MDPI

*Article*

# From Euclidean to Riemannian Centers of Classes: Information Geometry for SSVEP Classification

**Emmanuel Kalunga [1,2]*, Sylvain Chevallier [2], Quentin Barthélemy [3], Karim Djouani [1], Yskandar Hamamv[1] and Eric Monacelli [2,]**

[1] Department of Electrical Engineering/F'SATI Tshwane University of Technology, Pretoria 0001, South Africa
[2] Laboratoire d'Ingénierie des Systèmes de Versailles, Université de Versailles Saint-Quentin, 78140 Velizy, France
[2] Mensia Technologies, ICM, Hôpital de la Pitié-Salpêtrière, 75013 Paris, France
* Correspondence: emmanuelkalunga.k@gmail.com

Academic Editor: name
Version June 23, 2016 submitted to Entropy; Typeset by LaTeX using class file mdpi.cls

**Abstract:** Brain Computer Interfaces (BCI) based on electroencephalography (EEG) rely on multichannel brain signal processing. Most of the state-of-the-art approaches deal with covariance matrices, and indeed Riemannian geometry has provided a substantial framework for developing new algorithms. Most notably, a straightforward algorithm such as Minimum Distance to Mean yields competitive results when applied with a Riemannian distance or divergence. This applicative contribution aims at assessing the impact of several distances/divergences on real EEG dataset, as the invariances embedded in those distances/divergences have an influence on the classification accuracy. Riemannian centers of classes compare favorably with respect to Euclidean ones both in term of quality of results and of computational load. Riemannian distances cope with signal variabilities and reduce the adverse effect of artifacts in EEG signal.

**Keywords:** Riemannian geometry; Divergences; Centroids; covariance matrices; EEG; SSVEP; BCI

## 1. Introduction

Brain-Computer Interfaces (BCI) allow interaction with a computer or a machine without relying on the user's motor capabilities. In rehabilitation and assistive technologies, BCI offer promising solutions to compensate for physical disabilities. To record brain signals in BCI systems, the most common choice is to rely on electroencephalography (EEG) [1], as the recording systems are smaller and less expansive than other brain imaging technologies (such as MEG or fMRI). BCI systems rely on different brain signals, such as event-related desynchronization or evoked potentials. The former is observed in the premotor cortex when the subject imagines moving some part of his own body (also known as Motor Imagery paradigm) and the latter qualifies the brain response to a specific sensory stimulation, usually visual or auditory. This contribution concentrates on Steady-State Visually Evoked Potentials (SSVEP), which are potentials emerging when a subject concentrates his attention on a stimulus blinking with a given frequency. Shortly after the user concentrates on this stimulus, brain waves in visual cortex could be observed with matching frequencies. To date, BCI still faces challenges and a major limitation is the EEG poor spatial resolution. This limitation is due to the volume conductance effect [1], as the skull bones act as a non-linear low pass filter, mixing the brain source signals and thus reducing the signal-to-noise ratio. Consequently, spatial filtering methods have been developed or adopted, such as xDAWN [2] or Independent Component Analysis (ICA) [3]. Spatial filters are very efficient on clean datasets obtained from strongly constrained environment but they are sensitive to artifacts and outliers [4]. Instead, approaches based on covariance matrices, such

as CSP [5] and Canonical Correlation Analysis [6] offer a much more robust framework. Covariance matrices being Symmetric and Positive-Definite (SPD), tools offered by Riemannian geometry have been recently explored with promising results.

A classification technique referred to as *minimum distance to Riemannian mean* (MDRM) has been recently introduced to EEG classification [7]. It entirely relies on covariance matrices and the fact that they belong the manifold of SPD matrices. New EEG trials are assigned to the class whose average covariance matrix is the closest to the trial covariance matrix according to the affine-invariant Riemannian metric [8]. It is a simple, yet robust classification scheme outperforming complex and highly parametrized state-of-the-art classifiers. The limitations of using Euclidean metrics in the computation of distances between SPD matrices and their means have been demonstrated [9]. Using information geometry, a number of Riemannian distances and divergences have been developed and appropriately used on SPD matrices [9,10]. The present work applies some of these distances to SSVEP data, providing a practical analysis and a comparison with Euclidean metrics.

Moreover, most applications of Riemannian geometry to BCI are thus far focusing only on Motor Imagery (MI) paradigm. Riemannian BCI is well suited for MI experiment as the spatial information linked with synchronization are directly embedded in covariance matrices obtained from multichannel recordings. However, for BCI that rely on evoked potential such as SSVEP or event-related potential, as P300, frequency or temporal information are needed. In [11], the authors propose a rearrangement of the covariance matrices that embed the timing or the frequency information, thus allowing a direct application of the Riemannian framework. This contribution relies on this rearrangement to apply MDRM on covariance matrices of SSVEP signals. The signals are recorded in an application of assistive robotics where an SSVEP-based BCI is used in tandem with a 3D touchless interface based on IR-sensors as a multimodal system to control an arm exoskeleton [12].

The paper is organized as follows: Section 2 describes the framework for the classification of covariance matrices. It first presents the basic principles of classification 2.1. It then describes how means of covariance matrices are computed 2.2. Since the notion of distance is key in the computation of the mean, a number of useful distances and divergences are subsequently presented **??**. Section 2 ends with the presentation of the MDM algorithm which is the classification method used in this work 2.3. In Section 3, the classification results obtained on real EEG dataset are presented and discussed. Section 4 concludes this paper.

## 2. Classification of covariance matrices for SSVEP

### *2.1. fundamentals of classification*

Given labelled samples $x_i$ drown from two populations (positive and negative), a simple classification algorithm consists in assigning a previously unseen sample to the class with closer mean[ref: fig]. This implies a computation of means of classes and a measure of distances from the means. Assuming that the samples are embedded into a dot product space (i.e. with Euclidean geometry), the means can be computed as:

$$c_+ = \frac{1}{m_+} \sum_{\{i|y_i=+1\}} x_i, \tag{1}$$

$$c_- = \frac{1}{m_-} \sum_{\{i|y_i=-1\}} x_i, \tag{2}$$

where $y_i$ is the label of sample $x_i$; $m_+$ and $m_-$ the number of positive and negative samples respectively. An unseen sample $x$ is unsigned to the class whose mean is the closest. This simple geometric classification framework is the founding principle of more complex algorithms such as supporting vector machines. It can be formulated in terms of the dot product $\langle \cdot, \cdot \rangle$:

If $c := (c_+ + c_-)/2$ is the point lying halfway between $c_+$ and $c_-$, and $w := c_+ - c_-$ the vector connecting $c_+$ to $c_-$, the class of the unseen sample $x$ is determined by checking whether the vector $x - c$ connecting $c$ to $x$ makes an angle $\alpha < \pi/2$ with $w$ [ref: fig]. This is expressed as:

$$\begin{aligned} y &= \operatorname{sgn} \langle (x - c), w \rangle \\ &= \operatorname{sgn} \langle (x - (c_+ + c_-)/2), (c_+ - c_-) \rangle \\ &= \operatorname{sgn}(\langle x, c_+ \rangle - \langle x, c_- \rangle + b) \end{aligned} \tag{3}$$

