

UNIVERSITÉ LIBRE DE BRUXELLES
Faculté des Sciences
Département d'Informatique

Preparatory work for the Master Thesis

Towards the end of seclusion
for physically impaired
but cognitively intact individuals

Cédric Simar

Promotors : Prof. Gianluca Bontempi and Prof. Guy Chéron
Preparatory work for the Master Thesis in Computer Sciences

Contents

1	Introduction	1
1.1	Background and objectives of the thesis	1
1.2	Notations	2
1.3	Abbreviations	2
2	State of the art	3
2.1	Electroencephalography	3
2.1.1	General description	3
2.1.2	Frequency bands	4
2.1.3	Artefacts	5
2.2	Types of BCIs	6
2.2.1	Evoked potential	6
2.2.2	Motor imagery	8
2.3	Signal processing	9
2.3.1	Band-pass filter	9
2.3.2	Power Spectral Density	9
2.3.3	Spatial filtering	10
2.3.4	Covariance matrices	11
2.4	Riemannian geometry	12
2.4.1	Topological spaces	13
2.4.2	Manifolds	13
2.4.3	Riemannian differential manifolds	13
2.4.4	Matrix logarithm and exponential	14
2.4.5	Riemannian distance	15
2.4.6	Projection to the tangent space, logarithmic and exponential maps	15

2.4.7	Riemannian mean	16
3	EEG classification for BCI applications	19
3.1	BCI protocol	19
3.2	Classifiers	19
3.2.1	Support Vector Machine	19
3.2.2	Convolutional Neural Network	21
4	Preliminary classification tests	22
4.1	Problem description	22
4.2	Data	22
4.3	Methodology	23
4.3.1	Validation	23
4.3.2	Evaluation metrics	23
4.4	Models	23
4.4.1	PSD with Logistic Regression	23
4.4.2	Deep Learning on raw EEG data	24
4.4.3	Linear and Riemannian based SVM	25
4.5	Results	26
4.6	Discussion	26
A	Covariance matrices are symmetric positive definite	28

Chapter 1

Introduction

1.1 Background and objectives of the thesis

A Brain-Computer Interface (BCI) is a system that establishes a direct communication between a brain and a device without using the neuromuscular pathways. The design and implementation of a BCI system typically require a multidisciplinary approach in order to combine expertise from different fields such as neurophysiology, electronics, signal processing and algorithmics. In the framework of this thesis, I received the support and supervision of Prof. Gianluca Bontempi for the algorithmic aspect, Prof. Guy Chron for the neurophysiological aspect, Mathieu Petieau for the electronics and signal processing aspects as well as the logistics support from the Laboratory of Neurophysiology and Movement Biomechanics of the Faculty of Motor Sciences, Universit Libre de Bruxelles (ULB).

Currently, the main field of application of BCI systems is the medical domain. Encouraging results have been reported with the use of Neurofeedback therapy (NFT) for the treatment of Attention Deficit Hyperactivity Disorder (ADHD) [7] or to facilitate physical rehabilitation. Additionally, accurate and robust BCI systems would considerably enhance the quality of life of thousands of patients suffering from a locked-in syndrome (LIS) [40] or from severe motor paralysis caused by a trauma or a generative disease. Any progress in the development of such BCI systems is another hope that these patients will again be able to communicate, interact with or even regain control over their environment. Patients suffering from brain stem lesions caused by a stroke could again move autonomously by controlling their wheelchair or drink a sip of water by grabbing a glass with their prosthetic arm.

The objective of the Master thesis is to initiate the development of such BCI systems and obtain sufficiently encouraging results that can hopefully lead to a PhD thesis

The purpose of this work is to describe succinctly the state of the art in non-invasive electroencephalographic signal processing, types BCIs and the Machine Learning algorithms applied in this domain; implement and compare the performances of such algorithms

on a real dataset; determine a realistic approach to achieve the objective of the Master thesis.

1.2 Notations

It may be interesting to gather here the main notations, their meaning, and the page where it is defined in more detail or possibly a bibliographic reference.

$\ \cdot\ _{\mathcal{F}}$	Denotes the Frobenius norm
$tr(\cdot)$	Denotes the trace operator
$E \in \mathbb{N}$	Number of electrodes (channels)
$i, j \in \{1, \dots, E\}$	Indices of the EEG channels
$N \in \mathbb{N}$	Number of samples in a portion (epoch) of the EEG signal
$X \in \mathbb{R}^{E \times N}$	Matrix representation of an epoch
$X_y \in \mathbb{R}^{E \times N}$	Matrix representation of an epoch from class $y \in Y$
$x_i \in \mathbb{R}^C$	Vector representation of the EEG signal at channel i
$Y \in \mathbb{N}$	Class of X
$S \in \mathbb{R}^{E \times E}$	Covariance matrix estimated from a single epoch
$S_y \in \mathbb{R}^{E \times E}$	Covariance matrix from class $y \in Y$
$C \in \mathbb{R}^{E \times E}$	Average covariance matrix estimated from several epochs
$C_y \in \mathbb{R}^{E \times E}$	Average covariance matrix of class $y \in Y$

