

## THE RIEMANNIAN MINIMUM DISTANCE TO MEANS FIELD CLASSIFIER

M Congedo, P.L.C. Rodrigues, C. Jutten

GIPSA-lab, Université Grenoble Alpes, CNRS, Grenoble-INP

E-mail: Marco.Congedo@Gipsa-Lab.fr

**ABSTRACT:** A substantial amount of research has demonstrated the accuracy of the Riemannian minimum distance to mean (RMDM) classifier for brain-computer interface (BCI). This classifier is simple, fully deterministic, robust to noise, computationally efficient and prone to transfer learning. The use of the geometric mean in the Riemannian manifold of symmetric positive definite matrices has proved fundamental to obtain these characteristics. Recently the general family of power means living on this manifold, which includes the geometric mean, has been defined. In this article we extend the RMDM algorithm in an unsupervised and adaptive fashion using a sampling of power means, named *means field*. We show that the resulting Riemannian minimum distance to means field (RMDMF) classifier features superior performance. Our conclusion is supported by the analysis of 17 public databases covering two BCI paradigms, for a total of 335 individuals, using the open-source MOABB (Mother of all BCI Benchmark) framework. In order to promote reproducible research, our full code is released.

### INTRODUCTION

Riemannian geometry is a branch of differential geometry that studies smooth manifolds, curved spaces with peculiar geometries. In these spaces notions of angles, shortest path between two points, distances, center of mass of several points, etc., allow to study analytic properties of mathematical operators from a geometric perspectives, making them accessible to intuition [1]. In the field of brain-computer interface (BCI) the manifold of symmetric positive-definite (SPD) matrices [2] has proved very useful, since multivariate electroencephalography (EEG) data in finite time windows can effectively be mapped as points onto this manifold through the estimation of some form of their covariance matrix [3-6]. This approach has led to the introduction of classifiers with remarkable characteristics as compared to the state-of-the-art [6]; for a formal introduction to the SPD Riemannian manifold the reader is referred to [2], while for a primer and review of its use in the BCI context to [7,8].

Riemannian classifiers have proved accurate, general and robust to noise, largely superior to state-of-the-art competitors [9], winning five international BCI machine learning competitions in which they have competed [7].

In particular, the Riemannian minimum distance to mean (RMDM) classifier, while not the most accurate among Riemannian classifiers, stands out for its simplicity, computational efficiency and universality (it applies to all BCI paradigms). It is a fully deterministic and parameter-free classifier, thus no parameter needs to be tuned by cross-validation or other methods that may jeopardize its generalization. Further, it can be easily extended to the multi-user scenario [10], has proved accurate also in individuals affected by clinical conditions [11] and has proved apt to integrate transfer learning [12, 13] and adaptation strategies [14], as demonstrated by the calibration-less P300-based BCI video game *Brain Invaders* [14-16].

In its non-adaptive (test-training) form, the RMDM works as it follows: a training provides a set of SPD matrices encoding BCI trials for the available classes. For each class a center of mass of the available trials is estimated. Finally, in test mode, a BCI trial to be classified is encoded in the same way as an SPD matrix and is assigned to the class whose center of mass is the closest according to a suitable distance function acting on the manifold [7]. In adaptive mode instead, the centers of mass are initialized by a database of previous users for a naive user and/or a database of the same user for a non-naive user thanks to transfer learning strategies [12, 13], then the centers of mass are adapted to the user while the BCI is operated [14]. In any case, the good performance of the RMDM classifier derives from the adoption of an appropriate metric for the SPD manifold. The metric in turn determines both the distance function between two points and the definition of a center of mass for a cloud of points, which is also a function of the distance since it is defined as the point on the manifold minimizing the dispersion of the cloud around itself [7, 17]. So far, the hyperbolic (geometric) distance and the geometric mean as a center of mass, which arise adopting the Fisher-Rao (affine-invariant) metric, have been preferred, due to a number of desirable invariance properties they possess. As explained in [7], those are the extension to SPD matrices of the usual hyperbolic distance and geometric mean for scalars. Simply stated, their use instead of the much more common Euclidean distance and arithmetic mean has engendered the success of Riemannian classifiers dealing with covariance matrices [7, 8].