The offset $b$ vanishes if class means are equidistant to the origin [13]. Inserting (1) and (2) in (3) yields

$$y = \operatorname{sgn} \left( \frac{1}{m_+} \sum_{\{i|y_i=+1\}} \langle x, x_i \rangle - \frac{1}{m_-} \sum_{\{i|y_i=-1\}} \langle x, x_i \rangle + b \right) \tag{4}$$

Classifier (4) can be generalized as

$$y = \operatorname{sgn} \left( \sum_{i=1}^{m} y_i \alpha_i d(x, x_i) + b \right) \tag{5}$$

where $\alpha_i$ is the weight of sample $x_i$ and $d(\cdot, \cdot)$ a distance, a divergence of a kernel. In the case of two classes in the dot product space where all samples have the same weight, $y_i \in \{+1, -1\}$, $\alpha = 1/m$ and $d(\cdot, \cdot) = \langle \cdot, \cdot \rangle$.

Expression (5) corresponds to the decision function used in hyperplane classifiers [13]. This shows that even more complex classifiers rely on the calculations of class means (or centers) and their distances to individual samples. This being shown, in this article we focus on the simple classification approach of assigning a previously unseen sample to the class with closest mean.

In machine learning algorithms, samples are represented by their features which are determined through a feature extraction and selection process. In this work, samples are represented by their covariance matrices. Therefore means of classes and distances to mean will be means of covariance matrices and distance between them.

### 2.2. Means of Covariance matrices

Consider a multivariate variable $X \in \mathbb{R}^{C \times N}$ where $C$ is the number of variables and $N$ the number of samples, with $C > N$, the covariance matrix of $X$ which can be estimated as

$$\hat{\Sigma} = \frac{1}{N} X X^\intercal \tag{6}$$

is symmetric positive definite (SPD). The properties of SPD matrices constrain them to a convex cone:

(i) Symmetry: $\Sigma = \Sigma^\intercal$
(ii) Positive definiteness: $x^\intercal \Sigma x > 0, \forall x \in \mathbb{R}^C \backslash 0$
(iii) Strict positivity of diagonal element: $\Sigma_{ij} > 0 | i = j, \forall i, j \in \{1, \ldots, C\}$ i.e. positive variance.
(iv) Cauchy-Schwarz inequalities: $|\Sigma_{ij}| \leq (\Sigma_{ii} \Sigma_{jj})^{1/2}, \forall i, j \in \{1, \ldots, C\}$

The mean of SPD matrices can be computed as a center of mass modeled on Euclidean geometry: given a set of covariance matrices $\{\Sigma_i\}_{i=1,\ldots,I}$, the center of mass $\bar{\Sigma}$ of the set, is a covariance matrix that minimizes the sum of the squared distances to matrices $\Sigma_i$:

$$\bar{\Sigma} = \mu(\Sigma_1, \ldots, \Sigma_I) = \arg\min_{\Sigma \in \mathcal{M}_C} \sum_{i=1}^{I} d^2(\Sigma_i, \Sigma) \,, \tag{7}$$

where $d(\cdot, \cdot)$ is a measure of distance between two matrices. practically, $d(\cdot, \cdot)$ can either be a distance or a divergence. In the literature, this mean is at times designated as the *Frechet mean*, *Cartan mean*, or

*Karcher mean* [1] [15]. Cartan [16] had shown that a unique solution to (7) exists if all $\Sigma_i$ lie in a convex ball [section 16 of 16]. This applies also to closed convex cones.

Depending on the divergence or distance used, several means can be defined from (7). Those considered in this study are presented in the next lines and summarized in Table 1

#### 2.2.1. Distance and divergence

Divergences and distances are measures of dissimilarity between two points in a space. Here the Riemannian space will be considered. A distance function $d : \mathcal{M} \times \mathcal{M} \to \mathbb{R}^+$ has the following properties for all $\Sigma_1, \Sigma_2, \Sigma_3 \in \mathcal{M}$:

| | | |
|---|---|---|
| (i) | $d(\Sigma_1, \Sigma_2) \geq 0$ | non-negativity |
| (ii) | $d(\Sigma_1, \Sigma_2) = 0 \text{ iff } \Sigma_1 = \Sigma_2$ | identity |
| (iii) | $d(\Sigma_1, \Sigma_2) = d(\Sigma_2, \Sigma_1)$ | symmetry |
| (iv) | $d(\Sigma_1, \Sigma_3) \leq d(\Sigma_1, \Sigma_2) + d(\Sigma_2, \Sigma_3)$ | triangular inequality |

Divergences are very similar to distances with the difference that properties (iii) and (iv) do not have to be satisfied. In the context of Covariance matrices, divergences and distances should both induce a Riemannian metric on the manifold of SPD matrices.

#### 2.2.2. Euclidean distance

The Euclidean distance between two matrices is represented by the *Frobenius norm* of their difference:

$$d_{\mathrm{E}}(\Sigma_1, \Sigma_2) = \|\Sigma_1 - \Sigma_2\|_F \tag{8}$$

In (7), this yields the arithmetic mean:

$$\bar{\Sigma}_{\mathrm{E}} = \frac{1}{I} \sum_{i=1}^{I} \Sigma_i \tag{9}$$

The arithmetic mean is drown from a family of power means ($\Sigma_{t|t=1}$)

$$\Sigma_t = \left( \frac{1}{I} \sum_{i=1}^{I} \Sigma_i^t \right)^{\frac{1}{t}}, t \in [-1, +1]. \tag{10}$$

From the same family can be drown the *geometric mean* ($\Sigma_{t|t\to 0}$) and the *harmonic mean* ($\Sigma_{t|t=-1}$). We consider the arithmetic mean $\bar{\Sigma}_{\mathrm{E}}$, as a baseline. This averaging of covariance is usually not adequate in the space of SPD matrices for 2 main reasons. First the Euclidean distance and averaging do not grantee invariance under inversion: a matrix and its inverse are supposed to be at the same distance from the identity matrix. Secondly the Euclidean averaging of covariance SPD leads to a *swelling effect*: the determinant of the arithmetic mean of SPD matrices can be larger than the determinant of its individual components. And since the determinant of a covariance matrice is a direct measure of the dispersion of the multivariate variable, the swelling effect introduces a large distortion of the data dispersion [9]. For these reasons, other means that adapt to the geometry of convex cone of SPD matrices are used.

