



HAL
open science

Transfer learning for the Riemannian tangent space: Applications to Brain-Computer Interfaces

Alexandre Bleuzé, Jérémie Mattout, Marco Congedo

► **To cite this version:**

Alexandre Bleuzé, Jérémie Mattout, Marco Congedo. Transfer learning for the Riemannian tangent space: Applications to Brain-Computer Interfaces. ICEET 2021 - 2021 International Conference on Engineering and Emerging Technologies, Oct 2021, Istanbul, Turkey. pp.1-6, 10.1109/ICEET53442.2021.9659607. hal-03420920

HAL Id: hal-03420920

<https://hal.science/hal-03420920>

Submitted on 9 Nov 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.

Transfer learning for the Riemannian tangent space: Applications to Brain-Computer Interfaces

Alexandre Bleuzé
GIPSA-lab
CNRS, Grenoble INP
University Grenoble Alpes
Grenoble, France
Email: alexandrebleuze@hotmail.fr

Jérémie Mattout
Lyon Neuroscience Research Center
INSERM, CNRS
University Claude Bernard Lyon 1
Lyon, France
Email: jeremie.mattout@inserm.fr

Marco Congedo
GIPSA-lab
CNRS, Grenoble INP
University Grenoble Alpes
Grenoble, France
Email: marco.congedo@gipsa-lab.fr

Abstract—Transfer learning for training brain-computer interface (BCI) decoding algorithms is useful to reduce the calibration time, increase the accuracy, reduce the risk of overfitting and allow the application of machine learning methods that require a large amount of data, such as deep neuronal networks. In this article we propose a transfer learning method inspired by recent advances in Riemannian geometry. The method aligns vectors in the tangent space of a *source* and a *target* data set by means of Procrustes Analysis. We apply the method on a publicly available P300-BCI database. We show that using our method it is possible to transfer information reusing data from other subjects. The classification accuracy we obtain, as compared to the state of art, shows a clear transmission of information using the transfer learning method.

I. INTRODUCTION

A Brain-Computer Interface (BCI) is a computerized system for on-line prediction or classification of cognitive states and intentions of the user. A BCI extracts meaningful features from an incoming stream of neurophysiological signals and translates them into discrete events, for instance, commands for a machine, by means of a machine learning classification algorithm (1; 2; 3).

In this article we focus on BCIs based electroencephalography (EEG), which is not only completely non-invasive, safe and silent, but also affordable to the large public due to recent advances in micro-technology (4). Because of these characteristics EEG suits both a clinical and large public use, embracing a large spectrum of applications such as rehabilitation/substitution of motor functions, spellers, videogaming, cognitive ability training and more (5).

A major difficulty with EEG-based BCIs is the high *inter-subject* and *inter-session* variability. Typical BCIs operate in two phases: in the *training phase* the classifier is calibrated in a *supervised* fashion, that is, with examples of EEG data corresponding to labeled classes; the actual use of a BCI is named the *test phase* and is *unsupervised*, that is, the BCI classifies EEG data to infer the classes they belong to. Inter-subject variability refers to the fact that the training is highly subject-specific, thus in general a calibration is necessary for each user. Inter-session variability refers to the fact that the training achieved in one session is sub-optimal in other sessions of the same user, thus in practice a traditional BCI

requires a calibration at each and every use. Calibration is time and energy consuming, representing a burden both for healthy users and the clinical population. Thus, the research in the BCI domain is currently focusing on methods to reduce the length of the calibration or to bypass it completely. A possible strategy in case of a new user is to train the BCI classifier with data previously collected on other subjects. Starting from the second session of the same user, data collected in previous sessions can be used as well for this purpose. Methods accomplishing these tasks are said to operate *transfer learning*, of the between-subject and between-session type, respectively. Another term used in the literature for those methods is *domain adaption*.

