



HAL
open science

Subspace oddity - optimization on product of Stiefel manifolds for EEG data

Maria Sayu Yamamoto, Florian Yger, Sylvain Chevallier

► **To cite this version:**

Maria Sayu Yamamoto, Florian Yger, Sylvain Chevallier. Subspace oddity - optimization on product of Stiefel manifolds for EEG data. ICASSP, Jun 2021, Toronto, Canada. hal-03202357v2

HAL Id: hal-03202357

<https://hal.science/hal-03202357v2>

Submitted on 1 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SUBSPACE ODDITY - OPTIMIZATION ON PRODUCT OF STIEFEL MANIFOLDS FOR EEG DATA

Maria Sayu Yamamoto^{†§} Florian Yger^{†*} Sylvain Chevallier[§]

[†] Tokyo University of Agriculture and Technology, Koganei-shi, Tokyo, Japan

[§]LISV, Univ. Paris-Saclay, Versailles, France

[‡] LAMSADE, CNRS, Univ. Paris-Dauphine, PSL Research University, , Paris, France

Email: yamamoto18@sip.tuat.ac.jp, florian.yger@dauphine.fr, sylvain.chevallier@uvsq.fr

ABSTRACT

Dimensionality reduction of high-dimensional electroencephalography (EEG) covariance matrices is crucial for effective utilization of Riemannian geometry in Brain-Computer Interfaces (BCI). In this paper, we propose a novel similarity-based classification method that relies on dimensionality reduction of EEG covariance matrices. Conventionally, the dimension of the original high-dimensional space is reduced by projecting into one low-dimensional space, and the similarity is learned only based on the single space. In contrast, our method, *MULTiple SUBspace Mdm Estimation* (MUSUME), obtains multiple low-dimensional spaces that enhance class separability by solving the proposed optimization problem, then the similarity is learned in each low-dimensional space. This multiple projection approach encourages finding the space that is more useful for similarity learning. Experimental evaluation with high-dimensionality EEG datasets (128 channels) confirmed that MUSUME proved significant improvement for classification ($p < 0.001$) and also it showed the potential to beat the existing method relying on only one subspace representation.

Index Terms— High-dimensional EEG, Riemannian manifold, optimization, dimensionality reduction, brain-computer interfaces

1. INTRODUCTION

Brain-Computer Interfaces (BCI) is a technology that can identify users' intent by extracting their specific electrophysiological activity and enable users to interact with the external world [1]. Therefore, it is a promising tool to enable motor-impaired users to communicate without their muscle movements. For non-invasive interfaces, this activity is commonly measured with electroencephalography (EEG).

In recent years, the use of Riemannian geometry for analyzing EEG described by covariance matrices has risen as

a new generation of BCI system [2, 3, 4]. The EEG covariance matrices belong to the manifold of Symmetric Positive-Definite (SPD) and call for methods grounded in non-flat geometry to be handled rigorously. This approach shows outstanding performance in the BCI classification compared to the Euclidean approaches relying on spatial filters estimation or feature selection.

However, Riemannian-based BCI classifier has also been reported that it tends to be less effective as the number of electrodes increases, whereas it is effective with a limited number of electrodes [5]. This problem is due to as the dimension grew, the empirical covariance matrices would suffer from rank deficiency and may no longer be SPD. Moreover, classifiers trained on a high-dimensional datasets are prone to overfitting. From those facts, it is desirable to reduce the dimensionality of the covariance matrices to ensure that Riemannian algorithms are used efficiently, even with many channels.