1.3 Abbreviations

BCI	Brain-Computer Interface
CNN	Convolutional Neural Network
CSP	Common Spatial Pattern
CV	Cross Validation
EEG	Electroencephalography
EMG	Electromyography
EOG	Electrooculography
ERD	Event Related Desynchronisation
ERP	Event Related Potential
ERS	Event Related Synchronisation
MI	Motor Imagery
SCM	Sample Covariance Matrix
SNR	Signal to noise ratio
SPD	Symmetric Positive Definite
SSEP	Steady State Evoked Potential
SSVEP	Steady State Visual Evoked Potential
SVM	Support Vector Machine

2.1.2 Frequency bands

- Delta waves (< 4 Hz) has long been observed during stages of deep sleep in adults [43]. Additionally, Delta oscillation has been shown to play a role in attention selection [26] and its phase may also facilitates multiple unit activity (MUA) [49] and be involved in predictive timing [13].
- Theta waves (4 - 8 Hz) appear during an idling state but also play an important role in motor control. Indeed, Theta oscillation in the contralateral motor area were shown to be phase-locked with the onset of a movement, which acceleration also correlates with Theta power [36].
- Alpha waves (7 - 13 Hz) were the first observed [6] and most studied brain waves because of their noticable shape and amplitude. Alpha power increases significantly in the relaxed eyes-closed state and decreases in the eyes-open state. Oscillation in the 7 - 13 Hz frequency range are referred differently depending on the sensorimotor modalities (Alpha oscillation for the visual system, Mu oscillation for the sensorimotor system and Tau oscillation for the auditory system). It has been shown that the Mu oscillation decreases just before and during the execution of a movement. Interestingly, a similar behavior also occurs during the imagination [10] or the observation [21] of a movement, although to a lesser extent.

- Beta waves (13 - 30 Hz) appear during a concentration / attention state and also play a key role in the processing of sensorimotor information by the brain. Just as the Mu oscillation, it has been shown that Beta oscillation also decreases during the execution [36] or the observation ([52]; [11]; [5]) of a movement.
- Gamma waves (> 30 Hz) are involved in sensorimotor tasks [4], perception [23], attention [24], working memory [29] and associative memory [22]. Although there is no doubt about the important role of the Gamma oscillation in the brain, results often need to be interpreted with care since recordings of Gamma oscillation are often polluted with muscle artifacts as well as eye saccades and microsaccades [25].

Yet, the specific frequency ranges may slightly vary from person to person depending on the source and application.

2.1.3 Artefacts

Due to their weak amplitude, EEG signals exhibit a poor signal to noise ratio (SNR). Thus, in addition to synaptic activity, raw EEG signals also contain a certain proportion of noise, more specifically artifacts that have to be removed before features extraction can be performed with the finest precision. EEG artifacts can be either physiological or non-physiological (i.e. from recording instruments or patient environment). The main sources of physiological artifacts are eye blinks, cardiac contractions and muscle tensions that produce respectively electrooculography (EOG), electrocardiography (ECG) and electromyography (EMG) signals. A common source of non-physiological artifacts is the ambient power line noise arising from standard AC power supply frequencies (50 or 60 Hz depending on the country).

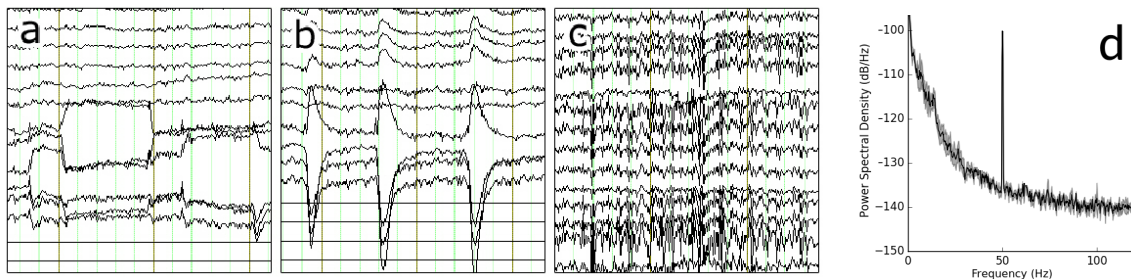


Figure 2.2: **a, b, c**, Physiological artifacts in EEG, respectively eye movement, eye blink and muscle tension. **d** Non-physiological artifact, power line noise at 50 Hz. [39]

Artefacts removal

Power line noise can commonly be removed using notch filters at 50 or 60 Hz. Notch filters are a type of band-stop filter [35] that attenuates a narrow range of specific frequencies (e.g. 49-51 Hz) without altering the rest of the frequency spectrum.

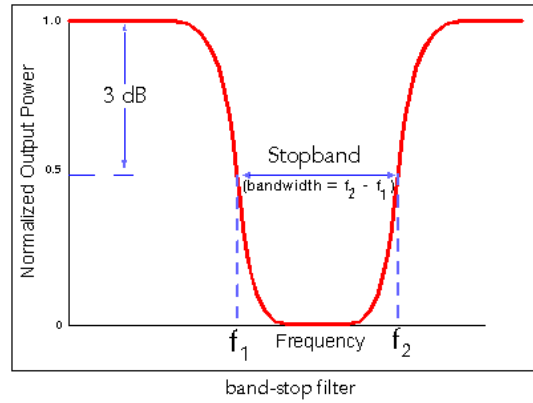


Figure 2.3: Band-stop filter [35].