We refer to the (known) domain we ought to use for learning as the *source* and the domain we want to apply the learning to as the *target*. Our goal is to adapt the data of the *target* subject, i.e., the subject who will be using the BCI, to the data of some *source* subject, i.e., a subject who has already used the BCI. After that we can train a classifier with the *source* data and use the trained classifier on adapted *target* data kept beforehand for testing. *Source* and *target* data sets may refer to different subjects or to different sessions, as aforementioned.

Recently, there have been some advances in this kind of transfer learning using Riemannian geometry. The first proposition aimed at domain adaptation using the differential geometry concept of parallel transport on the manifold of positive-definite matrices; the data points of both the *source* and the *target* are recentered to the origin of the manifold, which is the identity matrix (6). It turns out that such recentering is equivalent to data whitening. Similarly, in (7) the authors recentered the data to another common point, namely the midpoint between the centers of mass of the *source* and *target* data. As one may expect, (8) showed that these two methods yield the same performance. Another method in the same spirit, although not in a Riemannian framework, is **CORrelation ALignment (CORAL)** (9). The authors proposed an unsupervised domain adaptation that can be interpreted in the framework of parallel transport under certain conditions. It consists in whitening the *source* domain and recoloring it with the *target* domain covariance. This means parallel transporting from the *source* base-point to the identity and then to the *target*

domain.

Recentering is a translation procedure. Inspired by Procrustes analysis, the authors in (8) proposed to add two more matching steps after recentering: a stretching operation to match the dispersion of the points of the *source* and *target* set and a rotation. The Riemannian version of Procrustes analysis has been named by the authors Riemannian Procrustes Analysis (RPA). They showed that RPA allows superior transfer learning performance as compared to recentering only. The methods described so far carry out the transfer learning on the manifold and the data are then projected onto the tangent space for classification purposes (10). In (11) a recentering on the manifold is followed by the projection onto the tangent space, where the tangent vectors are submitted to a principal component analysis (PCA), independently for the *source* and *target* data sets. This method has the advantage of being unsupervised. However, in contrast with the very aim of transfer learning, PCA does not result in the alignment of the *source* and *target* tangent vectors. Furthermore, as acknowledged by the authors, PCA suffers from a sign ambiguity, which requires a manual correction of the sign of the principal vectors, making the method less viable in practice. Note that such a PCA was already proposed in (12), but for variable selection only.

In this paper we propose to align the tangent vectors using procrustes analysis in the tangent space. Under certain conditions this is equivalent to aligning the *source* and *target* tangent vectors by Maximum Covariance Analysis (MCA) (13). The transfer learning method we propose encompasses those above reported; while it can be conceived as a MCA on the feature vectors, thus as the bivariate extension of the PCA applied in (11), it is cast as a Procrustes analysis. This results in a tangent space version of the RPA applied in (8). Since our method applies in the tangent space, in contrast to RPA, it can be used in all circumstances, not just in the framework of Riemannian machine learning. For the same reason, it can be used with any kind of data, it generalizes well with any number of classes and not just in BCI, thus our method is of the utmost generality. It also does not need an iterative algorithm as it yields an exact solution. In this way, the algorithm is faster than RPA and holds the potential to be more precise. This is achieved at the price of a tangent space approximation.

We test the proposed transfer learning approach, named Tangent Space Alignment (**TSA**), on BIEEG.2012-GIPSA (14), a publicly available database comprising 25 subjects using a P300-based BCI. We compare the single-trial classification accuracy obtained with transfer learning to the *golden standard* (GS), represented by the usual training-test estimation of accuracy without transfer learning and to the **CORAL** transfer learning method.

II. METHOD

A. Riemannian geometry

As machine learning we employ the framework offered by Riemannian geometry (12; 15). This framework has proven very effective in the BCI field, both in terms of accuracy

and robustness, as demonstrated by the winning score obtained in five international BCI data classification competitions (16). Riemannian classifiers encode EEG data segments into positive-definite matrices, in practice, some form of covariance matrices. Those positive-definite matrices are then treated as points in a Riemannian manifold of non-positive curvature adopting the affine-invariant metric (17; 18). In such manifold, the notion of shortest path and distance between two points, center of mass of several points and projection onto a tangent space are naturally defined as shown in Fig. 1. This figure shows on the left-hand side the geodesic relying points P_1 and P_2 passing through its mid-point (mean) G (green curve), the tangent space at point G with tangent vectors to geodesic from G to P_1 and from G to P_2 (blue arrowed lines) and the distance $\delta(G, P_2)$. It also shows on the right-hand side the center of mass (also named geometric mean) G of points P_1, \dots, P_4 defined as the point minimizing the sum of the four squared distances $\delta^2(G, P_i)$, for $i = 1, \dots, 4$.