To address this limitation, several methods have been developed mainly outside of the BCI community and provide us with interesting ideas. In [6, 7], the authors developed a dimensionality reduction technique tailored for SPD matrices. This approach finds one subspace that is able to project the matrices such that the pairwise Riemannian distances in the original space would be similar in the space of the low-dimensional matrices. The problem is formulated as a supervised optimization problem on the *Grassman manifold*. It has been extended to an unsupervised setup to be a Riemannian PCA [8]. Those supervised and unsupervised approach are applied in a BCI setting in [9]. However, the experimental results do not show significant improvement for classification accuracy. This may be due to the fact that optimizing on a single subspace is eventually too restrictive. To overcome this shortcoming, Gao *et al.* [10] proposed an extensional work of [6] for multiple subspaces. The authors developed a similarity measure between matrices using notions point-to-set and set-to-set distance measure (PSSSD). PSSSD first projects each original SPD matrix into multiple low-dimensional SPD spaces (point-to-set). Then, low-dimensional representations are aggregated as set and a sim-

*FY acknowledges the support of the ANR as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute).

ilarity is learned by distance between each set (set-to-set). Since the model is formed with many parameter optimizations, PSSSD is more flexible than existing methods that apply a single Riemannian distance measure for metric learning. However, for BCI applications, datasets will be hardly large enough for learning so many parameters.

In this paper, we propose a novel similarity-based BCI classification inspired by PSSSD. All covariance matrices belonging to the high-dimensional manifold are projected into multiple low-dimensional subspaces, and the subspace which provide the highest classification performance is selected as the final classifier. This allows to find the most discriminant space for similarity learning from multiple low-dimensional representations. The proposed method was evaluated with two high-dimensional EEG datasets (128 channels) and the result showed superiority of the proposed method. To our knowledge, this is the first work that embeds EEG covariance matrix to multiple low-dimensional manifolds for classification.

This paper is organized as follows: Section 2 describes the principles of Riemannian geometry and our proposed method MUSUME. The experimental evaluation and the obtained results are described in Section 3. Then, Section 4 proposes a discussion while Section 5 concludes the paper.

2. METHODOLOGY

2.1. Riemannian geometry of EEG covariance matrix

Let $X \in \mathbb{R}^{M \times T}$ be an EEG signal which is bandpass filtered, with M channels and T time samples. The covariance matrix of X is defined as:

$$C_X = \frac{1}{T-1} X X^\top \quad (1)$$

Note that the estimated covariance matrix is empirically SPD, *i.e.* it can always be diagonalized with strictly positive eigenvalues. Using the right metric, the manifold of SPD matrix is smooth curved space that is locally differentiable. To manipulate SPD matrices while respecting their intrinsic geometry, we can rely on Riemannian geometry. There are several possible metrics [11] but the Affine Invariant Riemannian Metric (AIRM) is the most natural metric. AIRM distance δ_r between two SPD matrices C_1 and C_2 on manifold is defined as:

$$\delta_r(C_1, C_2) = \|\log(C_1^{-\frac{1}{2}} C_2 C_1^{-\frac{1}{2}})\|_F = \left(\sum_{i=1}^n \log^2 \lambda_i \right)^{1/2} \quad (2)$$

where λ_i are positive eigenvalues of $C_1^{-\frac{1}{2}} C_2 C_1^{-\frac{1}{2}}$ and $\|\cdot\|_F$ is the Frobenius norm of the matrix [12, Chap.6].

In our proposed similarity-based classification, we chose AIRM distance as a similarity measure.

2.2. MUSUME: Multiple Subspace Mdm Estimation

Our proposed similarity-based classification, MUSUME, is divided in two parts: dimensionality reduction and supervised similarity-based classification. In the dimensionality reduction part, high-dimensional SPD matrices are projected on multiple low-dimensional manifolds that can enhance the class separability. In the classification part, first we apply a classifier on each low-dimensional manifold. Then, the subspace with the highest classification performance for validation set is selected as the final classifier.

