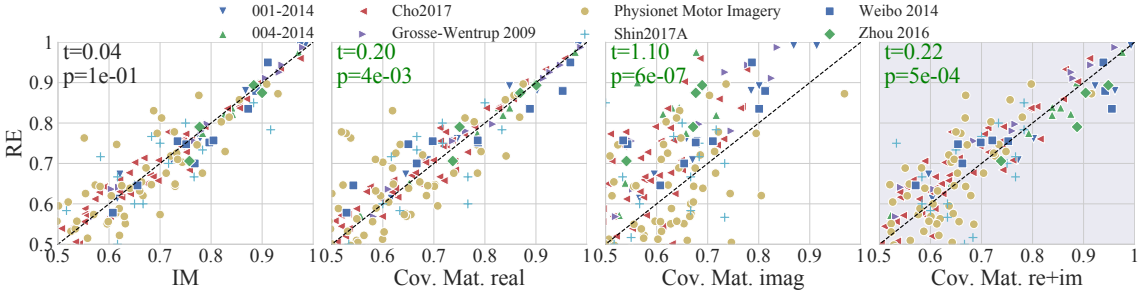
2) Classification: Fisher's Linear Discriminant Analysis (LDA) is chosen as our classification algorithm without regularization, and the accuracies are calculated with 5-fold cross-validation within each session.

3) Statistical analysis: Two statistical parameters are utilized: the p-values and the effect size (standard mean difference, SMD). We compute the p-value for the one-sided test within each dataset, whose null hypothesis is that the median accuracy of RE is not larger than that of another feature. As for more details, such as which model is employed to calculate p-values and which algorithm is used to combine p-values from multiple datasets, please refer to [5]. The effect size is taken as the SMD in the direction of the accuracies of RE minus the accuracies of another feature.

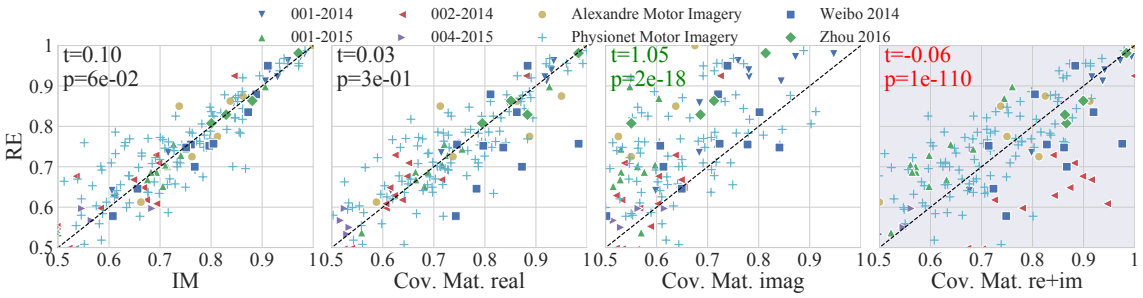
III. Results

Fig. 1 shows the comparison of accuracies using various features for different tasks. When comparing between RE and IM for both tasks, their performance tied with each other. However, from the results in the second column, we could note that RE is significantly better than the real part of complex TVs (Cov. Mat. real) when comparing between left and right hand MI signal, but not in the case of right hand versus feet. Interestingly, the subplots in the third column show that RE is significantly better than Cov. Mat. imag in both comparisons. Lastly, after using the concatenated features, there is no clear winner again. However, for some datasets, the novel features even improve the performance.

To figure out what happened to each dataset in the last column subplots of Fig. 1, the relevant meta-analysis has been conducted, and its results are as shown in Fig. 2. Subplot (a) shows that RE generally outperforms Cov. Mat. re+im for left versus right hand MI. However, when distinguishing between



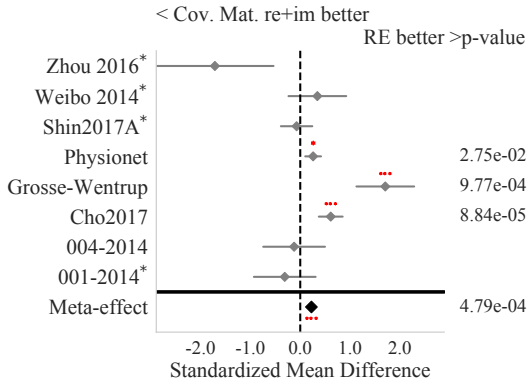
(a) Task: Left hand vs. Right hand



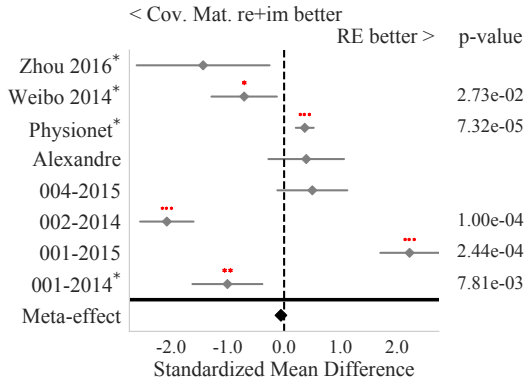
(b) Task: Feet vs. Right hand

Fig. 1: Accuracies comparison between using different features (Section II-C2). In each subplot, every point represents the classification results from one subject and the results from different datasets are distinguished from the shape and color of markers. Statistical parameters: overall p-value of one-tailed test **p** and overall effect size (SMD) **t** are computed based on the parameters from each dataset (Section II-E3). Text color: green - significant ($p < 0.05$ and $t > 0$) , black - not significant ($p >= 0.05$) , red - contradictory ($p < 0.05$ but $t < 0$)