Since the tangent space is an Euclidean space, by projecting observations onto it we can employ usual machine learning algorithms such as support vector machines or LASSO logistic regression, which have been shown to successfully capture the essence of BCI data (16). Deep neuronal networks can be used as well. In this work we employ a Linear Support Vector Classifier (SVC) (19). The general functioning of Riemannian classifiers is as it follows: an EEG epoch (a trial) is encoded in the form of a positive-definite matrix, typically, some form of covariance matrix. A representative point is taken on the manifold of positive-definite matrices, e.g., the center of mass of all observations, and all points are projected onto the tangent space defined at this point. Once vectorized, the points in the tangent space are treated as feature vectors, so as to be submitted to classical machine learning algorithm in order to determine a decision function (12; 15).

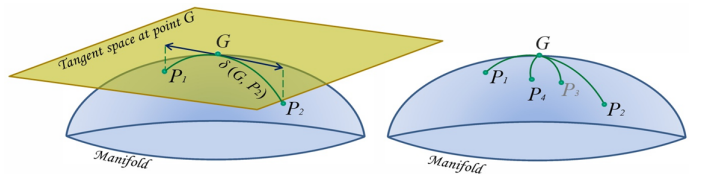


Fig. 1: Schematic illustration of the Riemannian manifold of positive-definite matrices.

B. EEG data encoding

In this work we focus on P300-based BCIs. In these BCIs there are two classes of event-related potentials (ERPs): + and 0 class. The + class corresponds to the response to stimulations of items the user wants to select. The 0 class corresponds to the response to all other stimulations. Following previous work on P300-based BCIs (20; 16), in order to encode efficiently the temporal information of ERP data we first define a "super-trial" as

$$X' = \begin{bmatrix} X \\ P_+ \end{bmatrix}, \quad (1)$$

where X is the single-trial $N \times T$ data matrix of N electrodes and T samples and P_+ is the $N \times T$ 'prototype' matrix for the $+$ class, usually taken as the mean of $+$ trials. The super trials are thus $2N \times T$ data matrices. Once obtained, they are encoded as positive-definite matrices by computing their sample covariance matrix

$$C = \frac{1}{T} X' X'^T, \quad (2)$$

where superscript T denotes matrix transposition. Large artefacts jeopardize the condition number of sample covariance matrix estimations, making them numerically close to singularity. In order to avoid numerical problems in ensuing computations, all covariance matrices are regularized using the Ledoit-Wolf shrinkage method (21).

C. Feature vectors

In Riemannian classifiers acting in the tangent space the feature vectors are the tangent vectors (12; 15). They are obtained by first estimating, for each data set separately, a center of mass of all available trials, also named Riemannian geometric mean (16). Unlike previous works, we compute the weighted geometric mean assigning to each trial a weight inversely proportional to the class numerosity. This way each class contributes equally to the determination of the center of mass, it does not matter if the classes are balanced or unbalanced (they are unbalanced in P300-based BCI). Then, all trials are recentered (parallel transported to the identity (6)), which, as discussed in the introduction, is a common step to all transfer learning procedures. Finally, the covariance matrices are projected onto the tangent space.