2.2.1. Dimensionality reduction

Let $\mathcal{D} = \{C_1, C_2, \dots, C_n\}$ be a set of n SPD matrices of size $M \times M$ and $y = [y_1, y_2, \dots, y_n]$ is their respective label (either +1 or -1). The form to project the original SPD matrices $C_i \in \mathbb{R}^{M \times M}$ into l low-dimensional manifolds are defined as:

$$\begin{cases} f_{U_1}(C_i) = U_1^\top C_i U_1 \\ \vdots \\ f_{U_l}(C_i) = U_l^\top C_i U_l \end{cases} \quad (3)$$

where $f_{U_s}(\cdot)$ is the s^{th} low-dimensional manifold projection and $U_s \in \mathbb{R}^{M \times p}$ is the s^{th} low-dimensional manifold projection matrix. Note that to ensure that the resulting low-dimensional spaces form valid SPD manifolds, we impose orthonormality constraints on U_s , *i.e.* $U_s^\top U_s = I_M$.

The optimal projection matrices $U = \{U_1, U_2, \dots, U_l\}$ are derived by solving the following optimization problem:

$$\begin{aligned} & \text{minimize} && \sum_{ij} y_i y_j N_{ij} \sum_{s=1}^l \delta_r^2(f_{U_s}(C_i), f_{U_s}(C_j)) \\ & \text{subject to} && U_1, U_2, \dots, U_l \in St(M, p) \end{aligned} \quad (4)$$

where $St(M, p) = \{U \in \mathbb{R}^{M \times p} : U^\top U = I_M\}$ is a manifold endowed with a Riemannian metric, so-called *Stiefel manifold* [13] and N_{ij} is the set of k -nearest neighbors of C_i with respect to C_j . The entries are as follows:

$$N_{ij} = \begin{cases} 1, & \text{if } C_i \in \mathcal{N}_k(C_j) \text{ or } C_j \in \mathcal{N}_k(C_i) \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

In order to keep the computation tractable, only the samples close to each other in the sense of δ_r within the original space are considered in the cost function. In practice, the cost function was optimized using a conjugate gradient method with an adaptive line search. The gradient of $\delta_r^2(\cdot)$ with respect to U_s can be computed as follows:

$$\begin{aligned} & \nabla_{U_s} (\delta_r^2(U_s^\top C_i U_s, U_s^\top C_j U_s)) = \\ & 4 \left(C_i U_s (U_s^\top C_i U_s)^{-1} - C_j U_s (U_s^\top C_j U_s)^{-1} \right) \\ & \log \left(U_s^\top C_i U_s (U_s^\top C_j U_s)^{-1} \right) \end{aligned} \quad (6)$$

This gradient is obtained by utilizing the fact that $\text{Tr}(\log(V)) = \log \det(V)$. The reader can refer to the appendix of [8] for the

details of the derivation and to [13] for details on optimization on manifolds.

2.2.2. Supervised similarity-based classification

Each low-dimensional manifold spanned by $\{U_1, U_2, \dots, U_l\}$ provide different representations of original SPD matrices. To make the most out of this abundance of representations, we apply the minimum distance to mean (MDM) classifier [14] on each low-dimensional manifold, then the subspace yielding the maximum classification accuracy for validation set is selected as the final classifier for test set.

3. EXPERIMENTAL EVALUATION

3.1. Data description

For evaluation, we used two different publicly available high-dimensionality Motor Imagery (MI) datasets. The first dataset is Munich MI dataset provided by Grosse-Wentrup *et al.* [15]. This set comprises EEG signals from 10 healthy subjects who performed left- and right-hand MI. EEG signals were recorded using 128 EEG channels placed according to the extended 10 – 20 system. A total of 150 trials per condition were carried out by each subject, each MI interval lasting 7 sec. In our experiment, we used 96 trials per condition as training set, 24 trials per condition as validation set, 30 trials per condition as testing set.