On the other hand, physiological artifacts removal requires a more complex analysis as well as the use of sophisticated algorithms. For example, although eye blinks carry important physiological information, they produce positive EEG potentials that impact significantly frequencies of interest in the power spectrum. EOG and ECG artefacts are commonly removed with algorithmic tools such as Independent Components Analysis [14]. To the best of our knowledge, no algorithm is currently able to reliably remove large EMG artefacts from EEG signals in a satisfactory manner. This issue is an active field of research and should be properly addressed in order to develop robust BCI systems that will perform well in an out-of-lab environment.

2.2 Types of BCIs

2.2.1 Evoked potential

Event related potential

Event related potential (ERP) are brief and reproducible electrophysiological responses synchronized with a stimulus. A BCI based on ERP will detect either the presence or the absence of such particular electrophysiological response within a specific and limited time range (time domain).

The most common ERP used in BCI application is the P300, a positive amplitude peak first reported by [44] observable around 300 ms after an infrequent or surprising stimulus. From the first P300 BCI [18] to more recent works, different paradigms have been developed. An example of a commonly used paradigm is the P300 matrix speller paradigm where alphabetical and numerical characters are displayed on a matrix. Rows and columns are intensified randomly and the user is able to select a character by focusing on it, generating a P300 response when the specific character is intensified.

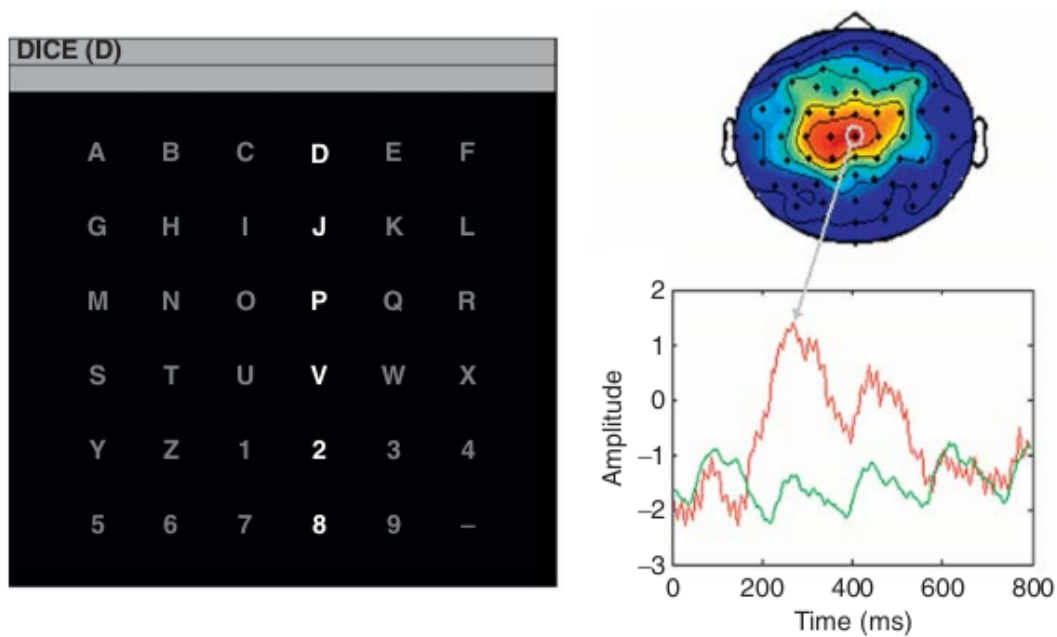


Figure 2.4: “The P300 matrix speller paradigm. *Left side*: The standard 6×6 matrix. The letter in parentheses at the top of the window is the current target letter ‘D’. A P300 should be elicited when the fourth column or first row is intensified. After the intensification sequence for a character epoch, the result is classified and online feedback is provided. *Right side*: (Top) A representative topography indicating a strong correlation over the central electrodes between the target letter intensifications and the EEG amplitude at 300 ms after the intensifications. (Bottom) The associated averaged P300 temporal response for the targets (red) and the non-targets (green) at electrode Cz”. Graphics and caption from (Krusienski and Wolpaw, 2009)

Although it has been shown that habituation, gaze direction [38] and eye movements [33] negatively impact the performances of a BCI based on the P300 ERP, very promising and enthusiastic results have been reported, notably with the use of Machine Learning algorithms (Rakotomamonjy and Guigue, 2008).

Error related potential

Improve the accuracy of BCI (Ferrez, 2007) Source: Error-related EEG potentials in brain-computer interfaces

Steady State Evoked Potential

Steady State Evoked Potentials (SSEP) are stable electrophysiological responses produced by the repetition of a stimulus over time. SSEP induce a repeated activation of specific cortical regions, resulting in the emergence of a signal oscillating at the same frequency as the stimulus repetition (frequency domain). In case of a visual stimulus,

SSEP are called Steady State Visual Evoked Potential (SSVEP).

When a person focuses on a repeating visual stimulus (e.g. a flashing light-emitting diode) a signal oscillating at the same frequency appears at a very localized region of the visual cortex (most noticable over the Oz electrode), which provides a robust and stable measure that can be used in a BCI application.