Let M be the geometric mean and C the covariance matrix of a trial, all the above operations amount to transformation

$$S = \text{Log}\left(M^{-1/2} C M^{-1/2}\right), \quad (3)$$

where $\text{Log}()$ is the matrix logarithm function. Finally, the feature vector is obtained as

$$z = \text{uvec}\left(S \circ Q\right), \quad (4)$$

where Q is a matrix holding 1 on the diagonal elements and $\sqrt{2}$ elsewhere, \circ is the Hadamard (element-wise) product and $\text{uvec}()$ vectorizes the upper triangular part of the argument. Weight $\sqrt{2}$ is given to the off-diagonal elements of S in order to preserve the Euclidean norm, i.e., to yield $\|S\|_F = \|z\|_2$.

D. Feature vector alignment

The method we propose is named Tangent Space Alignment (**TSA**). Let

$$Z_T = [z_{T,1} \cdots z_{T,P_T}] \text{ and } Z_S = [z_{S,1} \cdots z_{S,P_S}] \quad (5)$$

be the matrix holding in its columns a packet of P_T and P_S feature vectors taken from a *target* and *source* subject's run, respectively. Now let us consider the mean tangent vectors

$$\bar{Z}_T = [\bar{z}_{T,0}, \bar{z}_{T,+}] \text{ and } \bar{Z}_S = [\bar{z}_{S,0}, \bar{z}_{S,+}], \quad (6)$$

where $\bar{z}_{T,0}$ is the mean tangent vector for class 0 and $\bar{z}_{T,+}$ is the mean tangent vector for class $+$ for the *target* subject and similarly for the *source* subject. In order to re-align the tangent vectors of the *target* to those of the *source*, we solve the Procrustes problem

$$\arg \max_{P \in \mathcal{O}} \|P \bar{Z}_T - \bar{Z}_S\|_F, \quad (7)$$

where \mathcal{O} is the orthogonal group. For doing so, we consider the cross-product of the two packets, i.e.,

$$K = \bar{Z}_T \bar{Z}_S^T \quad (8)$$

and compute its singular value decomposition as

$$K = U^T \Delta V. \quad (9)$$

It is well-known that problem (7) has solution $P = VU^T$, where U and V are the orthogonal matrices in (9). Therefore the transformed feature vectors of the Procrustes problem for the *target* are obtained by

$$Z'_T = VU^T Z_T. \quad (10)$$

In practice, the number of columns in U and V is chosen so as to explain 99% of the cross-covariance.

E. Pipelines

For the assessment and comparison of our approach, we construct pairs of data sets Z_T and Z_S (5) following the processing method described in sections II-B, II-C and II-D. Using calibration data, we take each subject in a database as the *target* and all the others as the *source* in order to construct the cross-product K in Eq. (8) and solve the procrustes problem. Let $Z_{T_{est}}$ be a third dataset of feature vectors from the *target* subject using the test data. The golden standard pipeline is the classical calibration-test procedure. We use it as a comparison for all the following pipelines, where Z_T is considered as the training set and $Z_{T_{est}}$ as the testing set for classification:

1. **TSA**: Using the **TSA** method, we consider $Z'_T = VU^T Z_T$ as per (10). Finally, Z_S forms the training set and Z_T will not be used anymore. $Z'_{T_{est}} = VU^T Z_{T_{est}}$ is used as the testing set. A schematic description of the **TSA** transfer learning method is shown in Fig. 2.

2. **CORAL**: In order to allow a fair comparison with **TSA**, the *source* and *target* feature vectors used for **CORAL** are Z_S and Z_T as they are vectors lying in the Euclidean tangent space. In addition the test data $Z_{T_{est}}$ are also kept for testing.

F. Database

In this paper we employ a publicly available database for analysis. It has been recorded at GIPSA-lab, in Grenoble (France) and involved a BCI experiment using the P300-based Brain Invaders video game (22). The base BI.EEG.2012-GIPSA (14) contains EEG recordings from 25 healthy subjects. EEG signals were acquired by means of the NeXus-32 (MindMedia, Herten, Germany), a research-grade amplifier.