#### 2.2.3. Affine Invariant distance

The affine invariant distance between two points is defined by the length of the curve connecting these point on the Riemannian manifold. A differential manifold is a topological curved space that is locally similar to the Euclidean space and is differentiable globally. The convex cone of SPD matrices

---

[1] This appellations have been recently criticized by Karcher himself [14]

is a manifold that can be endowed with a Riemannian metric; such manifolds are called Riemannian manifold. Let $\mathcal{M}$ be a Riemannian manifold, and $T_\Sigma\mathcal{M}$ its tangent space defined on point $\Sigma$. A Riemannian metric $d$ is a family of inner product defined on the tangent spaces defined on each point $\Sigma$ of the Manifold. This inner product varies smoothly from point to point on the manifold,

$$d_\Sigma : T_\Sigma\mathcal{M} \times T_\Sigma\mathcal{M} \to \mathbb{R}$$

$d$ is a function that assigns, for each point $\Sigma \in \mathcal{M}$, an inner product in the tangent space $T_\Sigma\mathcal{M}$ . The Riemannian metric allows us to compute the length of vectors or distance between two point on the tangent space, and through appropriate mapping [17], the length of the corresponding *geodesic* (i.e. the shortest curve connecting two point) on the manifold $\mathcal{M}$. The *affine invariant distance* is the distance between two point of a Riemannian manifold endowed with an invariant Riemannian metric

$$\begin{aligned} g_\Sigma(\Theta_1, \Theta_2) &= \langle \Theta_1 | \Theta_2 \rangle_\Sigma \\ &= \left\langle \Sigma^{-\frac{1}{2}}\Theta_1\Sigma^{-\frac{1}{2}} | \Sigma^{-\frac{1}{2}}\Theta_2\Sigma^{-\frac{1}{2}} \right\rangle_{\mathbf{I}} \\ &= \text{trace}\left(\Sigma^{-\frac{1}{2}}\Theta_1\Sigma^{-1}\Theta_2\Sigma^{-\frac{1}{2}}\right) \end{aligned} \tag{11}$$

The inner product of the tangent vectors $\Theta_1$ and $\Theta_2$ at $\Sigma$ is invariant by the action of $\Sigma^{-\frac{1}{2}}$ transformation. The affine-invariant Riemannian distance is defined as:

$$d_{\text{AI}}(\Sigma_1, \Sigma_2) = \|\log(\Sigma_1^{-1}\Sigma_2)\|_F = \left[\sum_{c=1}^{C} \log^2 \lambda_c\right]^{1/2}, \tag{12}$$

where $\lambda_c$, $c = 1, \ldots, C$, are the eigenvalues of $\Sigma_1^{-1}\Sigma_2$.

Inserting (12) in (7) yields the mean $\bar{\Sigma}_{\text{AI}}$ associated to the affine-invariant metric. It is the solution to

$$\sum_{i=1}^{I} \log(\bar{\Sigma}_{\text{AI}}^{-1/2}\Sigma_i\bar{\Sigma}_{\text{AI}}^{-1/2}) = 0 \tag{13}$$

It has no close form solution and can be solved iteratively through a gradient descent algorithm [**?** ].

These distance and mean are invariant to affine transformations. Some of these invariances are particularly important to preserve the geometric topology of the Riemannian manifold of SDP matrices: Let $f$ be an affine invariant function defined on $\mathcal{M}$ (e.g. distance or mean),

(i) *Invariance under congruent transformation*

$$f(\Sigma_1, \Sigma_2) = f(\Sigma\Sigma_1\Sigma^\intercal, \Sigma\Sigma_2\Sigma^\intercal) \tag{14}$$

(ii) *Invariance under inversion*

$$f(\Sigma, \mathbf{I}) = f(\Sigma^{-1}, \mathbf{I}) \tag{15}$$

where $\mathbf{I}$ is the identity matrix. This implies that

$$f(\Sigma_1, \Sigma_2) = f(\Sigma_1^{-1}, \Sigma_2^{-1}) \tag{16}$$

Another interesting property of the affine-invariant metric is its invariance to left- and right-multiplication by a positive matrix.

$$f(\Sigma_1, \Sigma_2) = f(\Sigma\Sigma_1, \Sigma\Sigma_2) = f(\Sigma_1\Sigma, \Sigma_2\Sigma) \tag{17}$$

### 2.2.4. Log-Euclidean

The Log-Euclidean is another distance that takes into consideration the topology of Riemannian manifolds. It was introduced by Arsigny et al. to alleviate the complexity involved in the computation of the affine-invariant distance and its related mean [9]. The mean associated to the Log-Euclidean distance corresponds to an arithmetic mean in the domain of matrix algorithm. The distance between two SPD matrices is expressed as

$$d_{\mathrm{LE}}(\Sigma_1, \Sigma_2) = \|\log(\Sigma_1) - \log(\Sigma_2)\|_F \tag{18}$$

Its associated mean is defined explicitly:

$$\bar{\Sigma}_{\mathrm{LE}} = \exp\left(\sum_{i=1}^{I} \log(\Sigma_i)\right) \tag{19}$$

Unlike the affine-invariant mean, the Log-Euclidean mean as a closed form expression which gives it a large computational advantage. Moreover, the obtained mean is, to a large extent, similar to the affine-invariant mean:

(i) They have the same determinants which correspond to the geometric mean of the determinants of their building matrices:

$$|\bar{\Sigma}_{\mathrm{LE}}| = |\bar{\Sigma}_{\mathrm{AI}}| = \prod_{i=1}^{I}(|\Sigma_i|)^{1/I} = \exp(\frac{1}{I}\sum_{i=1}^{I}\log(|\Sigma_i|))$$

(ii) They are often equal in value, if not, $\mathrm{trace}(\bar{\Sigma}_{\mathrm{LE}}) > \mathrm{trace}(\bar{\Sigma}_{\mathrm{AI}})$

(iii) Log-Euclidean mean has properties close to affine-invariance (i.e. similarity-invariance instead of congruent-invariance).

### 2.2.5. Bregman divergences

Divergences have been considered for the computation of mean in applications of clustering and classification of SPD matrices due to the fact that they induce a Riemannian metric biven by (11). Consider a strictly convex and differentiable function $f : \mathbb{R} \to \mathbb{R}$; then $f(x) \geq f(y) + f'(y)(x-y)$ and $f(x) = f(y) + f'(y)(x-y) \Leftrightarrow x = y$ for all $x, y \in \mathbb{R}$. The Bregman divergence, [introducde by Bregman in 18] is the difference between the left and right sides of the inequality:

$$D_{\mathrm{f}}(x, y) = f(x) - f(y) - f'(y)(x-y). \tag{20}$$

$f$ is called a *seed function*. It is shown that $D_{\mathrm{f}}$ verifies the non-negativity and the identity properties. When the seed function is quadratic, it can also be symmetric. There is another set of properties that $D_{\mathrm{f}}$ verifies, they are reported in [18]. Geometrically, the Bregman divergence can be seen as the measure of the difference between $f(x)$ and its representation on the plane tangent to $f$ at $y$ as illustrated in figure 1.