feet and right hand MI, although there is no overall significant result, we could still notice that the concatenated features outperform the RE in three datasets with sizable effect and p-values lower than 0.03, which is rather surprising.



(a) Task: Left hand vs. Right hand



(b) Task: Feet vs. Right hand

Fig. 2: Meta analysis: p-value p and SMD are computed within each dataset. Black * means this dataset appears in both tasks. Red *, ** and *** represent $p < 0.05, 0.01, 0.001$ respectively

IV. DISCUSSION

Our goal was to verify whether those novel features derived via the Hermitian framework would be beneficial for classification accuracy. Based on the results in Section III, our novel features show increased performance in some situations, suggesting that complex covariance matrices may allow for a larger portion of the information in the EEG signal to be used.

In Fig. 1 the tied performance between RE and IM is regardless of the number of channels. These phenomena indicate that the information encoded in the IM is almost as effective for classification as that in RE. In other words, there is no information loss when using these new features, which is consistent with the fact that the HT does not alter the power spectrum of a signal. In addition, the subplots in the second column of Fig. 1 imply that RE is less effective when the location of stimulation sources are highly overlapped, such as when comparing feet and hand MI. Besides, although Cov. Mat. imag is significantly worse than RE with SMD > 0.8, it is still a predictive feature. The imaginary portion of the covariance corresponds to the covariance of the signal with its Hilbert transform, which has no obvious relationship to phase or amplitude.

Most surprising are the results in the last column of Fig. 1 and in Fig. 2. Although on average the Hermitian framework is not very effective, for some datasets the Hermitian-based signals (Cov. Mat. re+im) significantly outperform the Riemannian ones (RE) with very small p-values when classifying feet versus right hand imagery. In particular, for the case of *Weibo 2014*, this increase in performance is despite the addition of over 1800 additional features without any extra labels, which suggests that there is a rather prominent predictive component in the imaginary portion of the covariance.

V. CONCLUSION

In this paper, we propose a novel way to generate features which combine the Riemannian classification framework and the Hilbert transform together. Although the explanations of Cov. Mat. imag and concatenated features are still inadequate, their good performances convey the information that it would be worthy to explore them more which could also inspire researchers to find more relevant features in the future.

REFERENCES

[1] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers in Neuroscience*, vol. 6, p. 39, 2012.

[2] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms,," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2011.

[3] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Riemannian geometry applied to BCI classification," in *International Conference on Latent Variable Analysis and Signal Separation*. Springer, 2010, pp. 629–636.

[4] A. Barachant and S. Bonnet, "Channel selection procedure using Riemannian distance for BCI applications," in *Neural Engineering (NER), 2011 5th International IEEE/EMBS Conference on*. IEEE, 2011, pp. 348–351.

[5] V. Jayaram and A. Barachant, "MOABB: Trustworthy algorithm benchmarking for BCIs," *Journal of Neural Engineering*, vol. 15, no. 6, p. 066011, 2018.

[6] J. Onton, A. Delorme, and S. Makeig, "Frontal midline EEG dynamics during working memory," *Neuroimage*, vol. 27, no. 2, pp. 341–356, 2005.

[7] C. Brunner, R. Scherer, B. Graimann, G. Supp, and G. Pfurtscheller, "Online control of a brain-computer interface using phase synchronization," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, pp. 2501–2506, 2006.

[8] V. Jayaram, M. Hohmann, J. Just, B. Schölkopf, and M. Grosse-Wentrup, "Task-induced frequency modulation features for brain-computer interfacing," *Journal of Neural Engineering*, vol. 14, no. 5, p. 056015, 2017.

[9] A. Médl, D. Flotzinger, and G. Pfurtscheller, "Hilbert-transform based predictions of hand movements from EEG measurements," in *Engineering in Medicine and Biology Society, 1992 14th Annual International Conference of the IEEE*, vol. 6. IEEE, 1992, pp. 2539–2540.

[10] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *International Journal of Computer Vision*, vol. 66, no. 1, pp. 41–66, 2006.

[11] X. Pennec and N. Ayache, "Uniform distribution, distance and expectation problems for geometric features processing," *Journal of Mathematical Imaging and Vision*, vol. 9, no. 1, pp. 49–67, 1998.

[12] P. Y. Ktonas and N. Papp, "Instantaneous envelope and phase extraction from real signals: theory, implementation, and an application to EEG analysis," *Signal Processing*, vol. 2, no. 4, pp. 373–385, 1980.