Review of Riemannian Distances and Divergences, Applied to SSVEP-based BCI

Sylvain Chevallier, E.K Kalunga, Q. Barthélemy, E. Monacelli

Neuroinformatics manuscript No.
(will be inserted by the editor)

Review of Riemannian distances and divergences, applied to SSVEP-based BCI

S. Chevallier · E. K. Kalunga · Q. Barthélemy · E. Monacelli

Received: date / Accepted: date

Abstract

Background The first generation of brain-computer interfaces (BCI) classifies multi-channel electroencephalographic (EEG) signals, enhanced by optimized spatial filters. The second generation directly classifies covariance matrices estimated on EEG signals, based on straightforward algorithms such as the minimum-distance-to-Riemannian-mean (MDRM). Classification results vary greatly depending on the chosen Riemannian distance or divergence, whose definitions and reference implementations are spread across a wide mathematical literature.

Methods This paper reviews all the Riemannian distances and divergences to process covariance matrices, with an implementation compatible with BCI constraints. The impact of using different metrics is assessed on a steady-state visually evoked potentials (SSVEP) dataset, evaluating centers of classes and classification accuracy.

Results and Conclusions Riemannian approaches embed crucial properties to process EEG data. The Riemannian centers of classes outperform Euclidean ones both in offline and online setups. Some Riemannian distances and divergences have better performances in terms of classification accuracy, while others have appealing computational efficiency.

Keywords Riemannian geometry · covariance matrices · distances · divergences · EEG · SSVEP

E. K. Kalunga
Orderin
Cape Town, South Africa

Q. Barthélemy
Foxstream
69120 Vaulx-en-Velin, France

S. Chevallier (corresponding author)
E. Monacelli
Université Paris-Saclay, UVSQ, LISV,
78124, Vélizy -Villacoublay, France
E-mail: sylvain.chevallier@uvsq.fr

1 Introduction

Brain-computer interfaces (BCI) provide humans and animals with the ability to interact with machines without the use of their motor functions. They offer alternatives for the compensation of physical limitations (McFarland et al. 2017), with applications in rehabilitation and assistive technologies (Hosni et al. 2018). Brain signals could be recorded with various systems, with the electroencephalography (EEG) being widely adopted (Niedermeyer and Silva 2005) as it has a low cost and requires low computational power. BCI designs commonly rely on one of the three following paradigms (Nam et al. 2018; Quitadamo et al. 2008): the event-related (de)-synchronization (ERD/S), the event-related potentials (ERP) and the steady-state visually evoked potentials (SSVEP). ERD/S are detected in the premotor cortex and are correlated with motor imagery. ERP are the observed brain responses to an external stimulation of the sensory nervous system, often auditory or visual in BCI. SSVEP are elicited by the attentional focus on a steady pattern of a visual stimulus (Wang et al. 2004). Auditory or somatosensory patterns could be considered, but visual stimulation is the most common in BCI (Zhang et al. 2015). In SSVEP paradigm, a stimulus flickering at constant frequency is generated by LEDs or by alternating images on a computer screen. When a user's gaze is directed toward a flickering stimulus, it elicits cortical oscillations with matching fre-

quency. Those cortical oscillations are modulated by the selective attention (Vialatte et al. 2010). The experiments conducted in this paper focus on SSVEP-based BCI due to their general robustness and efficiency (Volosyak et al. 2011).

BCI face important challenges that are mainly caused by the poor spatial resolution of EEG. This limitation is due to the volume conduction effect (Niedermeyer and Silva 2005); the bone structure filters out the higher frequency part, mixing the brain sources signals and reducing the signal-to-noise ratio (SNR). Consequently, spatial filtering methods are developed to enhance SNR, such as common spatial patterns (CSP) (Johannes et al. 1999), independent component analysis (ICA) (Wang and James 2006), xDAWN (Rivet et al. 2009) or canonical correlation analysis (CCA) (Yang et al. 2017). Filtered signals define a feature space where machine learning methods are used to classify trials. While spatial filters are very efficient on clean datasets obtained in constrained environments, they are easily distorted by artifacts and outliers (Tomioka et al. 2007), and are specific to subjects and sessions.

All the above-mentioned spatial filters rely on complex computations based on covariance matrices estimated from the EEG signal. Covariance matrices have a specific structure and are defined on a subspace of the Euclidean space. Classical matrix computations – such as Euclidean distance, ℓ^2 norm or arithmetic mean – results in inaccurate results when applied directly on covariance matrices. It also induces numerical artifacts, like the *swelling effect* (Arsigny et al. 2007). The Riemannian framework accounts for the specific geometry of the space of covariance matrices. It embeds specific properties, like affine-invariance, resulting in equivalent distances between sensor-based covariance matrices or between source-based covariance matrices (Congedo et al. 2017a; Yger et al. 2016). This invariance to full-rank spatial filtering provides good generalization capabilities to classification algorithms. Recent works (Barachant et al. 2012; Kalunga et al. 2016; Meinel et al. 2019) demonstrate the benefit of Riemannian framework. Riemannian approaches are also applied in the winning strategy of five BCI competitions (Congedo et al. 2017a). For EEG classification, a very straightforward approach introduced in (Barachant et al. 2012) is the known as *minimum distance to Riemannian mean* (MDRM). The class centers are defined as the average of covariance matrices for each class and an unseen trial is assigned to the class with the nearest center, in the sense of a Riemannian distance. This straightforward approach is robust and yields better accuracy than more sophisticated classifiers of the literature (Barachant et al. 2012), especially on systems with few electrodes (Yang et al. 2016).

The space of covariance matrices is curved and proper distances are geodesics, *i.e.* curves that flow along the surface of the space. Different distances are defined in the mathematical literature, each with specific properties that could be applied to the space of covariance matrices (Arsigny et al. 2007). Still, there is no prior work that indicates the most appropriate distances to process EEG signals, or even to propose a thorough and reproducible analysis.

While most applications of Riemannian geometry on EEG are dedicated to motor imagery and ERP paradigms (Congedo et al. 2017a; Yger et al. 2016), this work focuses on SSVEP-based BCI. The experimental part is applied on a dataset acquired in the context of assistive robotics, aiming toward applications for cerebral palsy and post-stroke rehabilitation. The task follows a shared control paradigm, where an arm exoskeleton is operated through a controller shared between an SSVEP-based BCI and a 3D touchless interface based on infrared sensors (Martin et al. 2012; Kalunga et al. 2014).

The relevant properties to cope with brain signal variability have not been identified, and thus EEG signal processing is lacking the proper metrics to conduct deeper investigation. To address several open questions in EEG applications, such as handling specific noise, change in electrode positioning, inter and intra-individual variability, advances in Riemannian BCI could open new opportunities to develop applicative solutions. Such opportunities require a thorough analysis of existing distances and their properties, in a reproducible context.

Advances in Riemannian BCI have opened new opportunities to address several open questions in EEG applications, such as handling specific noise, change in electrodes positioning, coping with inter and intra-individual variability. However, the relevant metrics and their properties that could encompass the brain signal variability have not been formally identified. It is therefore important to identify the relevant metrics through an extended analysis in a reproducible context.

While we start to investigate these issues in a previous work (Kalunga et al. 2015), this paper extends this former approach and the contributions are the following:

- a review of classification methods for SSVEP-based BCI, showing remaining challenges such as resting-state class,
- a complete survey of the existing distances and divergences for covariance matrices for the first time in the literature,

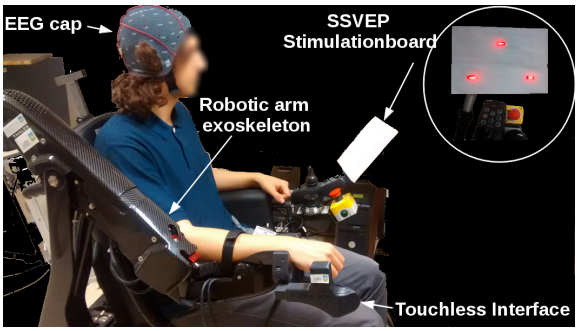


Fig. 1 SSVEP-based BCI setup for controlling an arm exoskeleton on an electric-powered wheelchair. The SSVEP stimulation board is made of 3 groups of LEDs that blink at different frequencies.

- an evaluation of the computational cost, numerical stability and a qualitative analysis of results of those distances and divergences,
- an open framework which includes the algorithms in Matlab (Kalunga 2018, 2015) and Python (Chevallier 2020; Bertrand-Lalo 2020), and the datasets (Chevallier 2017) to ensure the reproducibility of experiments.

This paper is organized as follows. Section 2 details the neurophysiological ground of the steady-state visually evoked potentials and reviews the existing methods for their classification. Section 3 describes the framework for the classification of covariance matrices, including the considered distances and divergences, and presents the minimum distance to mean (MDM), which is the classification algorithm used in this work. The real EEG dataset used for the experiment is presented in Section 4. In Section 5, the results of both offline and online analyses are shown and discussed. Section 6 concludes this paper.