The scalar divergence can be directly adapted to SPD matrices as:

$$D_{\mathrm{f}}(\Sigma_1, \Sigma_2) = \varphi(\Sigma_1) - \varphi(\Sigma_2) - \varphi'(\Sigma_2)(\Sigma_1 - \Sigma_2), \tag{21}$$

where the seed function $f$ is combined with a function $\lambda : \mathcal{M} \to \mathbb{R}^{\mathrm{C}}$ that maps an SPD matrix to a vector containing its eigenvalues: $\varphi = f \circ \lambda$.
Or, $\lambda$ can be the trace function, $\lambda : \mathcal{M} \to \mathbb{R}$ that maps an SPD matrix to its trace. For convenience, $f \circ \lambda$ will be referred to as $f(X)$ or $f(\Sigma)$. Depending on the seed function used, various divergences

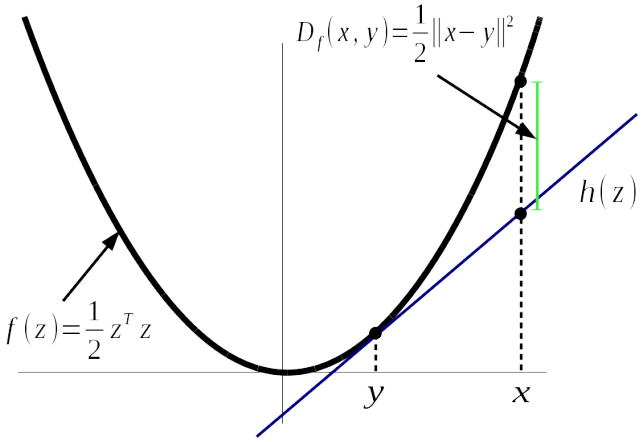



**Figure 1.** Geometry of the Bregman divergence with the seed function $f(z) = \frac{1}{2}z^\intercal z$. $h(z)$ is a hyperplane tangent to $f(z)$ at $y$. While it accuratly represent $f(y)$, it underestimate $f(x)$. The Bregman divergence measure how much the the representation of $f(x)$ on $h(z)$ *diverges* from $f(x)$.

can be defined defined from the Bregman divergence.

**Euclidean divergence**
The Frobenius norm is also a Bregman divergence in disguise. It is obtained when the seed function is the squared norm $f(x) = \frac{1}{2}\|x\|_2^2$ [19].

$$D_{\mathrm{E}}(\Sigma_1, \Sigma_2) = \|\Sigma_1 - \Sigma_2\|_F \tag{22}$$

The Euclidean mean of SPD matrices correspond to their arithmetic mean (9).

**Kullback-Leibler divergence**
Using the *Shannon entropy* $f(x) = \sum_i x_i \log x_i$ yields the *Kullback-Leibler* divergence [20]. It is also known as the *relative entropy* or *discrimination information*. The Kullback-Leibler divergence of SPD matrices $\Sigma_1, \Sigma_2 \in \mathcal{M}^C$ is given by:

$$D_{\mathrm{KL}}(\Sigma_1, \Sigma_2) = \frac{1}{2}\log\frac{\det(\Sigma_2)}{\det(\Sigma_1)} + \mathrm{tr}(\Sigma_2\Sigma_1) - C \tag{23}$$

The mean of SPD matrices $\Sigma_i \in \mathcal{M}^C$ induced by the Kullback-Leibler divergence is calculated iteratively [21].

**Log-det divergence**
Another function often used in Bregman divergences of symmetric matrices is the *logarithmic barrier* [19,21–23]

$$f(x) = -log(x) \rightarrow f(\Sigma) = -\log\det(\Sigma)$$

The corresponding divergence is called the *log-det* divergence and is given by [21]:

$$D_{\mathrm{ld}}(\Sigma_1, \Sigma_2) = \left\langle \Sigma_1, \Sigma_2^{-1} \right\rangle - \log\det(\Sigma_1\Sigma_2^{-1}) - C \tag{24}$$

135

The asymmetry of divergences result in the concept of right- and left-sided mean:

$$D_{\mathrm{f}}(\Sigma_1, \Sigma_2) \neq D_{\mathrm{f}}(\Sigma_2, \Sigma_1) \Rightarrow \arg\min_{\Sigma\in\mathcal{M}_C} \sum_{i=1}^{I} d^2(\Sigma_i, \Sigma) \neq \arg\min_{\Sigma\in\mathcal{M}_C} \sum_{i=1}^{I} d^2(\Sigma, \Sigma_i)$$

It is usually sufficient to consider a single sided divergence. In this work right sided-divergence and mean are used. In some cases however, the asymmetry can be undesirable. This has led to the symmetrization of some Bregman divergences. Often the symmetrization consist in an averaging of left- and right-sided divergence.

**S-divergence**

An example of a symmetric divergence is the S-divergence. It is obtained from the *Jensen-Shannon* divergence which is a symmetrized Bregman divergence:

$$\begin{aligned} D_{\text{J-S}}(\Sigma_1, \Sigma_2) &= \frac{1}{2}\left(D_{\text{f}}(\Sigma_1, \frac{\Sigma_1+\Sigma_2}{2}) + D_(\frac{\Sigma_1+\Sigma_2}{2}, \Sigma_2)\right) \\ &= \frac{1}{2}\left(\operatorname{tr} f(\Sigma_1) + \operatorname{tr} f(\Sigma_2)\right) - \operatorname{tr} f(\frac{\Sigma_1+\Sigma_2}{2}) \end{aligned} \tag{25}$$

The S-divergence is obtained by using the logarithmic barrier function for the positive definite cone $f(\Sigma) = -\log\det(\Sigma)$ as seed in $D_{\text{S-J}}$ [23]:

$$\div S(\Sigma_1, \Sigma_2) = \log\det(\frac{\Sigma_1+\Sigma_2}{2}) - \frac{1}{2}\log\det(\Sigma_1\Sigma_2) \tag{26}$$

Despite its symmetry, S-Divergence is not a metric. It does not satisfy the triangular inequality criterion. However, its square root has been shown to be a distance [23].

Other symmetric divergences can be obtained in the same fashion; for instance the *Jeffreys divergence* which is a symmetrized Kullback-Leibler divergence: $D_{\text{J}}(\Sigma_1, \Sigma_2) = D_{\text{KL}}(\Sigma_1, \Sigma_2) + D_{\text{KL}}(\Sigma_2, \Sigma_1)$ [23].

Another family of divergence is defined when the right- and left-sided divergence are mixed in a weighted manner. One such family is the *α-divergence* [24].