In [18] the authors have defined a one-parameter family of means generalizing to SPD matrices the power means

for scalars. Given a set of  $K$  scalars  $\{c_1, \dots, c_K\}$ , the power mean with real parameter  $h \neq 0$  is given by

$$g_{(h)} = \left( \frac{1}{K} \sum_k c_k^h \right)^{1/h}. \quad (1)$$

As for the scalar power means, the SPD matrix power means of [18] (see also [19-21]) interpolate between the harmonic mean ( $h = -1$ ) and the arithmetic mean ( $h = 1$ ), while the geometric mean that we have discussed corresponds to the limit of  $h$  evaluated at 0, from either side. This generality of power means is appealing in the BCI context; as suggested in [22], in EEG data the sensor measurement is affected by several noise components and varying the order  $h$  one can find an optimal mean depending on the signal-to-noise-ratio. In [22] we have tested the accuracy of the RMDM algorithm using 13 power means with  $h = \{\pm 1, \pm 0.8, \pm 0.6, \pm 0.4, \pm 0.2, \pm 0.1, 0\}$  (see Fig. 7 therein). The classical RMDM classifier corresponds to the power mean with  $h=0$  (geometric mean). We have found that the value of  $h$  offering the maximum accuracy gravitated around zero, but  $h=0$  was optimal only for three out of the 38 tested subjects. Instead, the optimal value of  $h$  was highly variable across individuals. Also, there was a significant positive correlation between the maximal accuracy and the value of  $h$  allowing such maximum. Thus, the higher the accuracy, which is an indirect measure of signal-to-noise ratio, the higher the optimal value of  $h$ .

Finding the optimal value of  $h$  for a given subject and session, as we have done in [22], is a supervised procedure. Therefore in seeking the optimal value we roll back to the problem of obtaining a classifier that is prone to overfitting, that lacks transfer learning and that is not capable of adaptation. Let us instead name a *means field* a sampling of power means in the interval  $h \in [-1, 1]$  such as the one used in [22]. Then, in this article we propose the Riemannian minimum distance to means field (RMDMF) classifier. It uses in an unsupervised and adaptive fashion all the means in the field for classifying. In particular, for a given unlabeled datum, the closest power mean, regardless of its class, is found, then the MDM is applied using the power means with the value of  $h$  that corresponds to the closest mean. Such method is unsupervised, in that it can be used blindly to any datum without any learning and is adaptive, in that the preferred value of  $h$  is allowed to change during the session.

We employ MOABB (Mother of All BCI Benchmark) [9] for testing the RMDMF against the RMDM classifier on 17 databases covering two BCI paradigms (motor imagery and P300), for a total of 335 individuals. MOABB is an open-source framework for objectively assessing the performance of BCI classifiers on large amount of data. The use of MOABB ensures that exactly the same processing pipeline is applied to all databases of the same type and that both the cross-validation procedure and the Riemannian classifiers operate exactly in the same way for all databases, regardless the BCI type.

## MATERIALS AND METHODS

Table 1 reports the main characteristics of the 17 databases we have used for testing. 12 concerns a motor imagery (MI) BCI, five concerns a P300 BCI. For some databases several sessions are available, therefore the actual number of EEG recordings analyzed is superior to the total number of subjects. Also, one may notice that the number of electrodes used in the experiments is highly variable, ranging from three to 128.

Table 1: Main characteristics of the databases used for the analysis. Legend: Ch.=number of channels; Sess=number of sessions; Ss=number of subjects. For extended names of the databases see Table 2.

