Detecting EEG outliers for BCI on the Riemannian manifold using spectral clustering

Maria Sayu Yamamoto, Khadijeh Sadatnejad, Toshihisa Tanaka, Md. Rabiul Islam, Yuichi Tanaka, Fabien Lotte

Detecting EEG outliers for BCI on the Riemannian manifold using spectral clustering

Maria Sayu Yamamoto[1,2], Khadijeh Sadatnejad[1], Toshihisa Tanaka[2], Md. Rabiul Islam[2], Yuichi Tanaka[2] and Fabien Lotte[1,2]

***Abstract*— Automatically detecting and removing Electroencephalogram (EEG) outliers is essential to design robust brain-computer interfaces (BCIs). In this paper, we propose a novel outlier detection method that works on the Riemannian manifold of sample covariance matrices (SCMs). Existing outlier detection methods run the risk of erroneously rejecting some samples as outliers, even if there is no outlier, due to the detection being based on a reference matrix and a threshold. To address this limitation, our method, *Riemannian Spectral Clustering* (RiSC), detects outliers by clustering SCMs into non-outliers and outliers, based on a proposed similarity measure. This considers the Riemannian geometry of the space and magnifies the similarity within the non-outlier cluster and weakens it between non-outlier and outlier clusters, instead of setting a threshold. To assess RiSC performance, we generated artificial EEG datasets contaminated by different outlier strengths and numbers. Comparing Hit-False (HF) difference between RiSC and existing outlier detection methods confirmed that RiSC could detect outliers significantly better ($p < 0.001$). In particular, RiSC improved HF difference the most for datasets with the most severe outlier contamination.**

I. INTRODUCTION

Brain-Computer Interfaces (BCI) can identify users' intent from their brain activity only, mostly measured by Electroencephalography (EEG) [1]. BCI is notably promising as a communication and control tool for motor impaired users but also for Human-Computer Interaction for healthy users. However, several limitations have prevented practical BCI use outside laboratories [2], including their sensitivity to outliers. Outliers may be generated by ocular artifacts such as eye blinks or muscle artifacts such as jaw clenching [3]. This can lead to recognizing erroneous user's BCI commands.

In recent years, describing EEG signals by Sample Covariance Matrices (SCMs) and analyzing them in their native geometry, the so-called Riemannian geometry (RG), has contributed to improve BCI reliability [4], [5]. In particular, using RG led to state-of-the-art classification performance in many BCI studies, and even won multiple brain signal classification competitions [5]. Even though removing outliers is crucial to improve BCI classification performance, only two outlier detection methods were proposed based on RG for BCI application: the *Riemannian Potato* and *Geometric Trimmed Averages*, (cf. Section II) [6], [7]. However, these methods suffer from one main limitation: they need a threshold to detect outliers and a reference matrix to characterize the structure of the SCMs dispersion on the Riemannian manifold. Thus, they may detect different outliers depending on those parameters. In other words, there is a risk that some true outliers may not be detected as outliers or that some non-outlier samples may be erroneously rejected as outliers.

In this paper, we thus propose a novel approach to detect outliers without using any threshold nor reference matrix. With our method, the dispersion of the SCMs on the Riemannian manifold is represented with a graph structure, and SCMs are clustered according to their similarity. Ultimately, clusters with a smaller number of elements are regarded as outliers. We compare the proposed method and the existing methods with intentionally contaminated EEG datasets we generated. This enables us to have a ground truth to assess objectively how well each method can detect outliers.

This paper is organized as follows: Section II describes the principles of RG, the existing methods and our new approach. The data used for evaluation is described in section III. Then, Section IV describes the results while Sections V and VI discuss and conclude the paper, respectively.