**Log-det α-divergence**

In this work, the $\alpha$-divergence used is defined by [21]:

$$D_{\text{f}}^{\alpha}(\Sigma_1, \Sigma_2) = \frac{4}{1-\alpha^2}\left[\frac{1-\alpha}{2}f(\Sigma_1) + \frac{1+\alpha}{2}f(\Sigma_2) - f\left(\frac{1-\alpha}{2}\Sigma_1 + \frac{1+\alpha}{2}\Sigma_2\right)\right], \alpha^2 \neq 1 \tag{27}$$

$D_{\text{f}}^{\alpha}$ can be expressed in terms of Bregman divergence as:

$$D_{\text{f}}^{\alpha} = \frac{4}{1-\alpha^2}\left[\frac{1-\alpha}{2}D_{\text{f}}\left(\Sigma_1, \frac{1-\alpha}{2}\Sigma_1 + \frac{1+\alpha}{2}\Sigma_2\right) + \frac{1+\alpha}{2}D_{\text{f}}\left(\Sigma_2, \frac{1-\alpha}{2}\Sigma_1 + \frac{1+\alpha}{2}\Sigma_2\right)\right], \alpha^2 \neq 1 \tag{28}$$

$\alpha$-divergences at $\alpha = \pm 1$ are obtained through the limit values $\lim_{\alpha\to\pm1} D_{\text{f}}^{\alpha}$.
Using the logarithmic-barrier function yields:

$$\begin{aligned} D_{\text{LD}}^{\alpha}(\Sigma_1, \Sigma_2) &= \frac{4}{1-\alpha^2}\log\det\left(\frac{1-\alpha}{2}\left(\Sigma_1\Sigma_2^{-1}\right)^{\frac{1+\alpha}{2}} + \frac{1+\alpha}{2}\left(\Sigma_2\Sigma_1^{-1}\right)^{\frac{1-\alpha}{2}}\right), \quad -1 < \alpha < 1 \\ D_{\text{LD}}^{1}(\Sigma_1, \Sigma_2) &= \operatorname{tr}\left(\Sigma_2^{-1}\Sigma_1 - \mathbf{I}\right) - \log\det\left(\Sigma_2^{-1}\Sigma_1\right) \\ D_{\text{LD}}^{-1}(\Sigma_1, \Sigma_2) &= \operatorname{tr}\left(\Sigma_1^{-1}\Sigma_2 - \mathbf{I}\right) - \log\det\left(\Sigma_1^{-1}\Sigma_2\right) \end{aligned} \tag{29}$$

$D_{\text{LD}}^{1}$ and $D_{\text{LD}}^{-1}$ are right- and left-sided Bregman divergences respectively.
At $\alpha = 0$, the log-det $\alpha$ divergence yields a symmetric divergence corresponding to the *Bhattacharrya* divergence [21,23].

**Table 1.** Distances, divergences and means considered in the experimental study.

| | Distance/Divergence | Mean | References |
|---|---|---|---|
| Euclidean | $d_{\mathrm{E}}(\Sigma_1,\Sigma_2) = \|\Sigma_1 - \Sigma_2\|_F$ | $\bar{\Sigma}_{\mathrm{E}} = \frac{1}{I}\sum_{i=1}^{I}\Sigma_i$ | |
| Log-Euclidean | $d_{\mathrm{LE}}(\Sigma_1,\Sigma_2) = \|\log(\Sigma_1) - \log(\Sigma_2)\|_F$ | $\bar{\Sigma}_{\mathrm{LE}} = \exp\left(\sum_{i=1}^{I}\log(\Sigma_i)\right)$ | [9] |
| Affine-invariant | $d_{\mathrm{AI}}(\Sigma_1,\Sigma_2) = \|\log(\Sigma_1^{-1}\Sigma_2)\|_F$ | Algorithm 3 in [25] | [8**?** ] |
| Kullback-Leibler | $D_{\mathrm{KL}}(\Sigma_1,\Sigma_2) = \frac{1}{2}\log\frac{\det(\Sigma_2)}{\det(\Sigma_1)} + \mathrm{tr}(\Sigma_2\Sigma_1) - C$ | Algorithm 1 in [21] | [21,26] |
| S-divergence | $D_{\mathrm{S}}(\Sigma_1,\Sigma_2) = \log\det(\frac{\Sigma_1+\Sigma_2}{2}) - \frac{1}{2}\log\det(\Sigma_1\Sigma_2)$ | Eq. (17-20) in [22] | [22,23] |
| $\alpha$-divergence | $D^{\alpha}_{\mathrm{LD}}(\Sigma_1,\Sigma_2)$ from Eq. (29) | Algorithm 1 in [21] | [21] |
| Bhattacharyya | $D_{\mathrm{B}}(\Sigma_1,\Sigma_2) = \left(\log\frac{\det\frac{1}{2}(\Sigma_1+\Sigma_2)}{(\det(\Sigma_1)\det(\Sigma_2))^{1/2}}\right)^{1/2}$ | Algorithm 1 in [21] | [21,27] |
| Wasserstein | | | |

### 2.2.6. Wasserstein

### *2.3. Minimum Distance to Mean classifier for SSVEP*

The considered classifier is described in section 2.1. It is given the name *Minimum Distance to Mean* or MDMclose, and was inspired from [7] where it is limited to Riemannian mean. Covariance matrices of EEG trials are classified based on their distance to the centers of the classes (i.e. means or centroids). To embed frequency information in the covariance matrices, we use a construction of matrices proposed in [11]. Let $X \in \mathbb{R}^{C\times N}$ be an EEG trial measured on $C$ channels and $N$ samples in a SSVEP experiment with $F$ stimulus blinking at different frequencies. The covariance matrices are estimated from a modified version of the input signal $X$:

$$X \in \mathbb{R}^{C\times N} \rightarrow \begin{bmatrix} X_{\mathrm{freq}_1} \\ \vdots \\ X_{\mathrm{freq}_F} \end{bmatrix} \in \mathbb{R}^{FC\times N}, \tag{30}$$

where $X_{\mathrm{freq}_f}$ is the input signal $X$ band-pass filtered around frequency $\mathrm{freq}_f$, $f = 1,\ldots,F$. Henceforth, all EEG signals will be considered as filtered and modified by Eq. (30). The associated covariance matrix $\Sigma \in \mathbb{R}^{FC\times FC}$ is estimated using the Schäfer skrinkage estimator [28].

For SSVEP classification, $K = F + 1$ classes are considered: one class for each target frequency, and one for the resting state. As described in Algorithm 1, from $I$ labelled training trials $\{X_i\}_{i=1}^{I}$ recorded per subject, $K$ centers of classes $\bar{\Sigma}^{(k)}$ are estimated (step 3). In this step, outliers matrices are removed to have a reliable mean estimation [**?** ]. A new unlabeled test trial $Y$ is predicted to belong to the class whose mean $\bar{\Sigma}^{(k)}$ is the closest to the trial covariance matrix, w.r.t. one of the distances from Table 1 (step 5).