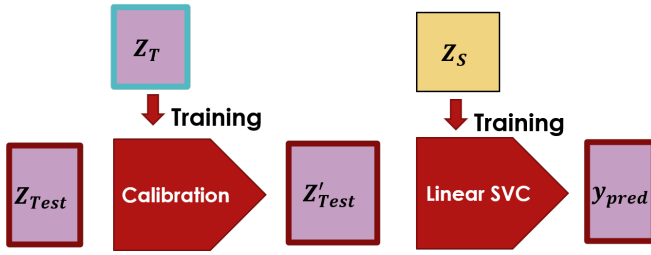


Fig. 2: **TSA** pipeline : A Linear SVC is trained on *source* data Z_S . Calibration *target* tangent vectors Z_T are used as alignment data. Finally, the test data Z_{Ttest} tangent vectors are aligned and then used as the testing set. See text for details.

The EEG cap was equipped with 16 Silver/Silver Chloride wet electrodes, placed according to the 10-20 international system. The electrode locations were F7, F3, F4, F8, T7, C3, CZ, C4, T8, P7, P3, PZ, P4, P8, O1 and O2. The ground was placed at the FZ location, The NeXus-32 machine does not use any reference electrode but a hardware common average reference. Data from three subjects were found to be corrupted and for another subject an insufficient number of trials were available. Therefore, 21 subjects were kept for subsequent analysis.

The P300 is an event-related potential (ERP). It peaks in-between 240-600 ms after stimulus onset. The classification task in this BCI paradigm is a 2-class problem, with the ratio of trials per class being 6:1 (non-target:target). Each subject performed both a calibration and a test run.

G. Computations

All analyses have been performed using in-house code written in the Julia programming language (23). The estimation of covariance matrices is done using the *CovarianceEstimation.jl* library¹. For the manipulation of data in the Riemannian manifold of positive-definite matrices we have used the *Diagonalizations.jl*, *PosDefManifold.jl* and *PosDefManifoldML.jl* packages (24) all publicly available². Finally the classification has been achieved with an SVC imported from the Python Scikit-Learn library³.

III. RESULTS

Since the classes are unbalanced, the balanced accuracy is used in all analysis. It is defined as

$$\text{Balanced accuracy} = \frac{\text{specificity} + \text{sensitivity}}{2}. \quad (11)$$

We show in Fig. 3 the balanced accuracies as a function of the number of trials used for alignment for **TSA** and **CORAL** compared with a simple train-test using the alignment data as training data for the golden standard algorithm. Golden standard has higher standard deviation since the curve is the average over the 21 subjects whereas for both alignment

methods the curve is the average of the $20 \times 21 = 420$ *source-target* pairs.

We observe that with a low number of trials used for alignment the **CORAL** method results in a better accuracy as compared to both the **TSA** method and the **GS** train-test evaluation. However, the higher the number of trials used for alignment, the better **TSA** becomes, whereas the **CORAL** method floors and is eventually surpassed. For the ensuing analysis, we focus on three distinct parts of these curves :

- The low alignment based on less than 120 trials
- The medium alignment based on in-between 120 and 240 trials.
- The high alignment based on more than 240 trials.

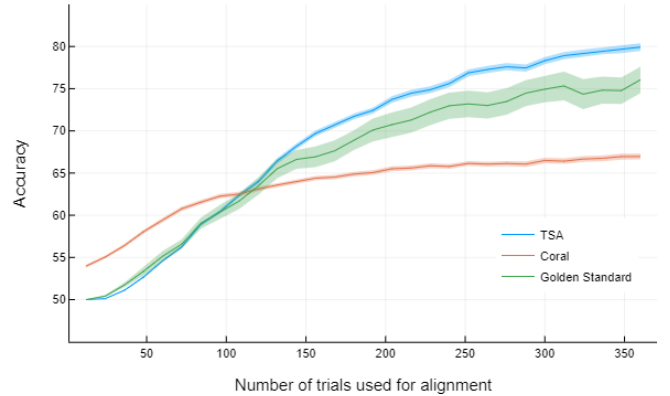


Fig. 3: Mean (line) and standard deviation (shaded area) balanced accuracy over the number of trials used for alignment for all methods presented in Section II.