Name	Type	Ch	Trials	Sess	Ss
Zhou 2016	MI	14	100	3	4
BNCI 2014-009	P300	18	4200	1	10
BNCI 2015-001	MI	13	200	2 or 3	13
BNCI 2014-002	MI	15	160	1	14
BNCI 2015-003	P300	10	5400	1	10
BNCI 2014-004	MI	3	120-160	5	9
BNCI 2015-004	MI	30	70-80	2	10
BNCI 2014-008	P300	10	4200	1	8
Alexandre MI	MI	16	40	1	9
Weibo 2014	MI	60	160	1	10
Brain Inv 2013a	P300	16	480	1 or 8	24
Cho 2017	MI	64	200	1	49
EPFL P300	P300	32	800	4	8
GW 2009	MI	128	300	1	10
Physionet MI	MI	64	40-60	1	109
Shin 2017a	MI	25	60	3	29
BNCI 2014-001	MI	22	144	2	9

The pipeline for MI databases included: filtering in the 8-32Hz band-pass region, computing the sample covariance matrix for all trials and evaluating the classifiers using (5-fold) cross-validation [3]. The pipeline for P300 included filtering in the 1-24Hz band-pass region and then, during (5-fold) cross-validation, estimating on the training set a spatial filter specifically conceived to enhance the signal-to-noise ratio of event-related potentials (ERPs) [23, 24] retaining the best eight discriminative components, filtering all the trials in the training and test set using this filter, computing the 16x16 extended sample covariance matrix used for ERP data [3, 14] on all trials and finally evaluating the classifiers.

Using these two pipelines we have run statistical tests in MOABB to compare the ROC-AUC classification accuracy of the RMDMF classifier vs. the RMDM classifier for all databases. The power means for the RMDMF classifier where computed for  $h = \{\pm 1, \pm 0.75, \pm 0.5, \pm 0.25, \pm 0.1, \pm 0.01\}$ . If several sessions for the same subject were available, the ROC-AUC score was averaged across-sessions to provide a unique score for each subject. For each database a paired permutation one-sided t-test has been carried out, enumerating all raw data permutations if the number of subjects was  $< 20$ , yielding in this case an exact test [25,26], otherwise employing the Wilcoxon signed-rank test, which basically is equivalent to a permutation test performed on the ranked

data. The  $p$ -values thus obtained for each database have been combined using the weighted *Liptak combination function* [26] (also known as Stouffer's combination function when expressed in terms of standard normal variables), which is given by

$$p = 1 - \Phi \left( \frac{\sum_{i=1}^n w_i \Phi^{-1}(1-p_i)}{\sum_{i=1}^n w_i^2} \right), \quad (2)$$

where  $\Phi$  is the standard normal cumulative distribution function,  $\Phi^{-1}$  its inverse function,  $n$  is the number of databases (17 in our case),  $p_i$  the  $p$ -value observed for the  $i^{\text{th}}$  database and  $w_i$  weights taken as the square root of the number of subjects in each database. This returned a single  $p$ -value for the global one-sided comparison RMDMF vs. RMDM. Such  $p$ -value is to be interpreted as the *probability to observe  $n$   $p$ -values under the Omnibus hypothesis*, i.e., given that the null hypothesis is true for all of them. Notice that the Liptak combination function (2) assumes that the  $p$ -values to be combined result from all pair-wise independent hypotheses, which in our case is verified since all the databases are independent. Notice also that this combination function is optimal (most powerful) when all tests have the same effect size, which, although rarely verified in practice, is a desirable property. The effect sizes were also determined, akin to meta-analysis studies, computing the *standardized mean difference* (SMD) for each database and combining them by the weighted arithmetic average using the same weights as those used for Liptak's  $p$ -value combination method. All statistical analysis tools here above described are already embedded in MOABB [9].

## RESULTS

Table 2 reports the  $p$ -values and SMDs obtained on the 17 databases along with the combined  $p$ -value and the average SMD. Figure 1 depicts the SMDs and their 95% confidence interval obtained on the 17 databases. One can see that for five databases the RMDMF classifier significantly outperforms the RMDM classifier. Among the remaining 12 databases, in nine of them RMDMF tends to perform better than RMDM (as seen by the positive SMD or, equivalently, by a  $p$ -value smaller than 0.5), while the opposite happens in three. There is no evidence that the RMDM significantly outperforms RMDMF for any databases. The weighted average SMD was 0.3 and the weighted combined  $p$ -value was extremely low ( $p=0,0000377$ ), allowing a firm rejection of the Omnibus hypothesis. Notice finally that, as expected, the confidence interval for the SMDs tends to be inversely proportional to the sample size.