II. METHODOLOGY

Let $X \in \mathbb{R}^{M \times N}$ be an EEG signal, with M channels and N time points. The SCM of X, P_X, is defined as:

$$P_X = \frac{1}{N-1}(X - \mu_X)(X - \mu_X)^T \tag{1}$$

where $\mu_X = \frac{1}{N}\sum_{n=1}^{N} X_n^{(:)}$ and $X_n^{(:)}$ is the n^{th} column of X. The space of SCMs is a subspace of a Riemannian manifold, i.e., a curved space equipped with an inner product on the tangent space at each point. Thus, when measuring the distance between two SCMs P_1 and P_2 on a Riemannian manifold, we should use the Riemannian distance R_d:

$$R_d(P_1, P_2) = \|\log({P_1}^{-\frac{1}{2}} P_2 {P_1}^{-\frac{1}{2}})\|_F = \left(\sum_{i=1}^{n} \log^2 \lambda_i\right)^{1/2} \tag{2}$$

where λ_i are positive eigenvalues of ${P_1}^{-\frac{1}{2}} P_2 {P_1}^{-\frac{1}{2}}$ and $\|.\|_F$ is the Frobenius norm of the matrix. The reader can refer to [8] for an in-depth review of Riemannian geometry.

This work was supported by the European Research Council with project BrainConquest (grant ERC-2016-STG-714567) and JSPS KAKENHI Grant Number 17H01760.

[1]Inria, 200 avenue de la vieille tour, 33405, Talence Cedex, France sayu.yamamoto, seyedehkhadijeh.sadatnejad, fabien.lotte@inria.fr

[2]Tokyo University of Agriculture and Technology, Koganei–shi, Tokyo 184–8588, Japan tanakat, ytnk@cc.tuat.ac.jp yamamoto18, rabiul@sip.tuat.ac.jp

A. Existing SCM outlier detection methods

Riemannian Potato is the first outlier detection method proposed for RG, for online EEG signal quality monitoring [6]. The detection algorithm requires two parameters: a reference matrix and a threshold. Samples whose distances from a reference matrix are larger than the threshold are rejected as outliers. The reference matrix is the Riemannian mean of all samples. The threshold th to reject samples is estimated as $th = \mu + 2.5\sigma$, where μ and σ are the mean and the standard deviation respectively of the Riemannian distances between each SCM and the reference matrix.

Geometric Trimmed Averages was proposed to improve the reliability of Tangent Space Mapping (TSM) based classifiers for BCI [7]. As an estimator for average, the plain geometric average derived from all samples is conventionally used but this estimation is affected strongly by outliers. In this method, the $d\%$ (a user-specified threshold) of samples that exhibit the largest Riemannian distance to the geometric mean (or geometric median) of all samples are eliminated as outliers. When evaluating a TSM-based BCI, all types of geometric trimmed averages including *mean-based trimmed mean* and *median-based trimmed median* exhibited higher classification accuracies than the plain geometric average.

B. New method: Riemannian Spectral Clustering (RiSC)

RiSC firstly computes the graph of SCMs, whose similarities is based on their pairwise Riemannian distance. Then, the nodes, i.e., SCMs, are clustered using spectral clustering. All clusters ,except the one with the largest number of elements, are rejected as outliers.

First, from a dataset containing s SCMs, the relations between them can be described by a fully connected similarity graph $G = \{V, E\}$, where $V = \{P_1, ..., P_s\}$ is a set of nodes, here SCMs, and E is a set of edges connecting V. The edges are weighted by the Gaussian similarity w_{ij} based on the Riemannian pairwise distance between P_i and P_j:

$$w_{ij} = \begin{cases} \exp\left(-\frac{R_d^2(P_i,P_j)}{2q^2}\right) & i \neq j, \\ 0 & i = j, \end{cases} \tag{3}$$

where $q \in [0, 1]$ controls the similarity between SCMs. In general, spectral clustering can be quite sensitive to changes in q values [9]. In the literature, the q parameter is often selected using rules of thumb [9]. However, we use the median of the length of the edges of the minimum spanning tree (MST) of a new graph G_1 that has the same structure as G, for automatic selection. Given a fully connected graph G_1 with s nodes and the Riemannian pairwise distance $R_d(P_i, P_j)$ as edge weight, its MST is a subgraph. This MST has the same s nodes as G_1 and all nodes are connected by a route that minimizes the total edge weights without any cycle. We assume the median of their length to be a reasonable radius of adjacency among non-outliers. Thus, using q defined in this way should magnify the similarity within the non-outlier cluster and weaken it between the non-outlier and the outlier clusters.