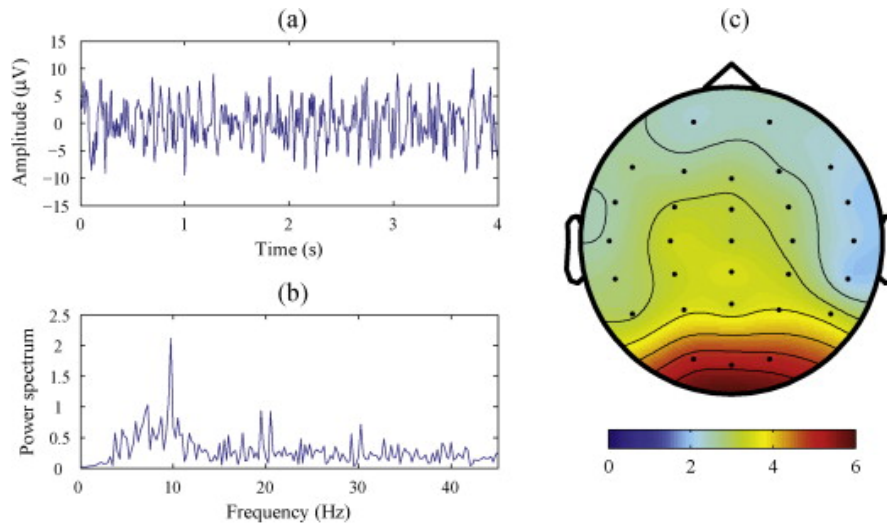


Figure 2.5: “SSVEP characteristics evoked at stimulus frequency of 10 Hz: (a) EEG waveform at channel Oz; (b) power spectrum at channel Oz; (c) scalp topography of power spectrum summed at 10 Hz and its harmonics”. Graphics and caption from (Zhang et al., 2015)

2.2.2 Motor imagery

BCIs based on motor imagery (MI) detect spatially localized changes in the power of specific frequency bands during the imagination of a movement. Such changes are called Event Related Desynchronisation (ERD) when there is an observable decrease of power and Event Related Synchronisation (ERS) when there is an observable increase of power in specific frequency bands.

A common practice in MI-based BCIs is to distinguish the imagination of a left- and right-hand movement. As described in 2.1.2, the imagination of a left- (right-) hand movement produces a noticable ERD of the Mu and Beta oscillations over the contralateral sensorimotor cortex, i.e. over the C4 (C3) electrode (see 2.1.1 for the electrodes position).

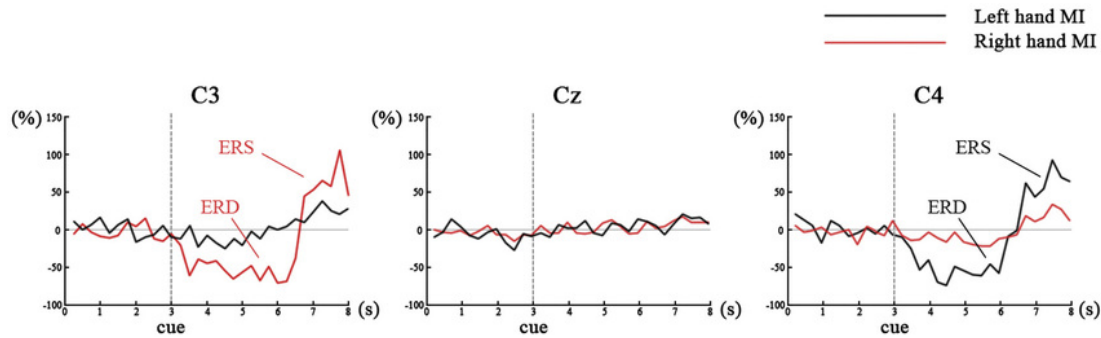


Figure 2.6: Plots of the averaged EEG power (relative to baseline) in the 8-12 Hz frequency band three seconds before and 5 seconds after starting imagining a left- or right-hand movement (clue) over the C3, Cz and C4 electrodes [45].

In this specific example, a MI-based BCI should be able to reliably detect the ERD of Mu and Beta oscillations over the corresponding electrodes and take the appropriate action. As opposed to BCIs based on Evoked Potentials, the MI mental task is not elicited by an external stimulus but generated voluntarily by the BCI user. Therefore, the BCI user also has to learn to reliably reproduce the MI mental task, which can take up to several months for complex tasks such as three-dimensional movement control [16].

2.3 Signal processing

2.3.1 Band-pass filter

A Band-pass filter (BPF) strongly attenuates frequencies outside a selected frequency range and passes the frequencies within that range. The selected frequency range is defined by the cutoff frequencies ω_1 and ω_2 (with $\omega_1 < \omega_2$). If the signal to be filtered is not zero mean (has a DC offset) and that the 0 frequency is excluded from the BPF frequency range, the BPF will remove the DC offset and the resulting filtered signal will thus be zero mean. A commonly used BPF is the Butterworth filter [9].