## DISCUSSION

Using MOABB we have presented results on 17 databases for a total of 335 individuals. Those results are therefore solid and powerful. For doing this we have added several databases to MOABB, including one of our

own database on the P300 video-game *Brain Invaders* [14-16] and we will continue this effort. MOABB is an ideal framework for testing classifiers objectively and we invite the community to contribute to it in terms of development and data. More in general, we urge the BCI community to promote publications on machine learning for BCI where either real on-line accuracy is reported or the classifiers are tested on large and diverse data. The Python code of this analysis and our data are available along with many other contributions<sup>1</sup>.

Table 2:  $p$ -values and standardized mean differences (SMD) obtained on the 17 databases along with the combined  $p$ -value and SMD. Small  $p$ -values (underlined for  $p<0.05$ ) indicate that the accuracy is significantly higher for the RMDMF as compared to the RMDM.

Name	Type	Ss	$p$	SMD
Zhou 2016	MI	4	0,188	0,679
BNCI 2014-009	P300	10	0,271	0,201
BNCI 2015-001	MI	13	<u>0,004</u>	0,964
BNCI 2014-002	MI	14	<u>0,041</u>	0,499
BNCI 2015-003	P300	10	0,374	0,136
BNCI 2014-004	MI	9	0,648	-0,120
BNCI 2015-004	MI	10	0,322	0,150
BNCI 2014-008	P300	8	0,078	0,659
Alexandre MI	MI	9	0,285	0,182
Weibo 2014	MI	10	<u>0,023</u>	0,711
Brain Invaders 2013a	P300	24	0,744	-0,132
Cho 2017	MI	49	0,085	0,259
EPFL P300	P300	8	0,836	-0,372
Grosse-Wentrup 2009	MI	10	<u>0,011</u>	0,687
Physionet MI	MI	109	<u>0,109</u>	0,090
Shin 2017a	MI	29	0,052	0,304
BNCI 2014-001	MI	9	<u>0,010</u>	0,998
<b>Combination</b>			<b><math>p=0,0000377</math></b>	<b>0,302</b>

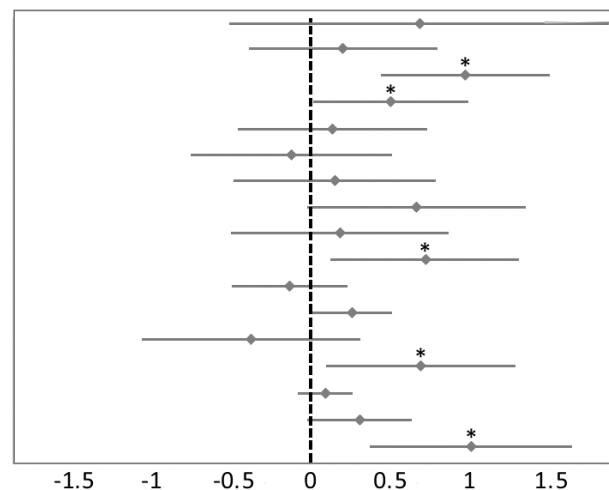


Figure 1: Standardized mean differences (diamond) and their 95% confidence interval (horizontal lines) for the 17 databases (from top to bottom in the same order as in table 1 and 2). A positive SMD value indicate that the accuracy of the RMDMF classifier is higher as compared to the RMDM, the opposite for a negative SMD value. \*= significant  $p$ -value (see table 2).