2 SSVEP-based BCI

2.1 Stimulus design

Steady-state visual evoked potential is a brain response to a repeated visual stimulus. It is measured in the primary visual cortex as oscillations at a frequency which corresponds to the visual stimulus frequency and its harmonics (Vialatte et al. 2010). In BCI, while various possible stimuli design are possible (Gergondet and Kheddar 2015), SSVEP is triggered by a flashing visual target using either LEDs or a computer screen. The LEDs induce larger oscillation amplitude than a computer screen (Zhu et al. 2010). A classical setup with LEDs is visible on Fig. 1. Computer screens also have a limitation in terms of the flashing frequencies that could be produced, as it should be a multiple of R, the screen refresh rate. This limitation is overcome by the sampled sinusoidal stimulation method that allows to simulate intermediate frequencies. These intermediate frequencies are obtained by modulating the luminance of the screen using the following equation (Chen et al. 2014):

$$s(t, f, \theta) = \frac{1}{2}\Big(1 + \sin\big(2\pi f\,(t/R) + \theta\big)\Big)\ ,$$

where $s(t, f, \theta)$ is the luminance of the screen at time t, varying with frequency f and phase θ. In this stimulation mode, a phase shift θ is added to differentiate between target stimulus with identical oscillation frequencies, allowing for more possible targets. Nonetheless, paradigms that require the use of phase information add a constraint on the BCI system: the SSVEP stimulation device should be synchronized with the EEG acquisition system, usually with wire cables.

Any frequency, between 2 Hz and 50 Hz, induces visible oscillations in the visual cortex (Herrmann 2001), the amplitude of the oscillations decreases as the frequency increase. Frequencies as high as 90 Hz can be used for SSVEP stimulation and could improve the visual comfort of the subject (Chen et al. 2015a), but are difficult to detect. Visual stimuli between 12 Hz and 25 Hz are the most common in SSVEP-BCI as this range provides a good compromise between detectability and reactivity (Zhu et al. 2010).

2.2 Classification with CCA

Classification of SSVEP brain signals relies on the identification of stimulus-specific frequency in the visual cortex. Before advances introduced by Riemannian frameworks, methods relying on CCA (Lin et al. 2007) achieved the highest classification performances. CCA finds a projection space that maximizes the cross-covariance of two sets of input signals while jointly minimizing their covariance (Kalunga et al. 2013). The common methodology is to find the canonical space between a multichannel EEG trial X, on the one hand, and the reference signal Y_f, on the other hand. Y_f is usually defined as:

$$Y_f = \begin{bmatrix} \sin(2\pi f t) \\ \cos(2\pi f t) \\ \vdots \\ \sin(2\pi N_h f t) \\ \cos(2\pi N_h f t) \end{bmatrix}\ ,$$

where N_h is the number of harmonics, and frequency f belongs to the set of stimulation frequencies. The goal

of the CCA is to maximize the correlation ρ_f between $w_X^T X$ and $w_Y^T Y_f$ as follows:

$$\rho_f = \max_{w_X, w_Y} \frac{w_X^T \Sigma_{XY_f} w_Y}{\sqrt{w_X^T \Sigma_X w_X w_Y^T \Sigma_{Y_f} w_Y}} , \tag{1}$$

where Σ_X and Σ_{Y_f} are the covariance matrices of X and Y_f, and Σ_{XY_f} is the cross-correlation matrix of X and Y_f. The estimated frequency $\hat{f}$ is computed from the values of ρ_f, as in Lin et al. (2007):

$$\hat{f} = \arg\max_f \{\rho_f\} . \tag{2}$$

It is possible to optimize the results by carefully selecting the reference signals and its harmonics, but there is no specific recommendation.

With Filter Bank Canonical Correlation Analysis (FBCCA) (Chen et al. 2015a), an input signal is decomposed in $n = 1 \dots N_b$ sub-bands with overlapping frequency bands. The correlation coefficients ρ_f^n are computed for each sub-band n and are combined as:

$$\tilde{\rho}_f = \sum_{n=1}^{N_b} w_n \, (\rho_f^n)^2 \quad \text{with} \quad w_n = n^{-a} + b , \tag{3}$$

where a and b are constants defining the weight w_n of each sub-band. To maximize the classification performance, an offline grid search is needed to select the correct hyper-parameters a, b, and N_b. The classification decision is reached with Eq. (2) as soon as the all the $\tilde{\rho}_f$ are computed. There is a significant improvement in terms of accuracy from the classical CCA with this filter bank scheme.

An important milestone has been reached with studies conducted by Nakanishi et al. (2014) and Chen et al. (2015b) for SSVEP-based spellers. These SSVEP-based BCI achieve the highest information transfer rate (McFarland et al. 2003) reported in the EEG literature: Nakanishi et al. (2014) brought the information transfer rate to an average of 166.91 bits/min and later, Chen et al. (2015b) raised it to an average of 270 bits/min.

These very high performances rely mostly on the improvement of two key aspects: the stimulation protocol, using phase modulation and combination of frequencies, and stimulus presentation, using a scheme similar to SMS writing. These approaches comes with two strong limitations: the first drawback is that there is no resting-state (also known as reject class, or no-SSVEP class). Even without any intent from the subject, the BCI system is selecting letters without any pause. The ability for a user to act at their own pace is crucial in Human-Machine Interface and in BCI, a system without any resting-state is hardly compatible for a real implementation of a BCI system. It should also be emphasized that handling the diversity of brain signals corresponding to resting-state is a challenging issue as stimulus frequency could still be observed in the EEG. The second limitation is the complexity of the protocol, which requires a hardware synchronization between the stimulation device and the EEG acquisition system for precise phase measurements. This limits potential applications.

A complete review of classification methods for SSVEP-based BCI can be found in Lotte et al. (2018).

3 Riemannian BCI

Spatial filters and classifiers in BCI rely on covariance matrices, as it could be seen in Eq. (1) for CCA. It is possible to operate directly on the covariance matrices of EEG signals, with formal tools of differential geometry and algorithms borrowed from the mathematical community of information geometry (Nielsen and Bhatia 2012). This section introduces the covariance matrices, their geometry and the two most important notions for EEG signal processing methods, which are the computation of distance between two covariance matrices and the estimation of the mean of a set of covariance matrices. Endowed with a distance and an average, it is possible to define all the required methods for processing and classifying EEG signals in the space of covariance matrices.

3.1 From EEG signals to covariance matrices

In the context of EEG signal, we will denote by X an EEG signal in $\mathbb{R}^{C \times N}$ where C is the number of channels, *i.e.* sensors that are placed on the subject's scalp, and N is the number of samples, which are the observations recorded by the system, with $N > C$. The covariance matrix is estimated with

$$\Sigma = \frac{1}{N} X X^T . \tag{4}$$

For noisy or truncated signal is possible to use more advanced estimator, an experimental analysis of covariance estimators applied to EEG signal is detailed in Kalunga et al. (2016). Covariance matrices are symmetric and positive-definite (SPD) and it is possible to define a manifold $\mathcal{M}$ of SPD $C \times C$ matrices (Bhatia 2009) as:

$$\mathcal{M} = \left\{ \Sigma \in \mathbb{R}^{C \times C} \mid \Sigma = \Sigma^T \text{ and } z^T \Sigma z > 0, \forall z \in \mathbb{R}^C \backslash 0 \right\} .$$

The eigenvalue decomposition of Σ is

$$\Sigma = U \, \mathrm{diag}(\lambda_1, \dots, \lambda_C) \, U^T ,$$

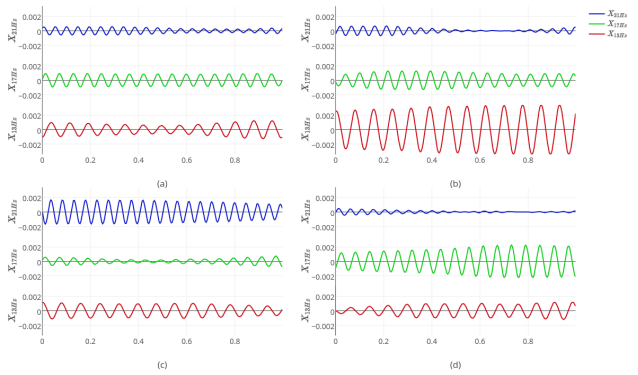


Fig. 2 Samples of EEG trials from Eq. (5). For each sub-band X_{freq_f} only the channel Oz is shown. Each subplot shows the first second of a trial from classes: resting-state (a), 13 Hz (b), 21 Hz (c), 17 Hz (d).