2.3.2 Power Spectral Density

In order to quantify an ERD over specific frequency bands (e.g. for Motor Imagery) or to estimate the power at a particular frequency (e.g. for SSVEP), signal transformation from the time domain to the frequency domain is necessary. To this end, a common technique in EEG signal processing is to compute the Power Spectral Density (PSD) of a signal. The PSD represents the strength of the signal variations distributed in the frequency domain. Given $x(t)$ a signal, f the signal frequency in Hertz, $\omega = 2\pi f$ the angular frequency and $j = \sqrt{-1}$. The Fourier Transform decomposes a signal (time

domain) into its constituent sinusoids (frequency domain) and is defined as:

$$\mathcal{F}[x(t)] = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt \quad (2.1)$$

The Auto-Correlation Function estimate how a signal is correlated with a copy of itself shifted in time. It is used to detect periodicity in a signal and is defined as:

$$R(\tau) = x(\tau) * x(-\tau) = \int_{-\infty}^{\infty} x(t)x(t + \tau) dt \quad (2.2)$$

The Power Spectral Density is commonly defined as the Fourier Transform of the Auto-Correlation Function:

$$S_x(\omega) = \mathcal{F}[R(\tau)] = \int_{-\infty}^{\infty} R(\tau)e^{-j\omega\tau} d\tau \quad (2.3)$$

The spectral power of $x(t)$ in the frequency band $[f_1, f_2]$ is computed by integrating over the frequency range as follow:

$$P_{[f_1, f_2]} = 2 \int_{f_1}^{f_2} S_x(\omega) df \quad (2.4)$$

In practice, the PSD is estimated over a time window using Welch's method [48] that computes an estimate of the PSD by partitioning the signal in overlapping windows, forming a modified periodogram based on a Discrete Fourier Transform (DFT) on each window using specific frequencies and averaging the values of the periodogram.

2.3.3 Spatial filtering

Spatial filtering consists in using a linear combination of signals from the original electrodes (channels) in order to derive a smaller number of new channels. Spatial filtering thus reduces data dimensionality and increases signal-to-noise ratio (SNR).

Formally, let $x_i \in \mathbb{R}^C$ be the vector representation of the EEG signal at channel i , w_i be the weight of channel i in the spatial filter and $X \in \mathbb{R}^{E \times N}$ be a matrix which i^{th} row is x_i and \tilde{x} the spatially filtered signal. Then,

$$\tilde{x} = \sum_i w_i x_i = wX \quad (2.5)$$

Typical unsupervised spatial filters used in EEG signal processing include Principal Component Analysis (PCA) [3] or Independent Component Analysis (ICA) [1].

Common Spatial Pattern

Common Spatial Pattern (CSP) is an algorithm first introduced in EEG signal analysis by [53] and later applied to BCI applications. As of today, CSP is still considered a state-of-the-art technique for BCIs based on oscillatory activity (such as MI). The CSP

algorithm optimizes spatial filters in a supervised way in order to maximize the variance of the filtered signals for one class and minimize their variance for the other class.

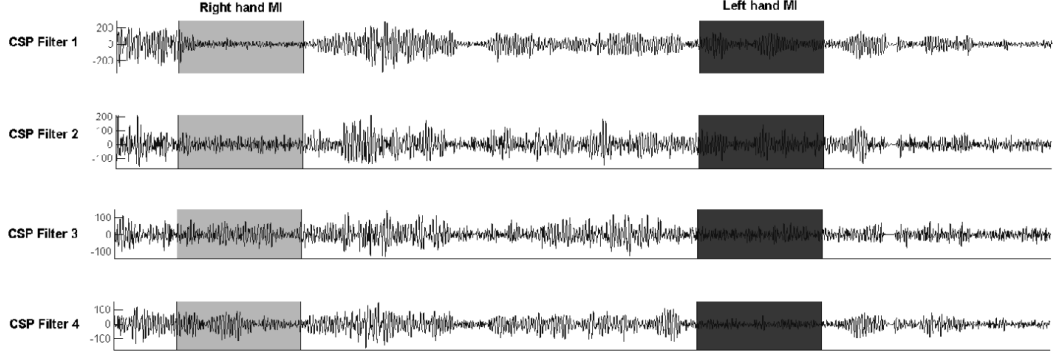


Figure 2.7: EEG signals spatially filtered using the CSP algorithm. CSP filters 1 and 2 maximize the variance of signals representing the imagination of a left-hand movement and minimize the variance of signals representing the imagination of a right-hand movement. Inversely for CSP filters 3 and 4. [28].

Formally, let w be the spatial filters used by the CSP algorithm, $X_y \in \mathbb{R}^{E \times E}$ be the matrix representation of an epoch from class $y \in Y$, $C_y \in \mathbb{R}^{E \times E}$ be the average covariance matrix of class y [8], wX_y be the spatially filtered signal from class y and thus $wX_yX_y^Tw^T$ be the variance of the spatially filtered signal from class y . Then, the CSP algorithm determines and uses w that maximizes and w that minimizes the function:

$$f(w) = \frac{wX_1X_1^Tw^T}{wX_2X_2^Tw^T} = \frac{wC_1w^T}{wC_2w^T} \quad (2.6)$$

As the variance of a signal band-pass filtered between the cutoff frequencies ω_1 and ω_2 represents the power of that signal in the $\omega_1 - \omega_2$ frequency range, the CSP algorithm actually optimizes spatial filters in order to produce maximal power difference between the two classes [28]. Thus, a common practice in BCIs based on MI is to use a BPF with cutoff frequencies at 8 and 30 Hz (covering the Mu and Beta frequency bands) on the raw EEG signal before applying the CSP algorithm [42].

2.3.4 Covariance matrices

As introduced in 2.3.3, the CSP algorithm requires a precise estimation of covariance matrices and their average in order to determine spatial filters w that yield optimal performances.