**Algorithm 1** Minimum Distance to Mean Classifier

Inputs: $X_i \in \mathbb{R}^{FC\times N}$, for $i = 1,\ldots,I$, a set of labelled EEG trials.
Inputs: $\mathcal{I}(k)$, a set of indices of trials belonging to class $k$.
Input: $Y \in \mathbb{R}^{FC\times N}$, an unlabeled test EEG trial.
Output: $k^*$, the predicted label of $Y$.

1: Compute covariance matrices $\Sigma_i$ of $X_i$
2: **for** $k = 1$ **to** $K$ **do**
3: Compute center of class : $\bar{\Sigma}^{(k)} = \mu(\Sigma_i : i \in \mathcal{I}(k))$
4: **end**
5: Compute covariance matrix $\Sigma$ of $Y$, and classify it : $k^* = \arg\min_k d(\Sigma, \bar{\Sigma}^{(k)})$
6: **return** $k^*$

## 3. Experimental Results

This section presents experimental results obtained applying Euclidean and Riemannian distances in SSVEP classification task. The first part of this section describes the data used and the second part provides the assessment of the classification for the considered distances and divergences.

### *3.1. SSVEP Dataset*

The experimental study is conducted on multichannel EEG signals recorded during an SSVEP-based BCI experiment [12]. EEG are recorded on $C = 8$ channels from 12 subjects. The subjects are presented with $F = 3$ visual target stimuli blinking respectively at 13Hz, 21Hz and 17Hz. It is a $K = 4$ classes setup combining $F = 3$ stimulus classes and one resting class (no-SSVEP). In a session, 32 trials are recorded: 8 for each visual stimulus and 8 for the resting class. The number of sessions recorded per subject varies from 2 to 5. For each subject, a test set is made of 32 trials while the remaining trials (which might vary from 32 to 128) make up for the training set.

### *3.2. Results and Discussion*

Discuss the invariance to right- and left-multiplication by positive matrices. It brings a significanc advantage over Euclidean metrics, in terms of electrode placement and unforseen displacement in electrodes position, and can even alleviate anatomical differences.

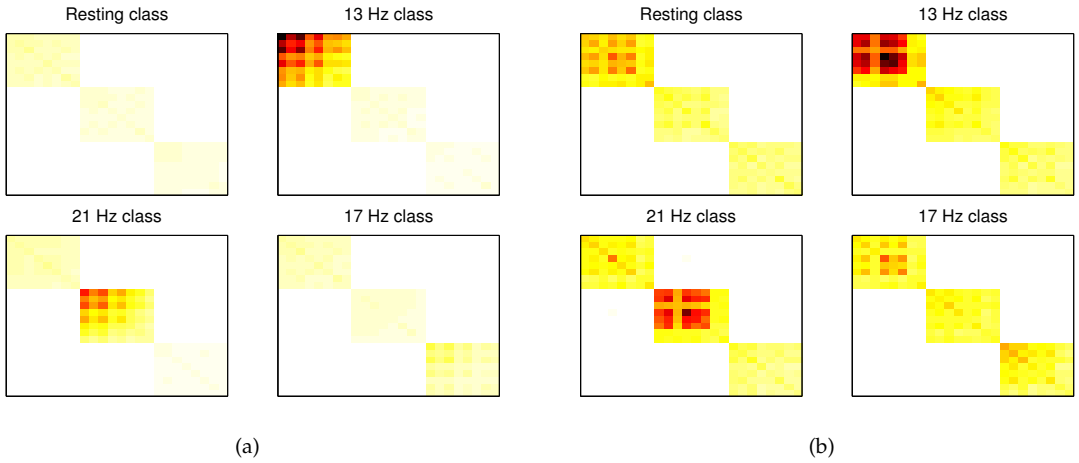
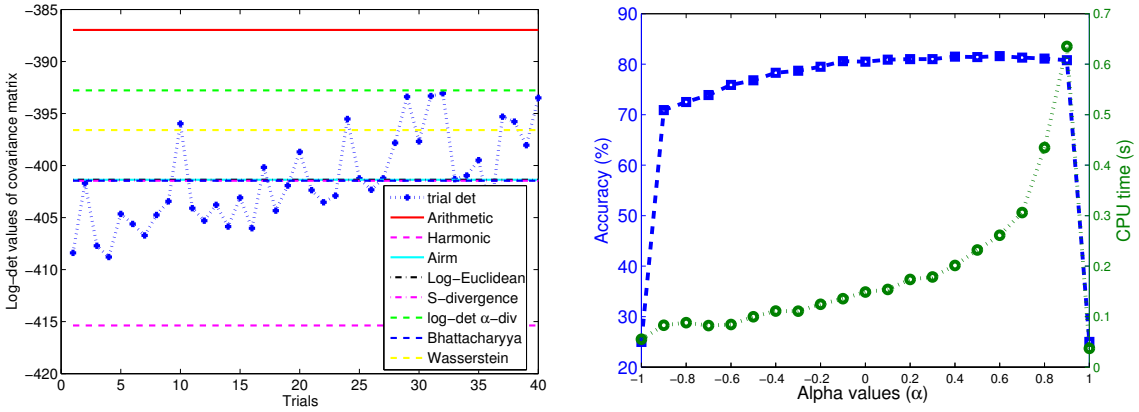



**Figure 2.** Representation of covariance matrices: each image is the covariance matrix mean $\bar{\Sigma}^{(k)}$ of the class $k$, for one session of the recording. The diagonal blocks show the covariance in different frequency bands, i.e. 13Hz in the upper-left block, 21Hz in the middle, and 17Hz in the bottom-right. Subjects with highest (a) and lowest (b) BCI performance.

The covariance matrices obtained from SSVEP data extended with Eq. (30) have interesting features, allowing the discrimination between signals of identical sources but with different frequencies. Fig. 2 shows the $K$ classes mean covariance matrices $\bar{\Sigma}^{(k)}$ from subjects with the highest (a) and lowest (b) classification accuracies. The three $8\times8$ diagonal blocks hold the covariance matrices of the $F = 3$ target frequencies. Inter-frequencies covariances blocks are almost null. In each mean covariance matrix, the block holding the covariance of the target frequency has the largest values. For the resting class, all $F$ blocks tend to have similar and small values. These features are more visible in the subject with the highest classification accuracy, and less visible in the one with lowest classification accuracy. It is interesting to see that features used for classification have a physiological meaning allowing an intuitive understanding, as opposed to *black-boxes* approaches such as LDA or SVM. EEG processing complexity is encoded by the distance and not by machine learning.