where $\lambda_1, \ldots, \lambda_C$ are the eigenvalues and U the matrix of eigenvectors of Σ. The unique symmetric square root $\Sigma^{\frac{1}{2}}$ and symmetric inverse square root $\Sigma^{-\frac{1}{2}}$ are defined as:

$$\Sigma^{\frac{1}{2}} = U \operatorname{diag}(\lambda_1^{\frac{1}{2}}, \ldots, \lambda_C^{\frac{1}{2}})\, U^T \ ,$$
$$\Sigma^{-\frac{1}{2}} = U \operatorname{diag}(\lambda_1^{-\frac{1}{2}}, \ldots, \lambda_C^{-\frac{1}{2}})\, U^T \ ,$$

and the exponential Exp and logarithm Log matrix operators are defined as:

$$\operatorname{Exp}(\Sigma) = U \operatorname{diag}(\exp(\lambda_1), \ldots, \exp(\lambda_C))\, U^T \ ,$$
$$\operatorname{Log}(\Sigma) = U \operatorname{diag}(\log(\lambda_1), \ldots, \log(\lambda_C))\, U^T \ .$$

Following Congedo et al. (2017a), specific covariance matrices for SSVEP are designed to embed frequency information. Let X be an EEG trial measured during an SSVEP experiment with F stimuli blinking at different frequencies. The covariance matrices are estimated from an extended version of the input signal X:

$$X \in \mathbb{R}^{C \times N} \rightarrow \begin{bmatrix} X_{\text{freq}_1} \\ \vdots \\ X_{\text{freq}_F} \end{bmatrix} \in \mathbb{R}^{FC \times N} \ , \tag{5}$$

where a band-pass filter around the frequency freq_f, with $f = 1, \ldots, F$, is applied on the input signal X to obtain X_{freq_f}. Fig. 2 show the extended signal of one electrode, filtered at $F = 3$ different frequencies. Hereafter, EEG signals always refer to the augmented signal of Eq. (5).

3.2 Distances and divergences

In Riemannian framework, several distances have been proposed for positive-definite matrices with different properties, yielding various impacts on signal processing and classification of brain signals. There is a large literature in differential geometry, probability and matrix calculus describing a great variety of distances for positive-definite matrices, whose definitions are outside the scope of this paper. In this large literature, there are many reformulations of existing distances and generalization. Some distances cannot be computed while others have very efficient implementations available; some are very sensitive to the perturbations while others have useful invariances for robustness to noise. This paper summarizes all this literature and provides the precise bibliographical references for the distance that could be of use for Riemannian BCI.

Divergences and distances are measures of dissimilarity between two points in a space. A distance function d has the following properties for all $\Sigma_1, \Sigma_2, \Sigma_3 \in \mathcal{M}$:

1. Non-negativity: $d(\Sigma_1, \Sigma_2) \geq 0$,
2. Identity: $d(\Sigma_1, \Sigma_2) = 0$ iff $\Sigma_1 = \Sigma_2$,
3. Symmetry: $d(\Sigma_1, \Sigma_2) = d(\Sigma_2, \Sigma_1)$,
4. Triangular inequality: $d(\Sigma_1, \Sigma_3) \leq d(\Sigma_1, \Sigma_2) + d(\Sigma_2, \Sigma_3)$.

Divergences, which are generalization of squared distance, do not necessarily satisfy properties (3) and (4).

In this work, to compare covariance matrices, we consider several existing distances, namely the Euclidean distance, the harmonic distance (Lim and Pálfia 2012), the affine-invariant Riemannian distance (Pennec et al. 2006), the Log-Euclidean distance (Arsigny et al. 2007), the Wasserstein distance (Villani 2008), and divergences, such as the Kullback-Leibler divergence (Nielsen and Nock 2009) (equivalent to the log-det divergence (Dhillon and Tropp 2007) for centered multivariate Gaussian distributions), the Jeffreys divergence (Jeffreys 1946), the S-divergence (Sra 2016) (equivalent to the Bhattacharyya divergence (Chebbi and Moakher 2012)), and the α-divergence family (Cichocki and Amari 2010; Nielsen et al. 2014). All these distances and divergences are detailed in Supplementary Material.

Table 1 provides consistent definitions of all these distances along with pointers to most efficient implementations. It is important to select the required properties carefully for a specific problem, as it could provide a measure that is robust to noise and undesired changes.

3.3 Riemannian mean

Considering the set of covariance matrices $\{\Sigma_i\}_{i=1\ldots I}$, the mean (also called barycenter or center of mass) $\bar{\Sigma}$

Table 1 Formula of distances, divergences and means considered in this article.

	Distance/Divergence	Mean	References
Arithmetic	$d_{\mathrm{E}} = \|\Sigma_1 - \Sigma_2\|_F$	$\bar{\Sigma}_{\mathrm{E}} = \frac{1}{I}\sum_{i=1}^{I} \Sigma_i$	
Harmonic	$d_{\mathrm{H}} = \|\Sigma_1^{-1} - \Sigma_2^{-1}\|_F$	$\bar{\Sigma}_{\mathrm{H}} = \left(\frac{1}{I}\sum_{i=1}^{I} \Sigma_i^{-1}\right)^{-1}$	(Lim and Pálfia 2012)
Log-Euclidean	$d_{\mathrm{LE}} = \|\mathrm{Log}(\Sigma_1) - \mathrm{Log}(\Sigma_2)\|_F$	$\bar{\Sigma}_{\mathrm{LE}} = \mathrm{Exp}\left(\frac{1}{I}\sum_{i=1}^{I} \mathrm{Log}(\Sigma_i)\right)$	(Arsigny et al. 2007)
Affine-invariant	$d_{\mathrm{AIR}} = \|\mathrm{Log}(\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2})\|_F$	Algorithm 3 in (Fletcher and Joshi 2004)	(Moakher 2005; Fletcher and Joshi 2004)
Kullback-Leibler	$D_{\mathrm{KL}} = \frac{1}{2}\left(\log\frac{\det(\Sigma_2)}{\det(\Sigma_1)} + \mathrm{tr}(\Sigma_2^{-1}\Sigma_1) - C\right)$	Algorithm 1 in (Chebbi and Moakher 2012)	(Chebbi and Moakher 2012)
Jeffreys	$D_{\mathrm{J}} = \frac{1}{2}\left(\mathrm{tr}(\Sigma_2^{-1}\Sigma_1) + \mathrm{tr}(\Sigma_1^{-1}\Sigma_2)\right) - C$	$\bar{\Sigma}_{\mathrm{J}} = \bar{\Sigma}_{\mathrm{E}}^{\frac{1}{2}}\left(\bar{\Sigma}_{\mathrm{E}}^{-\frac{1}{2}}\bar{\Sigma}_{\mathrm{H}}\bar{\Sigma}_{\mathrm{E}}^{-\frac{1}{2}}\right)^{\frac{1}{2}}\bar{\Sigma}_{\mathrm{E}}^{\frac{1}{2}}$	(Moakher and Batchelor 2006)
S-divergence	$D_{\mathrm{S}} = \log\det(\frac{\Sigma_1+\Sigma_2}{2}) - \frac{1}{2}\log\det(\Sigma_1\Sigma_2)$	Eq. (17-20) in (Cherian et al. 2011)	(Sra 2016; Cherian et al. 2011)
α-divergence	D_{LD}^{α} from Eq. (3) in (Chebbi and Moakher 2012)	Algorithm 1 in (Chebbi and Moakher 2012)	(Chebbi and Moakher 2012)
Wasserstein	$d_{\mathrm{W}} = \left(\mathrm{tr}\left(\Sigma_1 + \Sigma_2 - 2\left(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2}\right)^{1/2}\right)\right)^{1/2}$	Eq. (19) in (Álvarez-Esteban et al. 2016)	(Agueh and Carlier 2011; Álvarez-Esteban et al. 2016)

minimizes the dispersion of matrices Σ_i and could be expressed as:

$$\bar{\Sigma} = \mu(\{\Sigma_1, \ldots, \Sigma_I\}) = \arg\min_{\Sigma\in\mathcal{M}} \sum_{i=1}^{I} d^p(\Sigma_i, \Sigma) \ , \qquad (6)$$

for $p = 2$ $d^p(\cdot,\cdot)$ is a distance and for $p = 1$ it is a divergence. $\bar{\Sigma}$ is unique and is denoted with various names in the literature, such as geometric mean, or Frechet mean. It is sometimes called Cartan mean or Karcher mean (Ando et al. 2004; Lim and Pálfia 2012). Cartan had shown that a unique solution to (6) exists if all Σ_i lie in a convex ball (Cartan 1929).

The choice of the divergence or the distance defined in Eq. (6) has an important impact on the resulting mean, that will be explored hereafter. In particular, for the affine-invariant Riemannian (AIR) distance, the geometric mean is the unique matrix $\bar{\Sigma}_{\mathrm{AIR}}$ of the manifold $\mathcal{M}$ satisfying:

$$\sum_{i=1}^{I} \bar{\Sigma}_{\mathrm{AIR}}^{1/2} \mathrm{Log}(\bar{\Sigma}_{\mathrm{AIR}}^{-1/2}\Sigma_i\bar{\Sigma}_{\mathrm{AIR}}^{-1/2})\bar{\Sigma}_{\mathrm{AIR}}^{1/2} = \sum_{i=1}^{I} \mathrm{Log}_{\bar{\Sigma}_{\mathrm{AIR}}}(\Sigma_i) = 0 \ ,$$

where $\mathrm{Log}_{\bar{\Sigma}_{\mathrm{AIR}}}(\Sigma_i)$ is the logarithmic map projecting matrix Σ_i from manifold to tangent space at point $\bar{\Sigma}_{\mathrm{AIR}}$. Existence and unicity of this Riemannian mean can be proved (Moakher 2005; Afsari 2011; Massart and Chevallier 2017).