Let $X \in \mathbb{R}^{E \times N}$ be the matrix representation of an epoch and $N \in \mathbb{N}$ be the number of samples in a portion (epoch) of the EEG signal. Then, the usual covariance estimator is:

$$S = \frac{1}{N-1}X^TX \quad (2.7)$$

Let $S_y \in \mathbb{R}^{E \times E}$ be the covariance matrix from class $y \in Y$ and $|y|$ be the number of epochs of class y . Then, the average covariance matrix of class y , $C_y \in \mathbb{R}^{E \times E}$ is usually computed as follow:

$$C_y = \frac{1}{|y|} \sum S_y \quad (2.8)$$

Covariance matrices lie on a convex cone

Covariance matrices are symmetric positive definite (SPD) matrices (a proof is provided in appendix A).

Let $\mathcal{M}(n)$ be the set of $n \times n$ real matrices and $\mathcal{P}(n)$ be the space of $n \times n$ SPD matrices defined as:

$$\mathcal{P}(n) = \{A \in \mathcal{M}(n), A^T = A, x^T A x > 0, x \in \mathbb{N} \setminus \{0\}\} \quad (2.9)$$

$\mathcal{P}(n)$ is an open convex cone if, $\forall P, Q \in \mathcal{P}(n)$, $P + tQ$ is in $\mathcal{P}(n)$ $\forall t > 0$ [32]. Yet, if P and Q are SPD matrices, $P + tQ$ is also SPD $\forall t > 0$.

Thus, covariance matrices do not lie in a vector space but on a Riemannian manifold [32]. As illustrated in 2.8, the Riemannian geometry seems better suited than Euclidean geometry to manipulate covariance matrices in $\mathcal{P}(n)$.

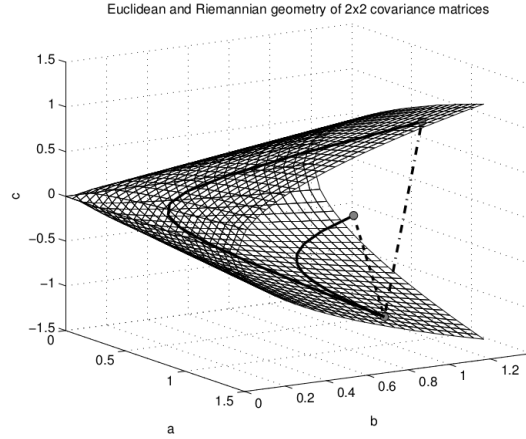


Figure 2.8: Euclidean (straight dashed lines) and Riemannian (curved solid lines) distances measured between points of $\mathcal{P}(2)$ [17].

2.4 Riemannian geometry

Before introducing the notions of Riemannian distance and mean, the fundamental concepts of topological spaces, manifolds and Riemannian differential manifolds have to be defined.

2.4.1 Topological spaces

The axiomatized definition of topological spaces from Hausdorff is used to propose a formal definition.

Let U be a set and $N(u)$ a function that maps a point $u \in U$ to a set of neighbourhoods of u . Then U is a so-called topological space if the following axioms are satisfied:

- The intersection of two neighbourhoods of u belongs to $N(u)$.
- For any set in $N(u)$, u belongs to it (u is located in its own neighbourhoods).
- Any subset of U that contains a neighbourhood of u , is itself a neighbourhood of u .
- For any neighbourhood $N_1 \in N(u)$, there exists another neighbourhood $N_2 \in N(u)$ such that N_1 is the neighbourhood of every point in N_2 .

2.4.2 Manifolds

A Hausdorff space is a topological space where two distinct points have disjoint neighbourhoods. A manifold is a Hausdorff space that can be approximated locally by an Euclidean space.

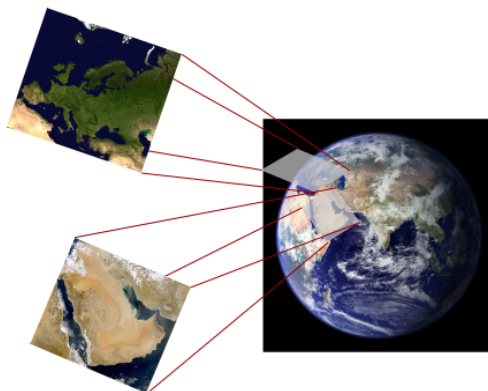


Figure 2.9: Earth can be seen as a Riemannian manifold and maps as the local linear approximation of the manifold [17].

Another example of Riemannian manifold are probability distributions. Each kind of probability distributions $\mathcal{N}(\mu, \sigma^2)$ can be seen as a manifold \mathcal{M} where the probability distribution's parameters μ and σ^2 act as coordinates in that topological space.

2.4.3 Riemannian differential manifolds

Riemannian manifolds are smooth manifolds (differentiable C^k -manifolds for which derivatives of all orders exist up to k) for which a scalar product is defined for any

point in the tangent space. Given a manifold \mathcal{M} and a point G , the tangent space is commonly denoted by $\mathcal{T}_G\mathcal{M}$ and represents the set of tangent vectors at point G in all directions.

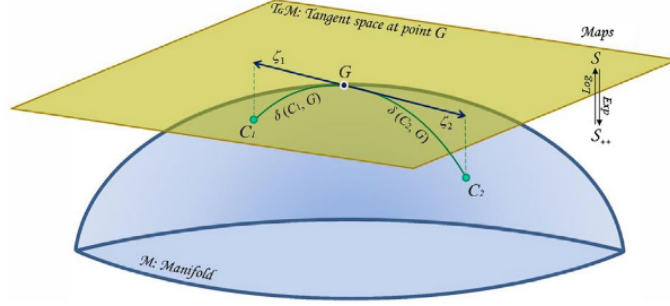


Figure 2.10: Schematic representation of the SPD manifold and the tangent space at point G [30].