The second dataset is the high-gamma dataset described in [16], hereinafter called Schirrneister MI dataset. The EEG signals were recorded with 128 EEG channels from 14 healthy subjects while they performed 4 classes MI, the left hand, right hand, both feet, and rest. In this work, we only used the left and right hand MI. In training sets, different number of trials are available from 80 to 224 per condition depending of the subject and the classes are balanced, thus we used 80% of training data as training set and remaining 20% is for validation. For the testing set, a total of 40 trials per condition is available. Each MI interval lasted 4 sec. In our experiment, we used 13 subjects except S02 since its training data was not balanced in two classes.

EEG signals in both datasets were bandpass filtered in the 8 – 32 Hz frequency band, using a 4th order Butterworth filter. To estimate covariance matrices, we used the filtered EEG signals from the whole MI interval. All pre-processing was conducted via the Mother of all BCI Benchmark (MOABB) [17] to ensure the replicability of this study.

3.2. Evaluation

To evaluate MUSUME, we compare its classification performance with two baselines:

- **MDM-noDR:** MDM without dimensionality reduction

Table 1. Classification accuracy (in %) for Munich MI dataset

	MDM-noDR	MDM-1DR	MUSUME
S01	96.67	100.00	100.00
S02	75.00	76.67	78.33
S03	48.33	53.33	56.67
S04	45.00	66.67	66.67
S05	51.67	51.67	51.67
S06	41.67	51.67	51.67
S07	66.67	68.33	65.00
S08	46.67	58.33	61.67
S09	36.67	46.67	45.00
S10	60.00	63.33	65.00
Ave.	56.83	63.67	64.17

- **MDM-1DR:** MDM with dimensionality reduction on one low-dimensional manifold ($l = 1$) with 20NN
- **MUSUME:** dimensionality reduction on four low-dimensional manifolds ($l = 4$) with 20NN

In order to compare how the difference in the number of spaces (single or multiple) affects the classification accuracy, we set up MDM-1DR by using our cost function with $l = 1$ instead of implementing the cost function proposed in [9].

We performed a one-way ANOVA for repeated measures to see statistical difference in three compared methods.

Applying MDM to multiple low-dimensional manifolds gives a wider range of possible classification accuracy than applying MDM to a single low-dimensional space. Therefore, we investigate the maximum accuracy for validation set among subspaces by changing the number of subspaces with $l = 1, 2, 4, 6, 8, 10$. We assumed this accuracy is the upper bound that MUSUME can achieve.

3.3. Results

The classification accuracy for Munich dataset is summarized in Table 1. The results in bold indicate the maximum score in the subject. MUSUME showed equal or increased accuracy for all subjects except one subject from MDM-noDR and especially it provided more than 10% improvement for 3 subjects (mean gain: $7.33 \pm 6.7\%$). In the comparison with MDM-1DR, MUSUME showed equal or increased accuracy for 8 subjects (mean gain: $0.50 \pm 2.0\%$).

Table 3 shows the result for Schirrneister MI dataset. The accuracy with MUSUME was equal or increased for 10 subject (mean gain: $2.98 \pm 4.3\%$) compared to MDM-noDR. In the comparison with MDM-1DR, MUSUME showed equal or increased accuracy for 10 subjects (mean gain: $0.77 \pm 2.3\%$).

To assess the statistical validity of those findings, we performed a one-way ANOVA for repeated measures of all 23 subjects. It revealed a significant effect of “Method” ($F(2, 44) = 13.7; p < 0.001$). Post-hoc analyses with Tukey’s honestly significant difference test showed that

Table 2. Upper-bound accuracy with respect to the number of low-dimensional manifolds.