A summary of distances and divergences considered in this study, as well as their associated means, is proposed in Table 1, and their properties are listed in Table 2.

3.4 Multi-class classification of covariance matrices for SSVEP

Minimum Distance to Mean (MDM) (Barachant et al. 2012) allows to directly classify covariance matrices in the Riemannian space. The training step consists in computing one center of class $\bar{\Sigma}_k$ for each class k as:

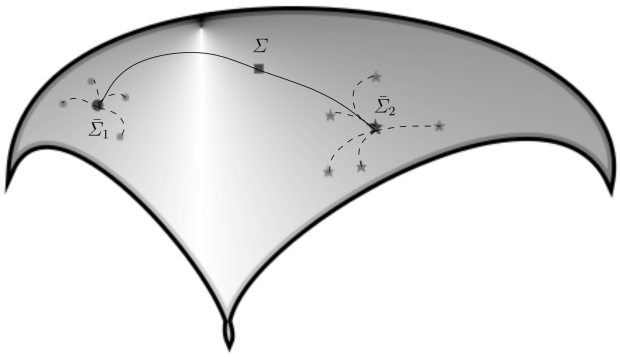


Fig. 3 Illustration of the MDM principle for a biclass problem in a curved space.

$$\bar{\Sigma}_k = \mu(\{\Sigma_i\}_{i\in\mathcal{I}(k)}) \ , \qquad (7)$$

where $\mathcal{I}(k)$ is the set of indices of training data belonging to class k, and $\mu(.)$ is the mean estimation function. For generalization, a new matrix Σ is assigned to the class of the closest mean, as illustrated on Fig. 3:

$$\hat{k} = \arg\min_{k=1\ldots K} d(\Sigma, \bar{\Sigma}_k) \ , \qquad (8)$$

where $d(.)$ is one of the distances from Table 1.

The classification of SSVEP signals discriminates between $K = F + 1$ classes, to take into account the resting-state class. From I labeled training trials $\{X_i\}_{i=1}^{I}$ recorded per subject, K centers of classes $\bar{\Sigma}_k$ are estimated. To obtain a correct estimation of the centers of classes, the artifacts are removed with the Riemannian potato (Barachant et al. 2013) used with a z-score threshold $z_{th} = 2.0$ and a number of iterations $L = 3$.

The MDM algorithm described above is suitable for offline and block-online BCI implementations (Chevallier et al. 2018). Offline implementation requires a reference time, that is aligned with the cue onset. This

Table 2 Properties of distances and divergences of SPD matrices.

Distance/Divergence	Restriction on eigenvalues	Invariances
Arithmetic	None	Rotation
Harmonic	None	Rotation
Log-Euclidean	Strictly positive	Rotation, inversion, similarity
Affine-invariant	Strictly positive	Rotation, inversion, similarity, affine transform
Kullback-Leibler	Strictly positive	Rotation, affine transform
Jeffreys	Strictly positive	Rotation, inversion, affine transform
S-divergence	Strictly positive	Rotation, inversion, affine transform
α-divergence	Strictly positive	Affine transform
Wasserstein	Positive	Rotation

reference time is used to estimate the covariance matrices, thus focusing on the SSVEP occurrences. In online and asynchronous setups, it is not possible to rely on such cue onset: the EEG epochs should be classified iteratively. An online computation scheme is applied on overlapping epochs, as described in Kalunga et al. (2016), and the length of the trial is determined with a dynamic stopping approach (Verschore et al. 2012). This avoids false detection induced by inattention of the subject or by external noise. The online computation scheme makes use of a curve direction criterion that shortens the synchronization latency by reducing the delays of the epoch-based decision process.

To conclude this section, MDM and MDM-online algorithms will be used in subsequent experiments to evaluate the classification accuracy of different distances and divergences.

4 Experimental SSVEP dataset

This section describes how the experiments have been conducted to obtain the data for the experimental analysis presented in the next section. The subject sits on an electric-powered wheelchair and an exoskeleton supports their right upper limb (Martin et al. 2012; Kalunga et al. 2014). During the experimental recording, the exoskeleton is powered on but not used. On the left side of the subject, a panel shows 3 groups of 4 LEDs, blinking each at specific frequency (see Fig. 1). Although the panel is on the left side, the user could see it without moving their head. The subject is asked to follow audio instruction, while moving and blinking freely. In the protocol, a trial starts by an audio cue prompting the subject to gaze either at a group of LEDs or at a fixation point positioned outside the LED panel. This fixation point is used to characterize the reject class, which is a reference state without any associated SSVEP command.

A g.Mobilab+ is used for recording EEG, with $C = 8$ channels. The electrodes are positioned in the cap at PO7, PO8, PO3, PO4, POz, O1, O2 and Oz, with an earlobe reference and a ground electrode in Fz. The impedance is kept below 10 kΩ throughout the experiments. The EEG sampling rate is 256 Hz, with a band-pass filtering between 0.5 Hz and 100 Hz, and no spatial referencing is performed to avoid obtaining low-rank matrix. An example of EEG recording, extended as per Eq. (5), is shown on Fig. 2.

Three distinct visual stimuli ($F = 3$) characterized by their respective blinking frequencies of 13, 21, and 17 Hz, are presented to the subject. Thus, the classification task is set up for $K = 4$ classes, that is the $F = 3$ classes of the stimuli and the resting-state (no-SSVEP). Each session is made of 32 trials, divided in 8 visual stimuli for each class including the resting-state class. A trial is 4 second long. These experiments gather 12 subjects and, for each subject, there is between 2 and 5 sessions. The data are split into $I = 32$ trials for training set and the 32 to 128 trials left are for the test set. This dataset is accessible at Chevallier (2017), with all the information to extract and process the EEG markers.

5 Experimental evaluation of distances and divergences

This section presents the analysis and experimental results of applying Euclidean and Riemannian distances and divergences in SSVEP classification task.

5.1 Assessment of centers of class

Covariance matrices of extended EEG signal, defined in Eq. (5), capture discriminative features of EEG generated by same brain sources but with different oscillating frequencies. The center of mass $\bar{\Sigma}_k$ for the K classes is shown on Fig. 4, with left (*resp.* right) part showing the subject with the highest classification results (*resp.* lowest). The diagonal blocks are the 8×8 covariance matrices computed for the $F = 3$ target frequencies, the off-diagonal blocks are almost null as there is a null inter-frequency covariance. The block with the largest values is almost always indicating the target frequency. In the case of the resting state, all F blocks have equivalent values, with small absolute value. Indeed, this is most visible for the subject with the highest classification rate and more difficult to see for the one with the lowest classification rate. The mean covariance for 13 Hz is very similar to the resting-state, and the first block of the mean covariance for 21 Hz displays non-negligible values. For this subject, the mean covariance matrices are very informative but are not discriminative. The subject with the highest classification accuracy has relatively lower within frequency correlations: each class has a higher within frequency correlation in its corresponding target frequency, and the resting-state class has all within frequency correlations equally reduced. Consequently, the 4 classes are thus easily separable.

Classifiers commonly used in BCI, such as linear discriminant analysis (LDA) or support-vector machine (SVM) (Hastie et al. 2009), can appear as *black-boxes* with decisions that are difficult to interpret. The presented classifier relies on features associated with a simple representation. This allows an intuitive understanding for the domain experts, such as medical doctors, neuroscientists, etc. One of the most interesting aspect is that there is a physiological meaning to the observed covariance matrices. Most of the complexity is embedded in a dedicated distance, allowing a simple interpretation, whereas machine learning algorithms produce complex decisions that are difficult to explain. Understanding why a subject shows some poor classification results is easily inferred from the values of the mean covariance matrices.

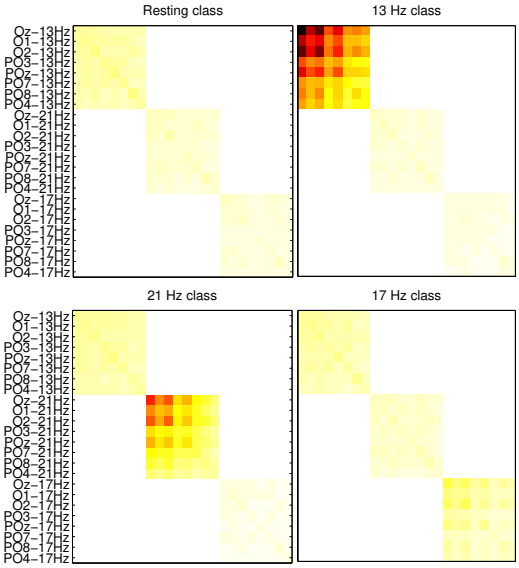
Fig. 5 shows the classification accuracy and CPU time, for all distances and divergences listed in Table 1, except for the Wasserstein which has a very high computational cost and a low accuracy. Firstly, Fig. 5a shows the computation time of averaging, to estimate the $K = 4$ centers of class $\bar{\Sigma}_k$ during MDM offline training. Secondly, Fig. 5b shows the computation time of distances and divergences elapsed on the trial classification of test data. The computation times are measured on every distance and divergence, then averaged across all subjects and trials. In an online setting where the target center of mass is updated at every trial, a classification using a specific metric would take the time to compute one distance/divergence plus the time to update the center of mass. These figures show that S-divergence and Bhattacharyya divergence, which are formally equivalent (Bhattacharyya 1943; Sra 2016) but with different implementations, yield similar accuracy but different computation times. They also constitute a reference for compatibility in applications with real-time constrains.