Another important point for spectral clustering is the choice of the number of clusters k. In this work, the optimal k is selected according to the eigengap heuristic [9]. The eigengap heuristic is based on the spectral decomposition of the Laplacian of graph G. The graph Laplacian L is a variation operator of G, obtained as $L = D - W$, where W is a symmetric matrix called weighted adjacency matrix, such that $(W)_{ij} = w_{ij}$, and D is a diagonal matrix whose diagonal elements are given by $(D)_{ii} = \sum_{i=1}^{s}(W)_{ij}$. Note that the graph Laplacian is always positive-semidefinite. The eigenvalues of L, $0 \leq \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_s$, are obtained by eigenvalue decomposition. The index of the i^{th} eigenvalue which indicates the maximum gap is set as k, i.e., $k = \arg\max_i(\lambda_{i+1} - \lambda_i)$.

After determining k using the eigengap heuristic, we conduct spectral clustering on L. First, we create a matrix $U \in \mathbb{R}^{s\times k}$ containing the first k eigenvectors of L, $u_1, ..., u_k$ as columns, and normalize the rows. Let $y_i \in \mathbb{R}^k$ $(i = 1, ..., s)$ be the vector corresponding to the i^{th} row of U, which represents the i^{th} SCM by k-dimensions. Then, all y_i vectors are clustered on $\mathbb{R}^k$ with the k-means algorithm into clusters $C_1, ..., C_k$. After clustering with k-means, all clusters except the one with the largest number of elements are regarded as outlier clusters. Thus, if there is no outlier, a single cluster should be detected, and no data will be rejected as outlier. This is in contrast to other existing methods, which always reject data, even when there is no outlier.

III. EXPERIMENT

For evaluation, we conducted two experiments. Experiment 1 investigated whether RiSC can correctly detect outliers for various datasets. Thus, we generated multiple intentionally contaminated datasets including artificial outliers. This provided us with a ground truth to objectively quantify the outlier detection performance of each method. Experiment 2 investigated RiSC performance on a real and clean EEG dataset, without any artificial outlier.

A. Real EEG Dataset Description

As real EEG dataset, we used dataset IIa from BCI competition IV, provided by TU Graz, Austria [10]. This set comprises EEG signals from nine subjects who performed left hand, right hand, both feet, and tongue Motor Imagery (MI). EEG signals were recorded using 22 EEG channels. In this dataset, the presence of eye movement artifacts was marked. However, here, EEG signals were band-pass filtered in the 7-30 Hz frequency band, using a 2nd order Butterworth filter. Thus, since eye movements artifacts usually affect frequency bands lower than 7 Hz [3], they should not affect the resulting SCMs. Training and testing sets are available for each subject. Both sets contain 72 trials for each class, each trial lasting 7 sec. Subjects performed MI within the time interval of $t = 3$ to 6 sec of each trial. In this work, we only used training EEG signals from left hand MI from subjects *A01*, *A02*, *A03*, and *A07* because these datasets contain the fewest ocular artifacts. For computing SCMs, we used EEG signals from the whole MI interval.