A Riemannian metric is defined as an inner product of the form [2]:

$$\mathcal{T}_G\mathcal{M} \times \mathcal{T}_G\mathcal{M} \rightarrow \mathbb{R}, \quad G \in \mathcal{M} \quad (2.10)$$

and is a metric tensor capturing the lengths of two vectors as well as the angle between them. Note that the two points involved must be projected to a tangent space at a same point G .

2.4.4 Matrix logarithm and exponential

The exponential of a matrix $A \in \mathbb{R}^{E \times E}$ is expressed as the convergent series [32]:

$$\exp(A) = \sum_{k=0}^{\infty} \frac{1}{k!} A^k \quad (2.11)$$

Since the matrix logarithm of A is the solution to the equation $\expm(X) = A$, a series that converges to this solution is:

$$\log(A) = - \sum_{k=1}^{\infty} \frac{1}{k} (I - A)^k \quad (2.12)$$

where $I \in \mathbb{R}^{E \times E}$ is the identity matrix.

As suggested by [2], the matrix logarithm and exponential of a covariance matrix $S \in \mathbb{R}^{E \times E}$ estimated from a single epoch can be efficiently implemented by referring to the following formulas:

$$\log(\mathbf{S}) = \mathbf{V} \log(\mathbf{\Lambda}) \mathbf{V}^T \quad (2.13)$$

$$\exp(\mathbf{S}) = \mathbf{V} \exp(\mathbf{\Lambda}) \mathbf{V}^T \quad (2.14)$$

where $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues and \mathbf{V} is the matrix containing the eigenvectors of \mathbf{S} .

2.4.5 Riemannian distance

As showed in 2.3.4, covariance matrices are symmetric positive definite (SPD) and thus lie in a Riemannian manifold. Let $A \in \mathbb{R}^{E \times E}$ and $B \in \mathbb{R}^{E \times E}$ be two covariance matrices (two points on the manifold), the distance between A and B can be approximated using:

- the Euclidean distance δ_e (the straight dashed lines on Figure 2.8) such that

$$\delta_e(A, B) = \|A - B\|_{\mathcal{F}} \quad (2.15)$$

- the LogEuclidean distance δ_l such that

$$\delta_l(A, B) = \|\log(A) - \log(B)\|_{\mathcal{F}} \quad (2.16)$$

However, by taking into account the particular topology space of covariance matrices, we can use the corresponding Riemannian tool in order to compute an exact measure of the distance between A and B .

Let α be a curve on the manifold with parameter $t \in [0, 1]$ and $\dot{\alpha}(t)$ be the tangent vector $\alpha(t)$ of norm $\|\dot{\alpha}(t)\|_{\alpha(t)}$. The length of the curve from $\alpha(0)$ to $\alpha(1)$ is defined as:

$$L(\alpha) = \int_0^1 \|\dot{\alpha}(t)\|_{\alpha(t)} dt$$

Let the geodesic be the curve of minimum length $\Gamma = \underset{\alpha}{\operatorname{argmin}} L(\alpha)$ that represents the distance between two points on the manifold. Let Γ the geodesic with parameter t from $\Gamma(0) = A$ to $\Gamma(1) = B$ and let $\Gamma(t)$ a point on the geodesic, the Riemannian distance δ_r is computed along the geodesic and defined as:

$$\delta_r(A, B) = L(\Gamma) = \int_0^1 \|\dot{\Gamma}(t)\|_{\Gamma(t)} dt \quad (2.17)$$

In [17], the Riemannian distance $\delta_r(A, B)$ (the curved solid lines on Figure 2.8) is equivalently defined as:

$$\delta_r(A, B) = \|\log(A^{-\frac{1}{2}} B A^{-\frac{1}{2}})\|_{\mathcal{F}} \quad (2.18)$$

2.4.6 Projection to the tangent space, logarithmic and exponential maps

For each point of the manifold, there is an associated tangent space where a scalar product is defined and [2] showed that the euclidean distance in the tangent space is a good approximation of Riemannian distance on the manifold itself.

Let C_{ref} be the reference point (covariance matrix) on the manifold \mathcal{M} where the tangent plane is computed. A good choice of C_{ref} is proposed by [2] to be the geometric mean

of the whole set of covariance matrices and motivated by the observation from [46] that the geometric mean is the point where the mapping on the tangent space leads to the better local approximation of the manifold [2].