5.2 Offline classification

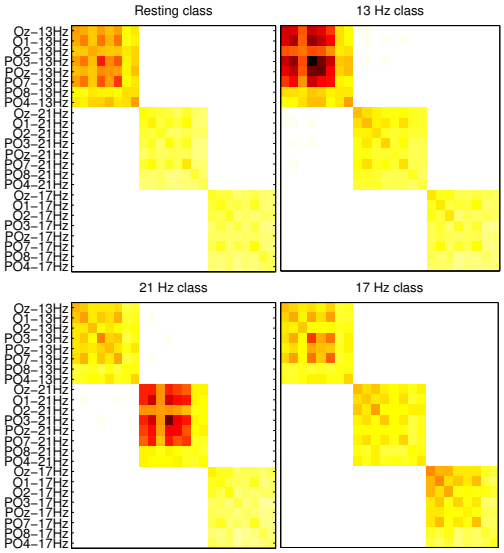
On Fig. 6, the performances achieved with offline MDM using the above-mentioned distances and divergences are compared to the three state-of-the-art methods: Lin et al. (2006), Nakanishi et al. (2014), and FBCCA of Chen et al. (2015a). All these methods are described in Section 2. Only classes corresponding to the stimulation frequencies ($F = 3$) are considered for the offline classification, $K = 3$, as CCA-based methods are not able to identify resting-state class. The harmonic number is set to $N_h = 6$ for the CCA based methods, this value is selected after a grid search on the training data. The signal is decomposed in N_b sub-bands for the FBCCA, where the sub-band n ranges from $n \times 8$ Hz to 93 Hz. The maximum value is ten times the frequency span of the stimulation: in the described experiment, 8 Hz corresponding to the span from 13 Hz to 21 Hz. The parameters N_b, a and b are selected by a grid search in the following range: $N_b = i_N$, with $i_N = 1, \ldots, 7$, $a = 0.25 \times i_a$, with $i_a = 0, 1, \ldots, 40$, $b = 0.25 \times i_b$, and with $i_b = 0, 1, \ldots, 4$. The selected values are $N_b = 3$, $a = 2$ and $b = 0$.

5.2.1 Centers of class and classification performances

When evaluating the results of offline MDM, the arithmetic distance yields the poorest performance, far lower than the results obtained with the state-of-the art methods, as it could be seen on Fig 6. This poor performance is attributed to the lack of invariance under inversion and to the *swelling effect* (see Supplementary Material). The determinant value is tightly linked to the dispersion of multivariate variables, here the frequency bands and the EEG electrodes, thus leading to low classification accuracy. The swelling effect of the arithmetic mean is shown on the TraDe plot (Congedo et al. 2017b) of Fig. 7, which displays the log-trace as a function of the log-determinant for the different covariance matrices taken from the trials of the 13 Hz class of



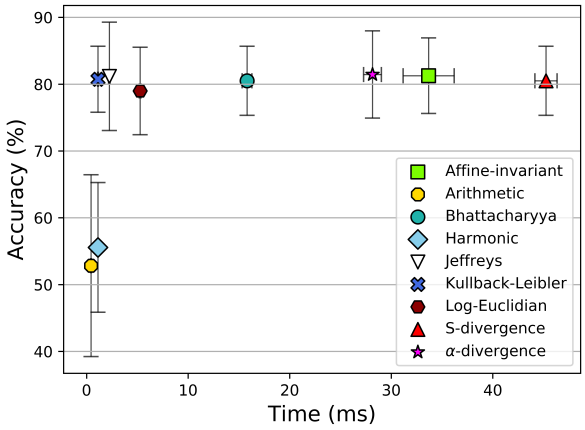
(a) Subject with highest BCI performance.



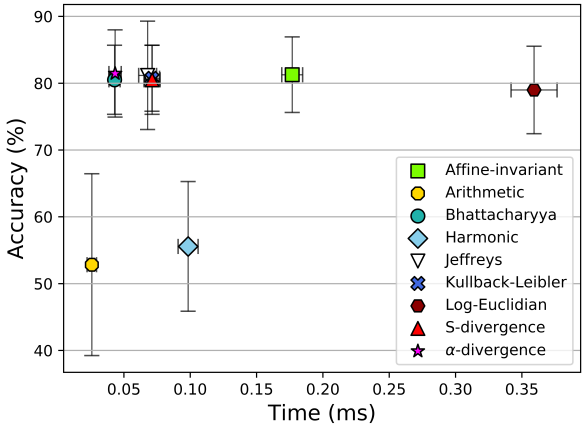
(b) Subject with lowest BCI performance.

Fig. 4 Representation of covariance matrices: each image is the covariance matrix mean $\bar{\Sigma}_k$ of the class k, for one session of the recording. The diagonal blocks show the covariance in different frequency bands, *i.e.* 13 Hz in the upper-left block, 21 Hz in the middle, and 17 Hz in the bottom-right.

the subject with the highest BCI performance. Markers represent the different means and the color map represents the distribution of covariance matrices used in the computation of the means. The determinant of the arithmetic mean is strictly larger than any other mean. The Log-Euclidean, the AIR, the S-divergence, the Jeffreys, and the Bhattacharyya means yield similar determinants, close to the mean value of determinants of individual matrices. The determinant of the harmonic mean shows a reverse phenomenon: there is a *shrinking* effect on the determinant of the individual matrices. In fact, the lack of self-duality of arithmetic mean, *i.e.* $\bar{\Sigma}_{\mathrm{E}}(\Sigma_i) \neq \bar{\Sigma}_{\mathrm{E}}^{-1}(\Sigma_i^{-1})$, can be reformulated as $\bar{\Sigma}_{\mathrm{E}} \neq \bar{\Sigma}_{\mathrm{H}}$, where $\bar{\Sigma}_{\mathrm{H}}$ is the dual of $\bar{\Sigma}_{\mathrm{E}}$ (Congedo et al. 2017b). With the generalization capability of the power means $\bar{\Sigma}_t$ with $t \in [-1, +1]$ (see Eq. (4) in Supplementary Material), the geometric mean ($t \to 0$) is right between arithmetic ($t = 1$) and harmonic ($t = -1$) means. This justifies that the shrinking factor of the harmonic mean is of the same order as the swelling factor of the



(a) Computation of average



(b) Distance and divergence computation

Fig. 5 Processing time and accuracy (mean and standard deviation across subjects and trials) for all distances and divergences. Computation of averages during offline training (a) and of distances/divergences during classification (b). For computation times, methods were run on Matlab 7.11.0.584 (R2010b), on a Linux Intel i7 machine.

arithmetic mean. The same phenomenon can be observed for the trace, with a shrinking effect stronger for the Harmonic mean than for the AIR mean. The Log-Euclidean mean and the AIR one conserve their determinant, because their determinant is the geometric mean of the determinants of individual matrices, see Supplementary Material, whereas the arithmetic mean conserves the trace, because its trace is the arithmetic mean of the traces of individual matrices. The equivalence between the S-divergence and the Bhattacharyya is also verified on this plot. Finally, the Log-Euclidean mean appears as a good trade-off for the estimation of such properties, close to the center of the distribution.

5.2.2 Riemannian distances and divergences evaluation

Using Riemannian distances significantly improves the classification performances, in comparison with the state-of-the-art CCA-based methods which have an average performance across subjects of 87.5%, 81.2% and 87.4% for Lin et al. (2006), Lin et al. (2006) and Chen et al. (2015a) respectively. These methods rely on CCA coefficients for the classification: this is a strong limitation, as they could not account for the reject class. As the reject class is not associated with any reference signal, no correlation coefficient could be computed for this class. Due to this limitation, CCA-based methods avoid to confront with the most challenging case, that is to identify when the user does not look at any stimulus. These approaches are thus not compatible with realistic BCI use case.

Various Riemannian distances and divergences achieve different classification performances. It is possible to determine the Riemannian distance or divergence that is best suited for the classification of EEG covariance matrices, and for SSVEP in particular.