Subspace number	1 subspace	2 subspaces	4 subspaces	6 subspaces	8 subspaces	10 subspaces
Upper bound accuracy[%]	69.04	70.76	71.30	71.18	71.82	71.79

Table 3. Classification accuracy (in %) for Schirrmester MI dataset

	MDM-noDR	MDM-1DR	MUSUME
S11	58.75	60.00	57.50
S12	97.50	98.75	98.75
S13	82.50	82.50	82.50
S14	52.50	61.25	67.50
S15	61.25	58.75	60.00
S16	52.50	52.50	51.25
S17	82.50	82.50	83.75
S18	60.00	67.50	65.00
S19	82.50	88.75	88.75
S20	53.75	57.50	58.75
S21	96.25	97.50	97.50
S22	71.25	72.50	76.25
S23	71.25	71.25	73.75
Ave.	70.96	73.17	73.94

our proposed method was significantly better than MDM-noDR ($MD = 4.87; p < 0.001$) but the statistical difference between MUSUME and MDM-1DR was not revealed ($MD = 0.65; p = 0.789$).

Table 2 shows the influence of the number of subspaces on the classification accuracy. The results in bold indicate the accuracy which beat the single subspace accuracy.

4. DISCUSSION

In the comparison with MDM-noDR, MUSUME showed a significant improvement globally for the classification accuracy with two datasets. It validates that our method can achieve both dimensionality reduction of covariance matrices and improves the classification performance at the same time. Furthermore, MUSUME showed higher classification accuracy than MDM-1DR for 10 out of 23 subjects, while MDM-1DR beat MUSUME only in 5 subjects. This may be explained because the MDM-1DR relies on only one representation of the covariance matrix, whereas MUSUME adopts the space showing the highest classification accuracy among multiple representations for the final decision, thus, the opportunity to meet higher classification accuracy increases. Therefore, we expected that if we increase the number of low-dimensional subspaces, it would offer more options to find useful subspaces for classification. As it is shown in Table 2, the upper-bound accuracy was higher than MDM-1DR. However, a significant increase was not revealed. This may be due to MUSUME overfitting for the training data since

the number of available training set is often limited in BCI datasets.

MDM-1DR is a non-convex (and even non-g-convex [18, chap.11]) problem which makes its solution very sensitive to the choice of the initialization. On the contrary, as MUSUME tries to optimize several subspaces in parallel in order to find the best of them, it is more robust to a poor initialization. Hence, MUSUME acts in the spirit of an ensemble method and this may explain why MUSUME performs in practice better than MDM-1DR.

5. CONCLUSION

In this paper, we proposed a novel similarity-based classification method for high-dimensional EEG covariance matrices. Our method, MUSUME, addresses the potential limitation of the current Riemannian approach, *i.e.* the inefficiency for high-dimensional EEG covariance matrix, with projection on multiple low-dimensional manifolds. The experimental result with high-dimensional EEG datasets confirmed its superiority than baseline methods. In particular, MUSUME showed significant improvement from MDM operated on the original high-dimensional covariance space. This suggests that MUSUME can reduce the dimension of EEG covariance matrices without unexpected loss of information, hence, it is useful for BCI applications recorded with many electrodes.

In this study, we started with the simple classification model to highlight the impact of multiple projections, but as a future task, we will aim to built classification models that enables us to integrate the covariance features extracted by multiple projection. It would be achieved by referring to other ensemble learning techniques or promoting diversity of the subspaces using a regularization [19].

The MDM is known as a classifier that is strongly impacted by outliers since it learns similarity based on distance from intra-class Riemannian mean. Therefore, to make MUSUME more robust, it would be interesting to use Riemannian geometric median [20] to estimate intra-class center on each subspace instead of using Riemannian mean or to incorporate an outlier detection method [21] in MUSUME.

6. ACKNOWLEDGMENTS

The authors would like to thank Dr. Fabien Lotte for his helpful comments and proofreading on this paper.