B. Contamination scenario

The contaminated datasets were generated by adding artificial artifacts to reference EEG trials by following the reference paper about probability models for time series additive outliers [12]. The reference EEG trials were randomly selected from the real dataset. In actual EEGs, outliers are often due to artifacts affecting specific channels and time periods. For instance, a common type of such artifacts is facial muscle artifacts (e.g., from frowning), which affect EEG recordings from frontal channels [3]. Thus, we generated contaminated EEG, i.e., outliers, by adding artifacts to frontal channels (Fz, FC3, FC1, FCz, FC2, and FC4), for a fixed time interval, to the reference EEG trials. Based on these trials, we produce c contaminated trials $Y \in R^{M \times N}$, such that the contaminated EEG signal $Y_t^{(m)}$ for channel m and time instance t is expressed by $Y_t^{(m)} = x_t^{(m)} + \gamma_t v_t$, where $x_t^{(m)}$ is a reference trial signal and $\gamma_t \in \{0, 1\}$ controls the proportion of time points to which we add artifacts. In this experiment, the simulated artifact v_t is added (i.e., $\gamma_t = 1$) with probabilities $\varepsilon = 0.10$, 0.30 or 0.50. We call ε the outlier strength. v_t is drown from a multivariate normal distribution $v_t \sim N(\mu, 2\sigma^2 I)$ in which $N(\cdot)$ is the normal distribution with mean μ and variance σ^2 of a reference channel selected randomly from each reference EEG trial. We used three different outlier numbers $c = 5, 10, 25$. For each generating condition, 30 datasets were randomly generated, for a total of 270 datasets per subject.

IV. Evaluation

The performance of RiSC is compared with both *Riemannian Potato* and *Geometric Trimmed Averages*. *Geometric Trimmed Averages* can use many types of trimming, as mentioned in Section II. Here, we used the *median-based trimmed average* since the median is more robust to outliers, and thus called this approach *Median-Based Trimming* (MBT). We used 95% trimming with MBT, a common threshold for statistical outlier detection.

In Experiment 1, the performance of each method was evaluated by Hit-False (HF) difference. The HF difference is calculated by subtracting the False Positive Rate (FPR) from the True Positive Rate (TPR) [11]. Here, the FPR is the percentage of non-outliers that were detected as outliers whereas the TPR is the percentage of actual outliers that were detected as outliers. To investigate the outlier detection performance of each method according to the outlier strengths and numbers, we performed a three-way ANOVA for repeated measures with factors Method (RiSC: *Riemannian Spectral Clustering*, RP: *Riemannian Potato*, MBT (95%): *Median based Trimming* (95%)), Outlier Strength, (10%, 30% or 50% strength), and Outlier Number, (5, 10 or 25 outliers).

In Experiment 2, each outlier detection method was evaluated by the number of trials detected as outlier.

A. Results: Experiment 1

The average HF difference for each intentionally contaminated dataset is summarized in Table I. The average HF difference with RiSC was globally higher than RP and MBT (95%) except for 10% strength, with either 5 or 10 outliers.

Three-way ANOVA for repeated measure revealed main effects of "*Method*" [$F(2, 238) = 1773; p < 0.001$], "*Outlier Strength*" [$F(2, 238) = 1297; p < 0.001$], and "*Outlier Number*" [$F(2, 238) = 628; p < 0.001$]. It also revealed interactions for "*Method X Outlier Strength*" [$F(4, 476) = 828; p < 0.001$], and "*Method X Outlier Number*" [$F(4, 476) = 1761; p < 0.001$] and "*Method X Outlier Strength X Outlier Number*" [$F(8, 952) = 170; p < 0.001$]. Post-hoc analyses of "*Method*", with Tukey's honestly significant difference test, showed that RiSC is significantly better than both RP [$MD = 27.80; p < 0.001$] and MBT (95%) [$MD = 24.08; p < 0.001$].

Fig. 1 shows the distribution of HF difference for each method and outlier strength. The HF difference with RiSC increased as the outlier strength increased. The HF difference with RP and MBT (95%) did not show any substantial change with the outlier strength.

Fig. 2 compares the HF difference for each outlier detection method and each outlier number. The HF difference of RiSC was constantly higher, regardless of the amount of outliers. In contrast, the performance of RP and MBT (95%) decreased with increasing the outlier number.