Following Rodrigues et al. (8), we use seriation plots to display the balanced accuracy for all possible *source-target* pair for the **TSA** and **CORAL** methods. We evaluate the attained balanced accuracies obtained for the low, medium and high alignment, respectively. The corresponding seriation plots are shown in Fig. 4. For each sub-plot, each row corresponds to a *target* subject and each column corresponds to a *source* subject. The rows and columns are sorted by ascending accuracy for targets obtained on the **TSA** method. Please note that in each row there is a dummy value (black), which corresponds to the case where the *source* subject is also the *target* subject. Additionally, a golden standard seriation plot has also been computed even though the *source* has no impact on the accuracy since it is not used. The aim of these plots is only to facilitate a visual comparison with **TSA** and **CORAL**.

We observe that for the **CORAL** method the accuracy of a given subject highly depends on both the *target* and the *source*. For the **TSA** method we observe that it mainly depends on *target* subject only. This shows that when enough alignment trials are available it is possible to achieve little dependency on the choice of the source data. This means that we are able to align the data between subjects without caring much about how "compatible" they are. This is an interesting feature for a BCI transfer learning algorithm as it means that if subjects are

¹<https://github.com/mateuszbaran/CovarianceEstimation.jl>

²<https://github.com/Marco-Congedo>

³<https://github.com/scikit-learn/scikit-learn>

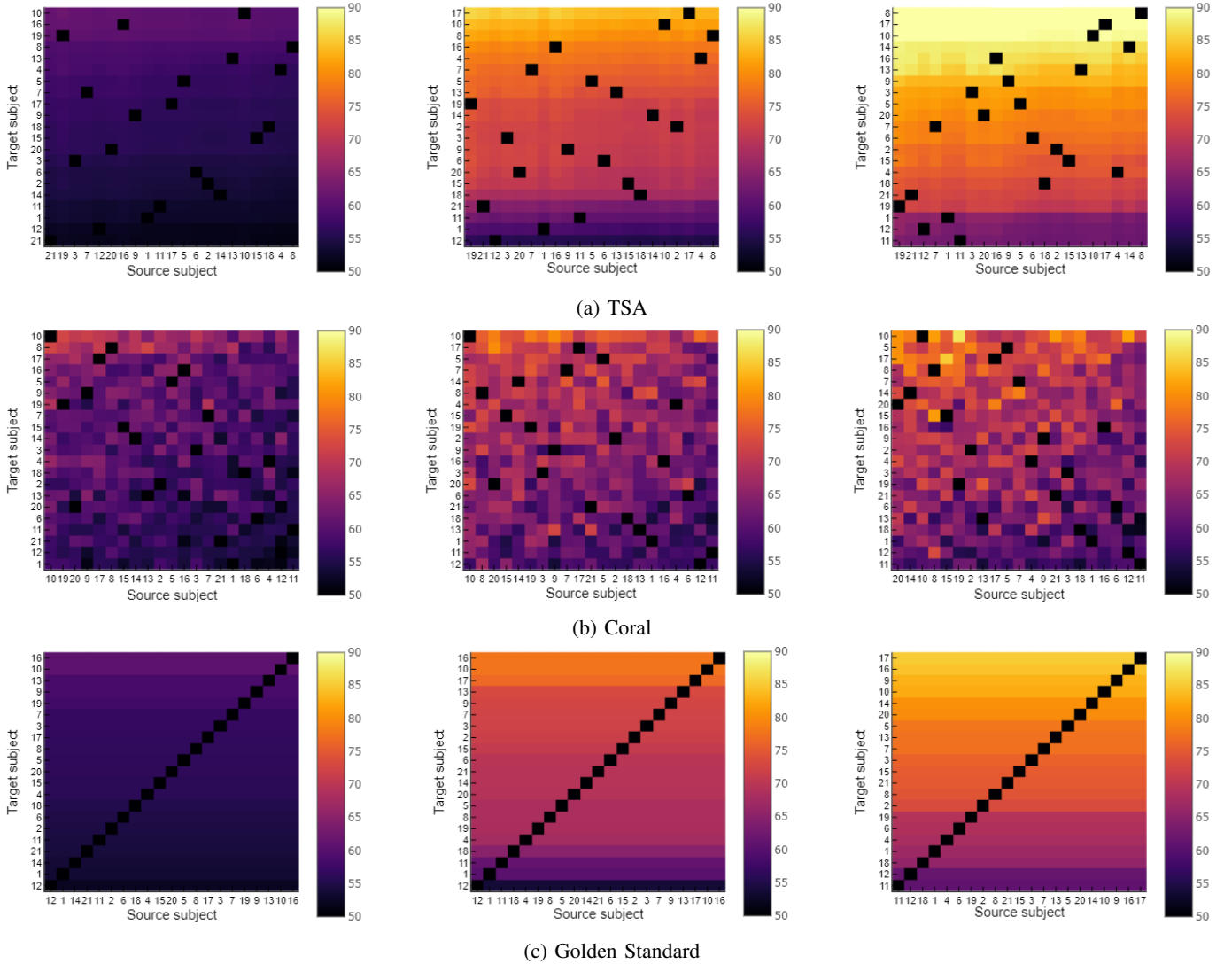


Fig. 4: Seriation plot obtained with TSA (first row), Coral (second row) and Golden Standard (third row) for, from left to right column, low, medium and high alignment situation, respectively.

aligned with enough data, their feature space will be aligned no matter how far they were originally. Possible uses of such a feature will be discussed in more details in the Discussion section.

Finally, we perform statistical analysis to compare the overall accuracy (averaging across number of trials used for alignment) obtained by each of the two pipelines with respect to the GS. For each *source* subject, we compare the accuracy between the three methods by applying paired Student's t-tests to the data shown in Fig. 3. For all *source* subjects **TSA** is superior to both **GS** and **CORAL**. The p-value is inferior to 0.001 for 17 out of 21 subjects comparing **TSA** to **CORAL** and for 7 out of 21 subjects comparing **TSA** to **GS**.