The α-divergence, being a generalization of the Kullback-Leibler divergence (see Supplementary Material), offers

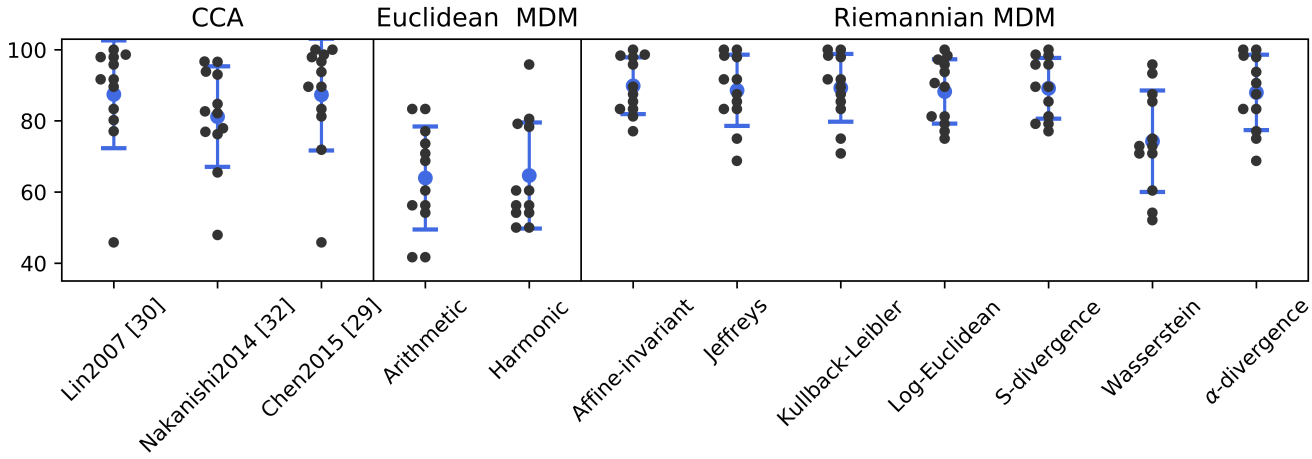


Fig. 6 Offline classification accuracy (in %) for $K = 3$ classes (no resting-state), by subject. Classification is performed with MDM using either Euclidean or Riemannian means (see Table 1). The results are benchmarked against the state-of-the-art CCA-based methods.

a particular advantage as it allows the assessment of a family of divergences by changing the value of α. It is possible to find the optimal value of α with respect to the classification accuracy for the investigated dataset. Through a cross-validation process with $-1 \leq \alpha \leq 1$ (see Fig. 9), $\alpha = 0.6$ is the optimal α for the current study. This procedure lasted 225.42 seconds and makes α-divergence the most costly method. This assessment goes through some of the divergences that are known to be particular cases of the α-divergence: the left- and right-sided Kullback-Leibler at $\alpha = -1$ and $\alpha = 1$ respectively, and the Bhattacharyya and S-divergence at $\alpha = 0$. The results of α-divergence reported on Fig 6 is obtained with the optimized value of $\alpha = 0.6$, and represents the best accuracy achievable with the family of α-divergence. The α-divergence yields the highest accuracy across distances and divergence considered in the current study. Unlike other distances and divergences, the α-divergence is not parameter-free and require an optimization step to determine the value of α. This is a limitation of the α-divergences for real BCI application, as it needs an extra processing step which is computationally demanding.

Despite the apparent good properties of the Wasserstein mean (see Fig. 7), the Wasserstein distance has the lowest accuracy among all Riemannian methods. While the Bhattacharyya distance and the S-divergence have different expressions in Table 1, they obtain similar results in terms of accuracy, and differ only in their computational time. For this reason, the Fig 6 only reports one of them.

The best trade-off regarding computation time and accuracy is provided by the Jeffreys divergence and mean, which achieves results close to the accuracy of the α-divergence or affine-invariance distance with a computation time that is an order of magnitude lower. The reported accuracy show that the Jeffreys divergence is a better estimator for classifiers than affine-invariant distance for all subjects but two.

To analyze the statistical significance of the improvement brought by Riemannian distances and divergences, we perform a paired t-test between pairs of results and we evaluate the p-values. The analysis reveals that the difference in performance of the two Euclidean distances, arithmetic and harmonic, are not statistically significant, with p-value $p > 0.05$. Also, despite the improvement in accuracy from state-of-the-art method, the test reveals that the improvement is not statistically significant ($p > 0.05$). However, it reveals a significant improvement between the Euclidean distances and the Riemannian ones, with p-values $p < 0.05$ (in the order of 10^{-5}). With the exception Wasserstein distance, the analysis shows no significant difference between the different Riemannian approaches. The most important point is that all the Riemannian approaches, with the exception of the Wasserstein distance, significantly improve the state-of-the-art ($p < 0.05$), that is MDM with any Riemannian distance or divergence outperforms the CCA-based methods for classifying SSVEP.

5.3 Online classification

Fig. 8 indicates the performances of online MDM with the different distances and divergences, using the $K = 4$ classes, including the resting-state class. Firstly, we observe that the online results are lower than the offline ones displayed on Fig. 6. There are two reasons: (i) Fig. 6 displays the results for only $K = 3$ classes, without the resting-state class; (ii) offline implementations capture more information than online ones (Chevallier

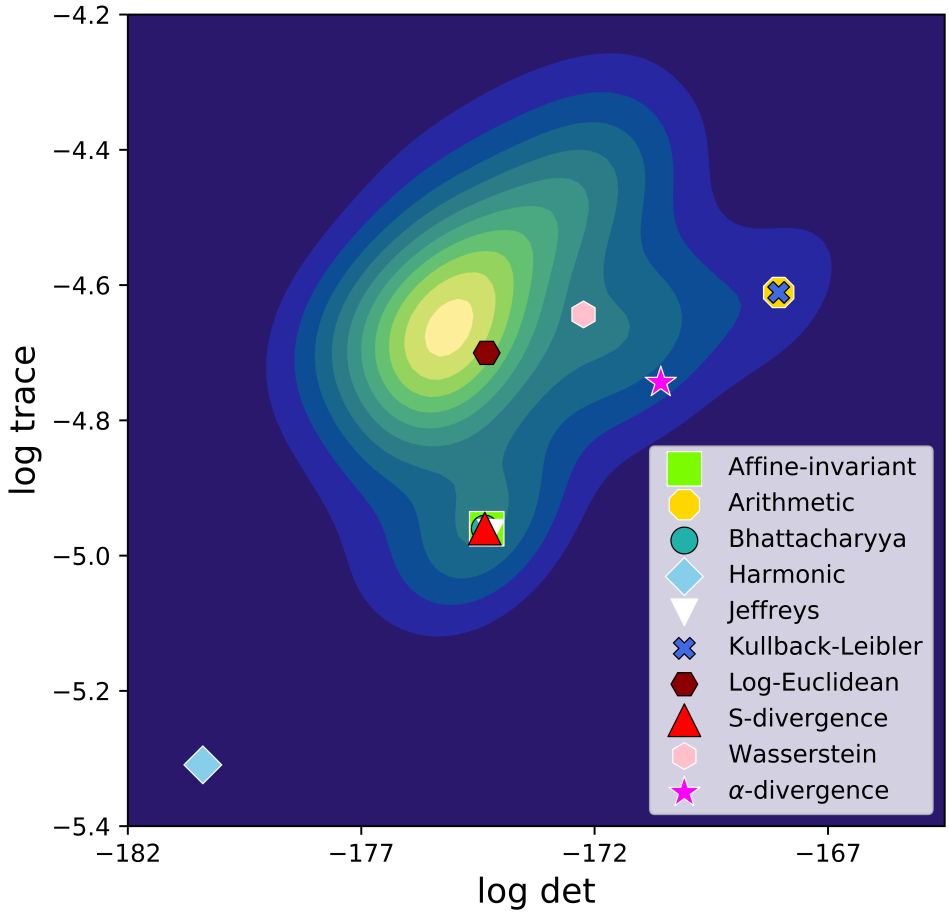


Fig. 7 The TraDe plot displays the log-trace as a function of the log-determinant of the different covariance matrices: the markers indicate the type of the mean and the color map shows the distribution of covariance matrices used in the computation of the means. Most of the means align diagonally, away from the individual matrices. A swelling effect is visible on the determinant for the arithmetic mean and a shrinking effect for the harmonic mean. A shrinking effect on the trace could also be seen for the harmonic mean and the AIR mean.

et al. 2018), leading to better accuracy. Secondly, online and asynchronous MDM with Riemannian metrics outperform those with Euclidean metrics. Wasserstein distance has the lowest accuracy among all Riemannian methods. These two observations are qualitatively similar to the observation made for the offline implementation.

The paired t-test shows that there is no statistically significant difference between MDM approaches with Riemannian metrics, apart from the Wasserstein that obtains significantly lower performances. It is to be noted that CCA-based methods, used for comparison in offline classification, are not adequate for online and asynchronous applications as they cannot separate SSVEP classes from the resting class. This limitation disqualifies them for any real implementation of BCI.

With these experimental results on real EEG, the benefit of dedicated Riemannian tools for processing covariance matrices is clearly demonstrated. As seen on offline experiments the Jeffreys divergence is a good choice, as it is a good trade-off between the accuracy and the computation time.