TABLE I
THE AVERAGE HF DIFFERENCE FOR EACH INTENTIONALLY CONTAMINATED DATASET [%]

Method	Outlier Number	Outlier Strength		
		10%	30%	50%
RiSC	5 outliers	2.67 ± 13.5	74.8±42.5	96.7±18.0
	10 outliers	0.67 ± 5.14	86.8±23.0	93.4±6.92
	25 outliers	84.4 ± 18.7	91.5±2.39	92.9±2.18
RP	5 outliers	54.2 ± 16.2	64.5±11.1	67.5±10.4
	10 outliers	38.8 ± 9.00	47.2±5.53	49.4±6.39
	25 outliers	17.3 ± 3.48	17.2±3.10	17.6±2.51
MBT(95%)	5 outliers	68.9 ± 14.7	78.6±5.36	80.0±0.00
	10 outliers	39.7 ± 1.79	40.0±0.00	40.0±0.00
	25 outliers	20.0 ± 0.00	20.0±0.00	20.0±0.00

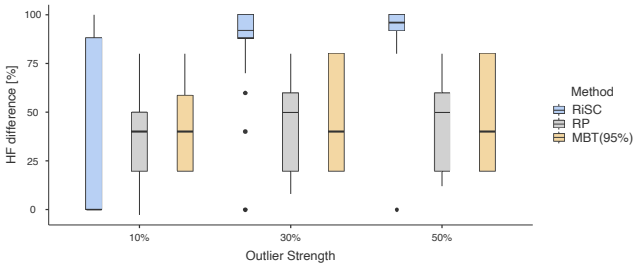

Fig. 1. Distribution of HF difference for each outlier strength, averaged across all outlier numbers.

B. Results: Experiment 2

In this experiment with clean, non-contaminated data, RP and MBT (95%) erroneously rejected some samples, while our proposed method did not detect any outlier (see table II).

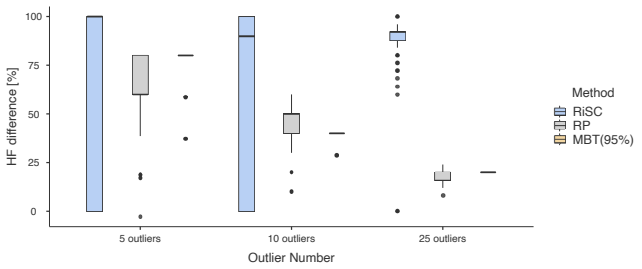

Fig. 2. Distribution of HF difference for each outlier number, averaged across all outlier strengths.

TABLE II
THE NUMBER OF SAMPLES DETECTED AS OUTLIERS IN THE NON-CONTAMINATED DATASET

Method	Subject			
	A01T	A02T	A03T	A07T
RiSC	0	0	0	0
RP	1	0	2	2
MBT (95%)	4	4	4	4

V. DISCUSSION

The purpose of Experiment 1 was to compare the performance among outlier detection methods on datasets with different types of contamination. Our method RiSC globally showed significantly higher HF difference than RP and MBT (95%), especially for the most contaminated dataset, e.g., with 50% strength and 25 outliers. However, the performance for more weakly contaminated datasets, notably with 10% strength and 5 outliers, was significantly lower than existing methods. This may be due to SCMs being based on EEG signal variance computation. Indeed, when estimating the signal variance over a long time window of EEG, short artifacts may have a small impact on the resulting overall variance. In other words, if the artifact affects a small proportion of EEG signals, the resulting SCM may be rather similar to non-outlier SCMs. Thus, its location on the manifold may become near the boundary of the non-outlier cluster. Thus, RiSC might incorrectly treat such weakly contaminated SCMs as part of the non-outlier cluster. On the other hand, RP and MBT (95%) both estimate the location of each SCM using their Riemannian distance from a reference matrix and then detect outliers with a threshold. Therefore, they can handle "weak" outliers interspersed near non-outliers. However, this comes at the risk of rejecting non-outliers as outliers, as shown by Experiment 2.