IV. CONCLUSIONS AND DISCUSSION

The contribution of this article is a novel transfer learning method for BCI purposes. It allows a very efficient transfer

of information from one subject to the other, with results surpassing a well-established method provided that a sufficient number of trials for alignment is used.

The **TSA** method here introduced demonstrates a good potential for transfer learning in the case of P300-based BCI data, where the mismatch between subjects' data distributions is pretty severe in general. In this study, the method is used to align vectorized tangent vectors of symmetric positive-definite matrices living in a Riemannian manifold, however it is not limited to this case and could be applied to any kind of feature vectors lying in an Euclidean space.

In this study **TSA** presents with systematic improvement of accuracy for all subjects, regardless their **GS** accuracy, as soon as about 120 trials are available for alignment. We observe that with the **TSA** method, the accuracy obtained for a given *target* is hardly dependent on *source* data. This kind of feature makes possible the creation of a database composed

of the alignment of multiple subjects. Such a database has the potential to help the field taking advantage of using deep neural networks which highly depend on data quantity in order to create a robust algorithm with reliable inter-session and inter-subject adaptation. It should be stressed that the method has been tested only on a single dataset. Therefore further testings have to follow in order to ensure the ability of **TSA** to execute accurate transfer learning in general.

In further studies, more databases will be used in order to ensure the effectiveness of the method in other BCI paradigms. Other studies may also investigate the accuracy of this method with more than two classes. Finally, research on cross-database transfer learning also seems promising since **TSA** method may align databases with different number of channels. This is due to the fact that it uses Singular Value Decomposition (SVD) in order to align data, and SVD is possible on rectangular cross-product matrices.

V. ACKNOWLEDGEMENTS

The authors are grateful to the members of the ViBS team (GIPSA-Lab) for their suggestions related to the analysis of the data.