6 Conclusion

The objective of this study is to investigate how fast and accurate systems that could be applied in a loosely controlled environment, instead of carefully tuned systems that are designed in strongly constrained environ-

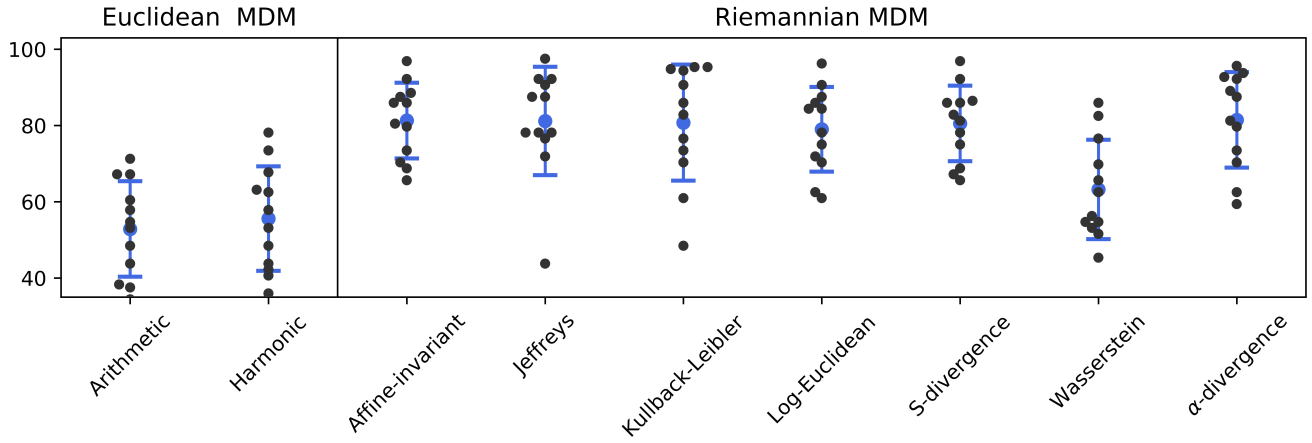


Fig. 8 Online classification accuracy (in %) for $K = 4$ classes, by subject. Classification is performed with MDM using either Euclidean or Riemannian means (see Table 1).

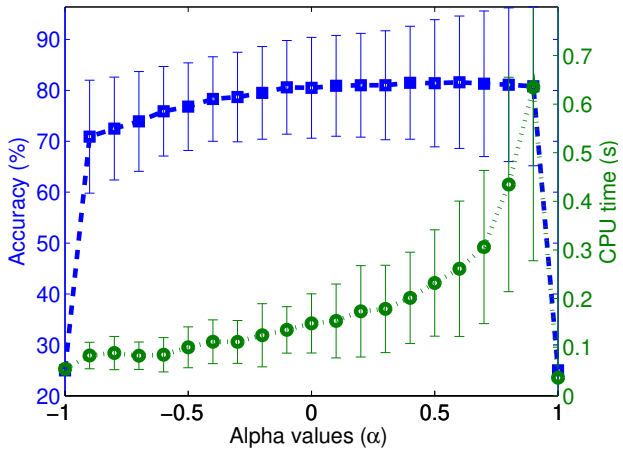


Fig. 9 Classification accuracy and CPU time (mean and standard deviation) obtained for the log-determinant α-divergence, with $-1 \leqslant \alpha \leqslant 1$. The values are averaged across all sessions and subjects.

ments. Riemannian approaches have been successfully applied on EEG signals for brain-computer interfaces, demonstrating their robustness to noise and to the intrinsic variation of EEG signals. A direct approach, like Minimum Distance to Riemannian Mean, yields comparable results with state-of-the-art without requiring a complex choice of parameters or a lengthy optimization phase. Covariance matrices belong to a curved space and, expressed in a Riemannian framework, several distances are defined in the literature. Most of these distances embed desirable invariances, avoiding costly calibration of spatial filters for each session and yielding estimation that are robust to artifacts and outliers. Different distances and divergences, along with their implementation are reviewed and investigated, including the Log-Euclidean distance, the affine-invariant Riemannian distance, the Kullback-Leibler divergence, the Jeffreys divergence, the S-divergence, the α-divergence and the Wasserstein distance.

This study shows that, for SSVEP classification, Riemannian approaches outperform previous state-of-the-art methods based on CCA. Riemannian approaches also address the issue of resting class, where the user is not interacting with the system, which is a challenging task in machine learning. A thorough analysis of computational cost and of numerical stability is proposed in the experimental section. The best accuracy is obtained with the α-divergence, after finding the optimal value of α. The Jeffreys divergence offers the best accuracy for the lowest computational cost. At last, to ensure the reproducibility of the results and to propose a framework for evaluation of new approaches to the communities of brain-computer interfaces and information geometry, all the data, including the algorithms to process them are available under an open source license.

The presented Riemannian framework can be extended to MI or ERP paradigms (Congedo et al. 2017a). Future works on this topic could consist of investigating whether distances and divergences behave similarly with different paradigms, in terms of classification accuracy.

On a methodological level, these first results call for an extension of covariance-based features to other VEP paradigms: phase-based SSVEP and code–modulated VEP (c-VEP). These paradigms achieve very high information transfer rates and could greatly benefit from the accuracy improvement provided by Riemannian geometry.

Finally, another opportunity ahead is to evaluate the possibility of making transfer learning between frequencies for a given subject. This approach would allow to calibrate the system on single frequency and generalize the classifier's training to a large number of frequen-

cies. This is particularly useful as it could help reduce the calibration time and allow subjects to keep their concentration for the real task.

Acknowledgements This work has been done without any support of the ANR or the ERC.

7 Compliance with Ethical Standards

At the time this study was carried out, Emmanuel Kalunga was an employee of VASTech company, Stellenbosch, South Africa; and Quentin Barthélemy was an employee of Mensia Technologies S.A., Paris, France. Sylvain Chevallier and Eric Monacelli declare that they have no conflict of interest. This study was conducted by the authors without any specific funding or grant.

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

8 Information Sharing Statement

All the data used in this paper could be freely downloaded from Chevallier (2017) or using the MOABB framework Barachant et al. (2019). The code for the experiments of this article is written in Matlab: the code is available on Kalunga (2018) for offline analysis, and on Kalunga (2015) for online analysis. A code written in Python is available on Chevallier (2020) for offline analysis, and on Bertrand-Lalo (2020) for online analysis.

References

Afsari B (2011) Riemannian L^p center of mass: existence, uniqueness, and convexity. Proc Amer Math Soc 139(2):655–673

Agueh M, Carlier G (2011) Barycenters in the Wasserstein space. SIAM J Math Anal 43(2):904–924

Álvarez-Esteban P, del Barrio E, Cuesta-Albertos J, Matrán C (2016) A fixed-point approach to barycenters in Wasserstein space. Journal of Mathematical Analysis and Applications 441:744–762

Ando T, Li CK, Mathias R (2004) Geometric means. Linear Algebra Appl 385:305–334

Arsigny V, Fillard P, Pennec X, Ayache N (2007) Geometric means in a novel vector space structure on symmetric positive-definite matrices. SIAM J Matrix Anal Appl 29(1):328–347

Barachant A, Bonnet S, Congedo M, Jutten C (2012) Multiclass brain–computer interface classification by Riemannian geometry. IEEE Trans Biomed Eng 59(4):920–928

Barachant A, Andreev A, Congedo M (2013) The Riemannian potato: an automatic and adaptive artifact detection method for online experiments using Riemannian geometry. In: Proceedings of TOBI Workshop IV, pp 19–20

Barachant A, Jayaram V, Chevallier S, Rodrigues P (2019) MOABB framework. URL https://github.com/NeuroTechX/moabb

Bertrand-Lalo R (2020) Online SSVEP classification in Python with TimeFlux. URL https://github.com/bertrandlalo/timeflux_ssvep

Bhatia R (2009) Positive definite matrices, vol 16. Princeton university press

Bhattacharyya A (1943) On a measure of divergence between two statistical populations defined by their probability distribution. Bull Calcutta Math Soc 35:99–109

Cartan E (1929) Groupes simples clos et ouverts et géométrie riemannienne. Journal de mathématiques pures et appliquées pp 1–34

Chebbi Z, Moakher M (2012) Means of Hermitian positive-definite matrices based on the log-determinant α-divergence function. Linear Algebra Appl 436(7):1872–1889

Chen X, Chen Z, Gao S, Gao X (2014) A high-ITR SSVEP-based BCI speller. Brain-Computer Interfaces 1(3-4):181–191

Chen X, Wang Y, Gao S, Jung T, Gao S (2015a) Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain-computer interface. J Neural Eng

Chen X, Wang Y, Nakanishi M, Gao X, Jung T, Gao S (2015b) High-speed spelling with a noninvasive brain–computer interface. Proceedings of the National Academy of Sciences 112(44):E6058–E6067

Cherian A, Sra S, Banerjee A, Papanikolopoulos N (2011) Efficient similarity search for covariance matrices via the Jensen-Bregman LogDet divergence. In: Int Conf Computer Vision, IEEE, pp 2399–2406