Based on those covariance matrices, the different distances and means of Table 1 are compared in terms of classification accuracy and average CPU time elapsed on a trial classification, which involves the computation of 4 means of class and a distance to each mean. Table 2 summarizes results obtained for each subject and each distance/divergence. Euclidean distance yields drastically low accuracy. This support the fact that using Euclidean distance and Arithmetic mean on SPD matrices is not appropriate. This is generally attributed to the invariance under inversion and the fact that the determinant of the Arithmetic mean of SPD matrices can be larger than the determinant of its parts; it is referred to as the *swelling effect*. Since the value of the determinant is a direct measure of dispersion of the multivariate variables (i.e. EEG channels and frequency bands), it leads to poor discrimination in the classification task. The swelling effect of Arithmetic mean is shown in Fig. 3(a): the determinant of the Arithmetic mean is strictly larger than other means, the Log-Euclidean, Affine-Invariant and Bhattacharyya ones yielding similar determinants, close to trials values.




(a) (b)

**Figure 3.** (a): Swelling effect of Arithmetic mean shown through log-determinant values. Training trials are taken from the 13Hz class of the subject with the highest BCI performance. Log-determinant values are given for each trial covariance (points), and for means of Table 1 (horizontal lines). (b): Classification accuracy and CPU time, obtained with $\alpha$-divergence for $-1 \leqslant \alpha \leqslant 1$.

Riemannian distances significantly improve classification performances, with $\alpha$-divergence yielding the best results (81.56%). The value of $\alpha$ was set to 0.6 through cross-validation. This procedure lasted 225.42 seconds and makes $\alpha$-divergence the most costly method, due to the optimization of its parameter $\alpha$. Log-Euclidean yields lower classification accuracy (average 78.98%) but could be computed faster than $\alpha$-divergence or Affine-Invariant distance. However, the Bhattacharyya distance has the lowest computational cost of the considered Riemannian distances (average CPU time 0.140s), with a higher average accuracy of 80.51%. So, it is good trade-off between efficiency and speed. The accuracies and CPU time of the $\alpha$-divergence at different values of $\alpha$ are shown in Fig. 3(b). It is seen that for $\alpha = \pm 1$, where $\alpha$-divergence represents a Bregman divergence associated with the log-determinant function, the classification accuracy are at the lowest accuracy (25%). For all other values of alpha, the expected accuracy is 78.85$\pm$3.3% and one can choose $-1 < \alpha < 1$ without any major impact on classification results.

This experiment on real EEG data shows that it is crucial to process covariance matrices with dedicated Riemannian tools, impacting the efficiency of the classification.

## 4. Conclusion

Riemannian approaches have been successfully applied on EEG signals for brain computer interfaces. Straightforward algorithms, such as Minimum Distance to Mean, provide competitive

results with state-of-the-art methods, without requiring meticulous parametrization or optimization. Working on covariance matrices in Riemannian spaces offers a wide choice of distances, embedding desirable invariances: it is thus possible to avoid the computation of user-specific spatial filters which are sensitive to artifacts and outliers. Nonetheless, the estimation of the Riemannian geometric mean has a strong impact on the classifier accuracy. This study investigates the performance of several distances and divergence on a real EEG dataset in the context of BCI based on the SSVEP paradigm. The experimental results indicate that the $\alpha$-divergence yields the best accuracy after the selection of the best $\alpha$ values, but the Bhattacharyya distance has the lowest computational cost while providing honorable accuracies.

**Supplementary Materials:** The following are available online at www.mdpi.com/link, Figure S1: title, Table S1: title, Video S1: title.

**Acknowledgments:** All sources of funding of the study should be disclosed. Please clearly indicate grants that you have received in support of your research work. Clearly state if you received funds for covering the costs to publish in open access.

**Author Contributions:** For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "X.X. and Y.Y. conceived and designed the experiments; X.X. performed the experiments; X.X. and Y.Y. analyzed the data; W.W. contributed reagents/materials/analysis tools; Y.Y. wrote the paper." Authorship must be limited to those who have contributed substantially to the work reported.

**Conflicts of Interest:** Declare conflicts of interest or state "The authors declare no conflict of interest." Authors must identify and declare any personal circumstances or interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. Any role of the funding sponsors in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript, or in the decision to publish the results must be declared in this section. If there is no role, please state "The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results".

## Bibliography

1. Niedermeyer, E.; Silva, F.H.L.d. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*; Lippincott Williams & Wilkins, 2005.
2. Rivet, B.; Souloumiac, A.; Attina, V.; Gibert, G. xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain-Computer Interface. *Biomedical Engineering, IEEE Transactions on* **2009**, *56*, 2035–2043.
3. Wang, S.; James, C.J. Enhancing Evoked Responses for BCI Through Advanced ICA Techniques. Advances in Medical, Signal and Information Processing (MEDSIP), 2006, pp. 1–4.
4. Tomioka, R.; Aihara, K.; Müller, K.R. Logistic regression for single trial EEG classification. Advances in neural information processing systems (NIPS), 2007, Vol. 19, pp. 1377–1384.
5. Johannes, M.G.; Pfurtscheller, G.; Flyvbjerg, H. Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clinical Neurophysiology* **1999**, *110*, 787–798.
6. Kalunga, E.; Djouani, K.; Hamam, Y.; Chevallier, S.; Monacelli, E. SSVEP enhancement based on Canonical Correlation Analysis to improve BCI performances. AFRICON, 2013, 2013, pp. 1–5.
7. Barachant, A.; Bonnet, S.; Congedo, M.; Jutten, C. Multiclass brain–computer interface classification by Riemannian geometry. *Biomedical Engineering, IEEE Transactions on* **2012**, *59*, 920–928.
8. Moakher, M. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications* **2005**, *26*, 735–747.
9. Arsigny, V.; Fillard, P.; Pennec, X.; Ayache, N. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications* **2007**, *29*, 328–347.
10. Amari, S.I. $\alpha$-Divergence Is Unique, Belonging to Both f-Divergence and Bregman Divergence Classes. *Information Theory, IEEE Transactions on* **2009**, *55*, 4925–4931.
11. Congedo, M.; Barachant, A.; Andreev, A. A New Generation of Brain-Computer Interface Based on Riemannian Geometry. *arXiv preprint arXiv:1310.8115* **2013**.
12. Kalunga, E.K.; Chevallier, S.; Rabreau, O.; Monacelli, E. Hybrid interface: Integrating BCI in multimodal human-machine interfaces. 2014 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), 2014, pp. 530–535.