<sup>1</sup> <https://sites.google.com/site/marcocongedo/science>

## CONCLUSION

In this article we have proposed an improvement to the Riemannian minimum distance to mean classifier for BCIs, leveraging on recent advances on mathematics (the definition of power means for symmetric positive definite matrices [18-21]) and signal processing (an efficient algorithm for estimating them [22]). Comparing the proposed RMDMF to the RMDM classifier on 17 databases using MOABB has yielded an average SMD equal to 0.3 and a combined  $p$ -value equal to 0,0000377. As it can be seen in Table 2, the effect is driven by MI databases, whereas for no P300 database the effect is significant. We believe this is due to the fact that in our pipeline for P300 an optimal spatial filter has been estimated during cross-validation, thus the signal-to-noise ratio for these data has been optimized; under these circumstances the benefit of using a means field populated by power means is likely lost. If this is confirmed by further analysis, the RMDMF classifier would prove better apt for working on raw data as compared to the RMDM, but of course it can be used also on spatially filtered data without losing performance, as our analysis suggests. Then the RMDMF, while preserving the other desirable properties of the RMDM (simplicity, computational efficiency, universality, ease of extension to the multi-user scenario, good attitude for transfer learning and for adaptation), truly would support parameter-free classification pipelines.

The computational cost of the RMDMF as compared to the RMDM is increased proportionally to the number of means used to populate the means field. Since the computational complexity of Riemannian classifiers is cubic on the size of the covariance matrices used to encode the EEG data, this does not represent a substantial additional cost. When the dimension of the covariance matrices is high, it can be reduced by well-known methods such as principal component analysis or by methods inspired by Riemannian geometry [27, 28]. As reported in [6], the performance of the MDM drops in high dimension, therefore in these situations a dimensionality reduction step in practice is necessary. The RMDMF may turn more robust than the RMDM also with respect to the dimensionality. Further research is necessary to verify these hypotheses.

For defining the RMDMF we have found effective a ‘closest mean’ approach. However, better strategies may exist to exploit the richness of the means field, such as, for example, pooling and majority voting. Further research is therefore needed to find optimal strategies for exploiting the Riemannian means fields. Not only the strategy can be optimized, but also the definition of the means field itself: while in this work we have populated the means field with power means, other means not belonging to this family may be added to the field, such as the log-Euclidean mean and a sample of the  $\alpha$ -divergence means, to which the Bhattacharyya mean belongs [29], making the means field even more rich.

## REFERENCES

- [1] Levi-Civita T. *Lezioni di Calcolo Differenziale Assoluto*, Alberto Stock, Roma (1925)
- [2] Bhatia R. *Positive Definite Matrices*, Princeton University Press, New Jersey (2007)
- [3] Barachant A, Bonnet S, Congedo M, Jutten C. Multi-class Brain Computer Interface Classification by Riemannian Geometry. *IEEE Trans. Biomed. Eng.* 2012; 59(4):920-928
- [4] Barachant A, Bonnet S, Congedo M, Jutten C. Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing.* 2013;112:172-178
- [5] Congedo M. *EEG Source Analysis, Habilitation à Diriger des Recherches*, University Grenoble Alpes. (2013)
- [6] Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A, et al. A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces: A 10-year Update. *J. Neural Eng.* 2018; 15(3): 031005
- [7] Congedo M, Barachant A, Bhatia R. Riemannian Geometry for EEG-based Brain-Computer Interfaces; a Primer and a Review. *Brain-Computer Interfaces* 2017; 4(3):155-174
- [8] Yger F, Berar M, Lotte F. Riemannian approaches in brain-computer interfaces: a review. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2017; 25(10):1753-1762
- [9] Jayaram V, Barachant A. MOABB: trustworthy algorithm benchmarking for BCIs. *J. Neural. Eng.* 2018;15(6):066011
- [10] Korczowski L, Congedo M, Jutten C. Single-Trial Classification of Multi-User P300-Based Brain-Computer Interface Using Riemannian Geometry. In *Proc. IEEE EMBS, Milano, Italy, Aug 2015, 1769-72*
- [11] Mayaud L, Cabanilles S, Van Langenhove A, Congedo M, Barachant A, Pouplin S, et al. Brain-computer interface for the communication of acute patients: a feasibility study and a randomized controlled trial comparing performance with healthy participants and a traditional assistive device. *Brain-Computer Interfaces* 2016; 3(4):197-215
- [12] Zanini P, Congedo M, Jutten C, Said S, Berthoumieu Y. Transfer Learning: a Riemannian geometry framework with applications to Brain-Computer Interfaces. *IEEE Trans. Biomed. Eng.* 2018; 65(5):1107-1116
- [13] Rodrigues PLC, Jutten C, Congedo M. Riemannian Procrustes Analysis: Transfer Learning for Brain-Computer Interfaces. *IEEE Trans. Biomed. Eng.* 2019; in press
- [14] Barachant A, Congedo M. A Plug&Play P300 BCI Using Information Geometry, *arXiv*, 2014; 1409.0107
- [15] Congedo M, Goyat M, Tarrin N, Varnet L, Rivet B, Ionescu G, et al. “Brain Invaders”: a prototype of an open-source P300-based video game working with