Chevallier S (2017) SSVEP data. URL https://github.com/sylvchev/dataset-ssvep-exoskeleton

Chevallier S (2020) Offline SSVEP classification in Python. URL https://github.com/alexandrebarachant/pyRiemann/tree/master/examples/SSVEP

Chevallier S, Kalunga E, Barthélemy Q, Yger F (2018) Brain Computer Interfaces Handbook: Technological and Theoretical Advances, CRC Press, chap 19 - Riemannian classification for SSVEP based BCI : offline versus online implementations, pp 371–396

Cichocki A, Amari S (2010) Families of alpha-beta-and gamma-divergences: Flexible and robust measures of similarities. Entropy 12(6):1532–1568

Congedo M, Barachant A, Bhatia R (2017a) Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review. Brain-Computer Interfaces 4:1–20

Congedo M, Barachant A, Koopaei E (2017b) Fixed point algorithms for estimating power means of positive definite matrices. IEEE Trans Signal Process 65:2211–2220

Dhillon IS, Tropp JA (2007) Matrix nearness problems with Bregman divergences. SIAM J Matrix Anal Appl 29(4):1120–1146

Fletcher P, Joshi S (2004) Principal Geodesic Analysis on Symmetric Spaces: Statistics of Diffusion Tensors. In: Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis, LNCS, vol 3117, Springer, pp 87–98

Gergondet P, Kheddar A (2015) SSVEP stimuli design for object-centric BCI. Brain-Computer Interfaces 2(1):11–

28
Hastie T, Tibshirani R, Friedman J (2009) The Elements of Statistical Learning. Springer Series in Statistics, Springer New York
Herrmann CS (2001) Human EEG responses to 1-100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. Experimental Brain Research 137(3-4):346–353
Hosni SM, Shedeed HA, Mabrouk MS, Tolba MF (2018) EEG-EOG based virtual keyboard: Toward hybrid brain computer interface. Neuroinformatics
Jeffreys H (1946) An invariant form for the prior probability in estimation problems. Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences 186:453–461
Johannes MG, Pfurtscheller G, Flyvbjerg H (1999) Designing optimal spatial filters for single-trial EEG classification in a movement task. Clin Neurophysiol 110(5):787–798
Kalunga E, Djouani K, Hamam Y, Chevallier S, Monacelli E (2013) SSVEP enhancement based on Canonical Correlation Analysis to improve BCI performances. In: AfriCon, 2013, pp 1–5
Kalunga E, Chevallier S, Barthélemy Q, Djouani K, Hamam Y, Monacelli E (2015) From Euclidean to Riemannian Means: Information Geometry for SSVEP Classification. In: Geometric Science of Information, 9389, Springer International Publishing, pp 595–604
Kalunga EK (2015) Online SSVEP classification in Matlab. URL https://github.com/emmanuelkalunga/Online-SSVEP
Kalunga EK (2018) Offline SSVEP classification in Matlab. URL https://github.com/emmanuelkalunga/Offline-Riemannian-SSVEP
Kalunga EK, Chevallier S, Rabreau O, Monacelli E (2014) Hybrid interface: Integrating BCI in multimodal human-machine interfaces. In: 2014 IEEE/ASME Int Conf Advanced Intelligent Mechatronics (AIM), pp 530–535
Kalunga EK, Chevallier S, Barthélemy Q, Djouani K, Monacelli E, Hamam Y (2016) Online SSVEP-based BCI using Riemannian geometry. Neurocomputing 191:55–68
Lim Y, Pálfia M (2012) Matrix power means and the Karcher mean. J Funct Anal 262(4):1498–1514
Lin Z, Zhang C, Wu W, Gao X (2006) Frequency Recognition Based on Canonical Correlation Analysis for SSVEP-Based BCIs. IEEE Trans Biomed Eng 53(12):2610–2614
Lin Z, Zhang C, Wu W, Gao X (2007) Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. IEEE Trans Biomed Eng 53(12):2610–2614
Lotte L, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A, Yger F (2018) A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update. J Neural Eng 15:031005
Martin H, Chevallier S, Monacelli E (2012) Fast calibration of hand movement-based interface for arm exoskeleton control. In: European Symposium on Artificial Neural Networks (ESANN), pp 573–578
Massart EM, Chevallier S (2017) Inductive means and sequences applied to online classification of eeg. In: International Conference on Geometric Science of Information, Springer, pp 763–770
McFarland D, Sarnacki W, Wolpaw J (2003) Brain–computer interface (bci) operation: optimizing information transfer rates. Biological psychology 63(3):237–251
McFarland D, Daly J, Boulay C, Parvaz M (2017) Therapeutic applications of BCI technologies. Brain-Computer Interfaces 4:1–2
Meinel A, Castaño-Candamil S, Blankertz B, Lotte F, Tangermann M (2019) Characterizing regularization techniques for spatial filter optimization in oscillatory EEG regression problems. Neuroinformatics 17(2):235–251
Moakher M (2005) A differential geometric approach to the geometric mean of symmetric positive-definite matrices. SIAM J Matrix Anal Appl 26(3):735–747
Moakher M, Batchelor PG (2006) Symmetric Positive-Definite Matrices: From Geometry to Applications and Visualization, Springer Berlin Heidelberg, Berlin, Heidelberg, chap 17, pp 285–298
Nakanishi M, Wang Y, Wang Y, Mitsukura Y, Jung T (2014) High-speed brain speller using steady-state visual evoked potentials. Int J Neural Syst 24(06):1450019
Nam CS, Choi I, Wadeson A, Whang M (2018) Brain-computer interfaces handbook: technological and theoretical advances, CRC Press, chap Brain-Computer Interface An Emerging Interaction Technology
Niedermeyer E, Silva FHLd (2005) Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Lippincott Williams & Wilkins
Nielsen F, Bhatia R (2012) Matrix Information Geometry. Springer Publishing Company, Incorporated
Nielsen F, Nock R (2009) Sided and symmetrized Bregman centroids. IEEE Trans Inf Theory 55(6):2882–2904
Nielsen F, Nock R, Amari S (2014) On clustering histograms with k-means by using mixed α-divergences. Entropy 16(6):3273–3301
Pennec X, Fillard P, Ayache N (2006) A Riemannian Framework for Tensor Computing. Int J Comput Vis 66(1):41–66
Quitadamo LR, Marciani MG, Cardarilli GC, Bianchi L (2008) Describing different brain computer interface systems through a unique model: A UML implementation. Neuroinformatics 6(2):81–96
Rivet B, Souloumiac A, Attina V, Gibert G (2009) xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain-Computer Interface. IEEE Trans Biomed Eng 56(8):2035–2043
Sra S (2016) Positive definite matrices and the S-divergence. Proceedings of the American Mathematical Society 144(7):2787–2797
Tomioka R, Aihara K, Müller KR (2007) Logistic regression for single trial EEG classification. In: Advances in neural information processing systems (NIPS), vol 19, pp 1377–1384
Verschore H, Kindermans PJ, Verstraeten D, Schrauwen B (2012) Dynamic stopping improves the speed and accuracy of a P300 speller. In: Artificial Neural Networks and Machine Learning–ICANN, Springer, pp 661–668
Vialatte FB, Maurice M, Dauwels J, Cichocki A (2010) Steady-state visually evoked potentials: focus on essential paradigms and future perspectives. Prog Neurobiol 90(4):418–438
Villani C (2008) Optimal transport: old and new, vol 338. Springer Science & Business Media
Volosyak I, Valbuena D, Luth T, Malechka T, Graser A (2011) BCI demographics II: How many (and what kinds of) people can use a high-frequency SSVEP BCI? IEEE Trans Neural Syst Rehabil Eng 19(3):232–239
Wang S, James CJ (2006) Enhancing Evoked Responses for BCI Through Advanced ICA Techniques. In: Advances in Medical, Signal and Information Processing (MEDSIP), pp 1–4

Wang Y, Zhang Z, Gao X, Gao S (2004) Lead selection for SSVEP-based brain-computer interface. In: Int Conf IEEE Engineering in Medicine and Biology Society, vol 2, pp 4507–4510

Yang B, Zhang T, Zhang Y, Liu W, Wang J, Duan K (2017) Removal of electrooculogram artifacts from electroencephalogram using canonical correlation analysis with ensemble empirical mode decomposition. Cognitive Computation 9(5):626–633

Yang Y, Bloch I, Chevallier S, Wiart J (2016) Subject-specific channel selection using time information for motor imagery brain–computer interfaces. Cognitive Computation 8(3):505–518

Yger F, Berar M, Lotte F (2016) Riemannian approaches in brain-computer interfaces: a review. IEEE Trans Neural Syst Rehabil Eng 25(10):1753–1762

Zhang Y, Zhou G, Jin J, Wang X, Cichocki A (2015) Ssvep recognition using common feature analysis in brain–computer interface. J Neurosci Methods 244:8–15

Zhu D, Bieger J, Molina GG, Aarts RM (2010) A survey of stimulation methods used in SSVEP-based BCIs. Intell Neuroscience 2010:1–12