REFERENCES

- [1] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, and T. Vaughan, "Brain-computer interfaces for communication and control," *Neurophysiol Clin.*, 2002.
- [2] R. Rao, "Brain-Computer Interfacing - An Introduction," *Cambridge University Press, Cambridge*, 2013.
- [3] F. Lotte, L. Bourgain, A. Cichoki, M. Clerc, M. Congedo, and A. R. et al., "A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update," *J Neural Eng*, vol. 15, 2018.
- [4] J. W. Y. Kam and S Griffin and A. Shen and S. Patel and H. Hinrichs and H. J. Heinze et al., "Systematic comparison between a wireless eeg system with dry electrodes and a wired eeg system with wet electrodes," *NeuroImage*, vol. 184, pp. 119–129, 2019.
- [5] J. Wolpaw and E. W. Wolpaw, "Brain-Computer Interfaces: Principles and Practice," *Oxford University Press, Oxford*, 2012.
- [6] P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthoumieu, "Transfer learning: a Riemannian geometry framework with applications to Brain-Computer Interfaces," *IEEE Trans Biomed Eng*, vol. 65, 2018.
- [7] O. Yair, M. Ben-Chen, and R. Talmon, "Parallel Transport on the Cone Manifold of SPD Matrices for Domain Adaptation," *IEEE Trans Signal Process*, vol. 67, pp. 1797–1811, 2019.
- [8] P. L. C. Rodrigues, C. Jutten, and M. Congedo, "Riemannian Procrustes Analysis : Transfer Learning for Brain-Computer Interfaces," *IEEE Trans Biomed Eng*, vol. 66, pp. 2390–2401, 2018.
- [9] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," 2015.
- [10] M. Congedo, A. Barachant, and A. Andreev, "A new generation of brain computer interface based on riemannian geometry," 2013.
- [11] G. Maman, O. Yair, D. Eytan, and R. Talmon, "Domain adaptation using Riemannian geometry of SPD matrices," *ICASSP*, 2019.
- [12] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass Brain-Computer Interface Classification by Riemannian Geometry," *IEEE Trans Biomed Eng*, vol. 59, pp. 920–928, 2012.
- [13] M. Congedo, "EEG Source Analysis," 2013, HDR presented at Doctoral School EDISCE, University of Grenoble.
- [14] G. V. Veen, A. Barachant, A. Andreev, G. Cattan, P. Rodriguez, and M. Congedo, "Building brain invaders: Eeg data of an experimental validation," 2019.
- [15] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications," *Neuro-Computing*, vol. 112, pp. 172–178, 2013.
- [16] M. Congedo, A. Barachant, and R. Bhatia, "Riemannian geometry for eeg-based brain-computer interfaces; a primer and a review," *Brain-Computer Interfaces*, vol. 4, pp. 155–174, 2017.
- [17] R. Bhatia, "Positive definite matrices," *Princeton University Press.*, 2007.
- [18] M. Moakher, "A differential geometric approach to the arithmetic and geometric means of operators in some symmetric spaces," *SIAM J Matrix Anal Appl.*, vol. 26, pp. 735–747, 2005.
- [19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, and O. e. a. Grisel, "Scikit-learn: Machine learning in Python," *J Mach Learn Res*, vol. 12, pp. 2825–2830, 2011.
- [20] A. Barachant and M. Congedo, "A plugplay p300 bci using information geometry," 2014.
- [21] O. Ledoit and M. Wolf, "Honey, i shrunk the sample covariance matrix," *The Journal of Portfolio Management*, vol. 30, pp. 110–119, 2004.
- [22] M. Congedo, M. Goyat, N. Tarrin, L. Varnet, B. Rivet, and G. I. et al., "Brain Invaders: a prototype of an open-source P300- based video game working with the OpenViBE platform," *5th International Brain-Computer Interface Conference 2011 (BCI 2011)*, pp. 280–283, Sep 2011.
- [23] J. Bezanson, A. Edelman, S. Karpinski, and V. Shah, "A fresh approach to numerical computing," *SIAM review*, pp. 59, 65–98, 2017.
- [24] M. Congedo and S. Jain, "A Julia Package for manipulating Brain-Computer Interface Data in the Manifold of Positive Definite Matrices," in *Conf Proc IEEE Int Conf Syst Man Cybern*, 2019.