- the OpenViBE platform. In Proc. Int. BCI Conf., Graz, Austria, 2011, 280-283
- [16] Andreev A, Barachant A, Lotte F, Congedo M. Recreational Applications in OpenViBE: Brain Invaders and Use-the-Force. In: Clerc M, Bougrain L, Lotte F (Ed.). Brain-Computer Interfaces 2: Technology and Application. Wiley-iSTE, New York 2016, pp. 241-258
- [17] Moakher M. A differential geometric approach to the arithmetic and geometric means of operators in some symmetric spaces. *SIAM J. Matrix Anal. Appl.* 2005; 26(3):735-747
- [18] Lim Y, Pálfia M. Matrix Power means and the Karcher mean. *J. Funct. Anal.* 2012; 262:1498-1514
- [19] Pálfia M. Operator means of probability measures and generalized Karcher equations. *Adv. Math.* 2016; 289:951-1007
- [20] Lawson J, Lim Y. Weighted means and Karcher equations of positive operators, *PNAS* 2013; 110(39):15626-32
- [21] Lawson J, Lim Y. Karcher means and Karcher equations of positive definite operators. *Trans. Am. Math. Soc.* 2014; 1:1-22
- [22] Congedo M, Barachant A, Kharati Koopaei E. Fixed Point Algorithms for Estimating Power Means of Positive Definite Matrices. *IEEE Trans. Sig. Process.* 2017; 65(9):2211-2220
- [23] Rivet B, Souloumiac A, Attina V, Gibert G. xDAWN algorithm to enhance evoked potentials: application to brain-computer interface. *IEEE Trans. Biomed. Eng.* 2009; 56(8): 2035-2043
- [24] Congedo M, Korczowski L, Delorme A, Lopes da Silva F. Spatio-temporal common pattern: A companion method for ERP analysis in the time domain. *J. Neurosci. Methods.* 2016; 267: 74-88
- [25] Edgington ES. Randomization tests, 3rd ed., M. Dekker, New York, USA (1995)
- [26] Pesarin F. Multivariate Permutation Tests with applications in Biostatistics, John Wiley and Sons, New York, USA (2001)
- [27] Rodrigues PLC, Bouchard F, Congedo M, Jutten C. Dimensionality Reduction for BCI classification using Riemannian geometry. In Proc. Int. BCI Conf., Graz, Austria, Sept. 2017, 80-85
- [28] Congedo M, Rodrigues PLC, Bouchard F, Barachant A, Jutten C. A Closed-Form Unsupervised Geometry-Aware Dimensionality Reduction Method in the Riemannian Manifold of SPD Matrices. In Proc. Int. EMBS Conf., Jeju Island, South Korea, July 2017, 3198-3201
- [29] Congedo M, Afsari B, Barachant A, Moakher M. Approximate Joint Diagonalization and Geometric Mean of Symmetric Positive Definite Matrices. *PLoS One.* 2015; 10(4): e0121423