13. Scholkopf, B.; Smola, A.J. *Learning with kernels: support vector machines, regularization, optimization, and beyond*; MIT press, 2001.
14. Karcher, H. Riemannian center of mass and so called karcher mean. *arXiv preprint arXiv:1407.2087* **2014**.
15. Lim, Y.; Pálfia, M. Matrix power means and the Karcher mean. *Journal of Functional Analysis* **2012**, *262*, 1498–1514.
16. Cartan, E. Groupes simples clos et ouverts et géométrie riemannienne. *Journal de mathématiques pures et appliquées* **1929**, pp. 1–34.
17. Pennec, X.; Fillard, P.; Ayache, N. A Riemannian Framework for Tensor Computing. *International Journal of Computer Vision* **2006**, *66*, 41–66.
18. Bregman, L.M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics* **1967**, *7*, 200–217.
19. Dhillon, I.S.; Tropp, J.A. Matrix nearness problems with Bregman divergences. *SIAM Journal on Matrix Analysis and Applications* **2007**, *29*, 1120–1146.
20. Nielsen, F.; Nock, R. Sided and symmetrized Bregman centroids. *IEEE transactions on Information Theory* **2009**, *55*, 2882–2904.
21. Chebbi, Z.; Moakher, M. Means of Hermitian positive-definite matrices based on the log-determinant $\alpha$-divergence function. *Linear Algebra and its Applications* **2012**, *436*, 1872–1889.
22. Cherian, A.; Sra, S.; Banerjee, A.; Papanikolopoulos, N. Efficient similarity search for covariance matrices via the Jensen-Bregman LogDet divergence. 2011 International Conference on Computer Vision. IEEE, 2011, pp. 2399–2406.
23. Sra, S. Positive definite matrices and the S-divergence. *Proceedings of the American Mathematical Society* **2016**, *144*, 2787–2797.
24. Nielsen, F.; Nock, R.; Amari, S.i. On clustering histograms with k-means by using mixed $\alpha$-divergences. *Entropy* **2014**, *16*, 3273–3301.
25. Fletcher.; Joshi, S. Principal Geodesic Analysis on Symmetric Spaces: Statistics of Diffusion Tensors. In *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*; Springer, 2004; Vol. 3117, *LNCS*, pp. 87–98.
26. Kang, H.; Nam, Y.; Choi, S. Composite common spatial pattern for subject-to-subject transfer. *Signal Processing Letters, IEEE* **2009**, *16*, 683–686.
27. Nielsen, F.; Bhatia, R. *Matrix Information Geometry*; Springer Publishing Company, Incorporated, 2012.
28. Schäfer, J.; Strimmer, K. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology* **2005**, *4*.


© 2016 by the authors. Submitted to *Entropy* for possible open access publication under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

| | Euclidean | | | | | | Riemannian | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Arithmetic | | Harmonic | | Geometric | | Log-Euclidean | | Affine-invariant | | $\alpha$-divergence | | Bhattacharyya | | Kullbacl-Leibler | | S-divergence | | Wassestein | |
| Sub. | acc (%) | time(s) | acc (%) | time(s) | acc (%) | time(s) | acc (%) | time(s) | acc (%) | time(s) | acc (%) | time(s) | acc (%) | time(s) | acc (%) | time(s) | acc (%) | time(s) | acc (%) | time(s) |
| 1 | 53.12 | 0.025 | 48.44 | 0.015 | 00.00 | 0.000 | 71.88 | **0.150** | **73.44** | 0.194 | 59.37 | 0.155 | 68.75 | 0.225 | 53.12 | 0.030 | 68.75 | 0.220 | 54.69 | 0.630 |
| 2 | 43.75 | 0.020 | 31.25 | 0.025 | 00.00 | 0.000 | 78.13 | 0.160 | 79.69 | 0.190 | 79.69 | 0.200 | **81.25** | **0.065** | 54.69 | 0.045 | 81.25 | 0.255 | 54.69 | 0.285 |
| 3 | 67.19 | 0.020 | 67.19 | 0.015 | 00.00 | 0.000 | 85.94 | 0.120 | 85.93 | 0.205 | **95.31** | 0.155 | 85.94 | **0.100** | 71.88 | 0.060 | 85.94 | 0.200 | 76.56 | 0.280 |
| 4 | 54.68 | 0.030 | 42.19 | 0.010 | 00.00 | 0.000 | 84.38 | 0.225 | 87.50 | 0.315 | **89.07** | 0.250 | 85.94 | **0.100** | 48.44 | 0.025 | 85.94 | 0.120 | 65.62 | 0.310 |
| 5 | 37.50 | 0.020 | 26.56 | 0.015 | 00.00 | 0.000 | 62.50 | **0.115** | 68.75 | 0.290 | **73.44** | 0.140 | 65.62 | 0.125 | 67.19 | 0.030 | 65.63 | 0.110 | 45.31 | 0.660 |
| 6 | 34.37 | 0.015 | 39.06 | 0.020 | 00.00 | 0.000 | 84.38 | 0.120 | 85.94 | 0.210 | **87.50** | 0.145 | 82.81 | **0.100** | 62.50 | 0.030 | 82.81 | 0.130 | 53.13 | 0.300 |
| 7 | 60.42 | 0.027 | 47.92 | 0.013 | 00.00 | 0.000 | 87.50 | 0.267 | 88.54 | 0.410 | **91.66** | 0.417 | 86.46 | **0.137** | 54.17 | 0.043 | 86.46 | 0.243 | 69.79 | 0.777 |
| 8 | 67.19 | 0.035 | 68.75 | 0.015 | 00.00 | 0.000 | 90.63 | 0.215 | **92.19** | 0.290 | **92.19** | 0.290 | **92.19** | **0.125** | 71.88 | 0.050 | 92.19 | 0.165 | 85.94 | 0.335 |
| 9 | 57.81 | 0.035 | 42.19 | 0.015 | 00.00 | 0.000 | 70.31 | 0.275 | 70.31 | 0.380 | **75.00** | 0.300 | 67.19 | **0.134** | 60.94 | 0.050 | 67.19 | 0.160 | 62.50 | 0.310 |
| 10 | 38.28 | 0.035 | 28.12 | 0.013 | 00.00 | 0.000 | 75.00 | 0.254 | 80.47 | 0.514 | **82.03** | 0.510 | 78.13 | **0.160** | 67.97 | 0.028 | 78.13 | 0.263 | 51.56 | 0.650 |
| 11 | 48.44 | 0.025 | 46.88 | 0.010 | 00.00 | 0.000 | 60.94 | 0.144 | 65.63 | 0.235 | 57.81 | 0.150 | **75.00** | **0.105** | 39.06 | 0.040 | 75.00 | 0.195 | 56.25 | 0.575 |
| 12 | 71.25 | 0.032 | 53.75 | 0.022 | 00.00 | 0.000 | 96.25 | **0.292** | 96.69 | 0.534 | 95.62 | 0.634 | **96.88** | 0.300 | 76.88 | 0.042 | 96.88 | 0.466 | 82.50 | 1.042 |
| **Avg.** | 52.83 | 0.027 | 45.19 | 0.016 | 00.00 | 0.000 | 78.98 | 0.194 | 81.27 | 0.314 | **81.56** | 0.279 | 80.51 | **0.140** | 60.72 | 0.039 | 80.51 | 0.210 | 63.21 | 0.513 |

**Table 2.** Subject classification accuracies (acc(%)) and average CPU time (time(s)) elapsed for the classification of a single trial. Classification is performed with MDM using either Euclidean or Riemannian means (see Table 1).