In contrast to RiSC, the HF difference with RP decreased with increasing the outlier number. This may be because the more outliers, the larger the Standard Deviation (SD). Since SD is one parameter defining the rejection threshold, the higher the SD, the higher the threshold, which reduces the amount of rejected outliers. Since MBT (95%) determines outliers as a fixed percentage of data, its performance did not change when changing the outlier strength.

The result of Experiment 2 revealed RiSC is a robust outlier detection method that does not detect normal samples as outlier incorrectly. On the other hand, RP and MBT (95%) detected some supposedly clean SCMs as outliers.

Both experiments suggested that rejecting SCMs that are only slightly different from normal SCMs risks to remove some normal samples as outliers. To determine whether we should take such a risk, we need further investigations on how such SCMs affect classification accuracy.

VI. CONCLUSIONS

In this paper, we proposed RiSC, an outlier detection method on the Riemannian manifold based on spectral clustering. Our proposed method addressed one limitation of existing methods, i.e., the necessity of a threshold and a reference matrix, by clustering SCMs into a non-outlier cluster and outlier clusters according to their similarity.

Results of comparisons with existing methods showed the superiority of RiSC. RiSC indeed performed overall significantly better with various contamination conditions and did not detect any SCM as outlier on the dataset without outlier. In particular, the performance increased on highly contaminated datasets. This suggests that RiSC could be useful in the future for BCI applications outside laboratories which are expected to suffer from various artifact sources.

This paper evaluated our method on MI-BCI data, however, RiSC is generic and should work on any type of EEG signals, e.g., Steady-State Visual Evoked Potentials, Event Related Potentials, or Sleep EEG. Thus, as future works, we will investigate the proposed method with other EEG signals. We will also explore whether RiSC can be used for improving BCI classifier training on noisy EEG data.

REFERENCES

[1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan, "Brain-computer interfaces for communication and control," Clinical Neurophysiology, 113(6), pp. 767–791, 2002.

[2] D.J. Krusienski, M.G. Wentrup, F. Galán, D. Coyle, K.J. Miller, E. Forney, and C.W. Anderson, "Critical issues in state-of-the-art brain–computer interface signal processing," J Neur Eng, 8(2):025002, 2011.

[3] M. Fatourechi, A. Bashashati, R.K. Ward, and G.E. Birch, "EMG and EOG artifacts in brain computer interface systems: A survey," Clinical Neurophysiology, 118(3), pp. 480–494, 2007.

[4] M. Congedo, A. Barachant, and R. Bhatia, "Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review," Brain-Computer Interfaces, 4(3), pp. 155–174, 2017.

[5] 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), pp. 1753–1762, 2016.

[6] A. Barachant, A. Andreev, and M. Congedo, "The riemannian potato: an automatic and adaptive artifact detection method for online experiments using riemannian geometry," in TOBI Workshop, 2013.

[7] T. Uehara, S. Matteo, T. Tanaka, and S. Fiori, "Robust averaging of covariances for EEG recordings classification in motor imagery brain-computer interfaces," Neural computation, 29(6), pp. 1631-1666, 2017.

[8] S. Lang, Differential and Riemannian Manifolds, ser. Graduate Texts in Mathematics. Springer, 1995.

[9] U. V. Luxburg. "A Tutorial on spectral clustering," Statistics and computing, 17(4), pp. 395-416, 2007.

[10] M. Naeem, C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller, "Seperability of four-class motor imagery data using independent components analysis," J. Neur. Eng., 3(3), pp. 208-216, 2006.

[11] S. Mason, J. Kronegg, J. Huggins, M. Fatourechi, and A. Schlögl, "Evaluating the performance of self-paced BCI technology," Technical report, Neil Squire Society, 2006.

[12] R.A. Maronna, R.D. Martin, V.J. Yohai, and M. Salibián-Barrera, "Robust statistics: theory and methods (with R)", Wiley, 2019.