7. REFERENCES

- [1] Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan, “Brain-computer interfaces for communication and control,” *Clinical Neurophysiology*, vol. 113(6), pp. 767–791, 2002.
- [2] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger, “A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update,” *Journal of Neural Engineering*, vol. 15, no. 3, pp. 031005, 2018.
- [3] Florian Yger, Maxime Berar, and Fabien Lotte, “Riemannian approaches in brain-computer interfaces: a review,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25(10), pp. 1753–1762, 2016.
- [4] Marco Congedo, Alexandre Barachant, and Rajendra Bhatia, “Riemannian geometry for eeg-based brain-computer interfaces; a primer and a review,” *Brain-Computer Interfaces*, vol. 4, no. 3, 2017.
- [5] Florian Yger, Fabien Lotte, and Masashi Sugiyama, “Averaging covariance matrices for EEG signal classification based on the csp: an empirical study,” in *2015 23rd European Signal Processing Conference (EUSIPCO)*. IEEE, 2015, pp. 2721–2725.
- [6] Mehrtash T Harandi, Mathieu Salzmann, and Richard Hartley, “From manifold to manifold: Geometry-aware dimensionality reduction for SPD matrices,” in *European Conference on Computer Vision*. Springer, 2014, pp. 17–32.
- [7] Mehrtash Harandi, Mathieu Salzmann, and Richard Hartley, “Dimensionality reduction on SPD manifolds: The emergence of geometry-aware methods,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 1, pp. 48–62, 2017.
- [8] Inbal Horev, Florian Yger, and Masashi Sugiyama, “Geometry-aware principal component analysis for symmetric positive definite matrices,” in *Asian Conference on Machine Learning*, 2016, pp. 1–16.
- [9] Pedro Rodrigues, Florent Bouchard, Marco Congedo, and Christian Jutten, “Dimensionality reduction for BCI classification using riemannian geometry,” in *7th Graz Brain-Computer Interface Conference*, 2017.
- [10] Zhi Gao, Yuwei Wu, Mehrtash Harandi, and Yunde Jia, “A robust distance measure for similarity-based classification on the SPD manifold,” *IEEE Transactions on Neural Networks and Learning Systems*, 2019.
- [11] Sylvain Chevallier, Eemmanuel K Kalunga, Quentin Barthélemy, and Eric Monacelli, “Review of Riemannian distances and divergences, applied to SSVEP-based BCI,” *Neuroinformatics*, pp. 1–14, 2020.
- [12] Rajendra Bhatia, *Positive definite matrices*, vol. 24, Princeton university press, 2009.
- [13] P-A Absil, Robert Mahony, and Rodolphe Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, 2009.
- [14] Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten, “Multiclass brain–computer interface classification by riemannian geometry,” *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 920–928, 2011.
- [15] Moritz Grosse-Wentrup, Christian Liefhold, Klaus Gramann, and Martin Buss, “Beamforming in noninvasive brain–computer interfaces,” *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1209–1219, 2009.
- [16] Robin Tibor Schirrmester, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball, “Deep learning with convolutional neural networks for eeg decoding and visualization,” *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.
- [17] Vinay Jayaram and Alexandre Barachant, “MOABB: trustworthy algorithm benchmarking for BCIs,” *Journal of Neural Engineering*, vol. 15, no. 6, pp. 066011, 2018.
- [18] Nicolas Boumal, *An introduction to optimization on smooth manifolds*, 2020.
- [19] Ikko Yamane, Florian Yger, Maxime Berar, and Masashi Sugiyama, “Multitask principal component analysis,” in *Asian Conference on Machine Learning*, 2016, pp. 302–317.
- [20] Thomas Fletcher, Suresh Venkatasubramanian, and Sarang Joshi, “The geometric median on riemannian manifolds with application to robust atlas estimation,” *NeuroImage*, vol. 45, no. 1, pp. S143–S152, 2009.
- [21] Maria Sayu Yamamoto, Khadijeh Sadatnejad, Toshihisa Tanaka, Md Rabiul Islam, Yuichi Tanaka, and Fabien Lotte, “Detecting EEG outliers for BCI on the riemannian manifold using spectral clustering,” in *42nd Annual International Conferences of the IEEE Engineering in Medicine and Biology Society (EMBC’2020)*, 2020.