Let $A \in \mathbb{R}^{E \times E}$ and $B \in \mathbb{R}^{E \times E}$ be two covariance matrices, the scalar product in the tangent space at C_{ref} is defined as [2]:

$$\langle A, B \rangle_{C_{ref}} = tr (AC_{ref}^{-1}BC_{ref}^{-1}) \quad (2.19)$$

Let $logm$ and $expm$ denote respectively the matrix logarithm and matrix exponential, the explicit projection operators are given by:

$$\phi(C) = Log_{C_{ref}}(C) = C_{ref}^{1/2} logm(C_{ref}^{-1/2}C_pC_{ref}^{-1/2}) C_{ref}^{1/2} \quad (2.20)$$

$$\phi^{-1}(C) = Exp_{C_{ref}}(C) = C_{ref}^{-1/2} expm(C_{ref}^{1/2}C_pC_{ref}^{1/2}) C_{ref}^{-1/2} \quad (2.21)$$

Let k_r be a kernel function (or Riemannian metric) that computes the scalar product of two covariance matrices projected on the tangent space and let $logm(C_{ref}^{-1/2}SC_{ref}^{-1/2})$ be denoted as a new function $logm_{C_{ref}}(S)$ for simplicity.

$$k_R(A, B; C_{ref}) = \langle \phi(A), \phi(B) \rangle_{C_{ref}} \quad (2.22)$$

$$= tr [Log_{C_{ref}}(A) C_{ref}^{-1} Log_{C_{ref}}(B) C_{ref}^{-1}] \quad (2.23)$$

$$= tr [logm(C_{ref}^{-1/2}AC_{ref}^{-1/2}) logm(C_{ref}^{-1/2}BC_{ref}^{-1/2})] \quad (2.24)$$

$$= tr [logm_{C_{ref}}(A) logm_{C_{ref}}(B)] \quad (2.25)$$

The $logm_{C_{ref}}(C)$ operator returns a matrix of the same shape as C . Thus, the memory complexity is linear to the size of the training set if the computations are cached for all covariance matrices.

2.4.7 Riemannian mean

Similarly to the Riemannian distance, the Riemannian mean can be approximated using the Euclidean or LogEuclidean geometry. A noticable drawback of this method is that it induces a *swelling effect* [47], i.e. a distortion implied by the fact that the determinant of the average of two matrices can be bigger than both of their determinants [17].

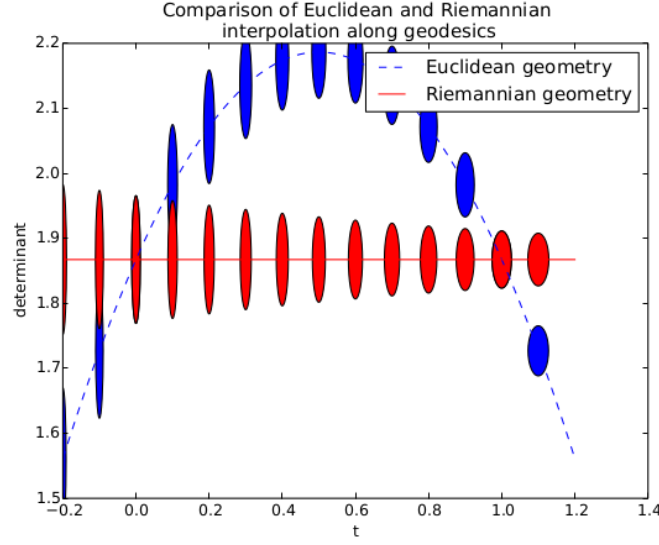


Figure 2.11: Swelling effect [51]

A proper way to compute the mean of covariance matrices without inducing a *swelling effect* is to define, by analogy to the Euclidean mean, the Riemannian mean (also named geometric mean or Frechet barycentre) of $I \geq 1$ SPD matrices as follows:

$$\mathcal{G}(\Sigma_1, \dots, \Sigma_I) = \underset{\Sigma \in P(n)}{\operatorname{argmin}} \sum_{i=1}^I \delta_R^2(\Sigma, \Sigma_i) \quad (2.26)$$

There exists no explicit formulation to compute the best reference matrix that minimizes the sum of the Riemannian distances to all training points on the manifold and the sum of euclidean distances on the tangent space.

Thus, we have to use the following iterative optimization procedure consisting in projecting all training covariance matrices onto the tangent plane based on the current reference matrix, computing their arithmetic mean, and finally replacing the current reference matrix by a new one by projecting the arithmetic mean back to the manifold. The iterative procedure described hereabove can be formalized as follows:

- $t \leftarrow 0$
- Initialize $\mathcal{R}^{(1)} = \frac{1}{P} \sum_{p=1}^P C_p$
- While $\|\bar{S}\|_F \geq \epsilon$
 - $\bar{S} = \frac{1}{P} \sum_{p=1}^P \operatorname{Log}_{\mathcal{R}^{(t)}}(C_p)$
 - $\mathcal{R}^{(t+1)} = \operatorname{Exp}_{\mathcal{B}^{(t)}}(\bar{S})$
 - $t \leftarrow t + 1$

Where $Log_{\mathcal{B}^{(t)}}$ and $Exp_{\mathcal{B}^{(t)}}$ are the logarithmic and exponential maps at iteration t , respectively.

Chapter 3

EEG classification for BCI applications

3.1 BCI protocol

In order to use a BCI application, two phases are necessary:

- an offline training phase when EEG data signals are gathered in a controlled environment following a strict acquisition protocol. This phase is essential to create a clean dataset without outliers that is be used to optimally calibrate the classifier.
- an online (operational) phase. During this second phase, the EEG signals from the user are being measured, preprocessed and relevant features are extracted. These EEG features are classified by the classifier calibrated in the offline phase and then translated into a specific command. Finally, a feedback (i.e. neurofeedback) is provided to the user that will inform him whether the desired command was recognized or not.

3.2 Classifiers

3.2.1 Support Vector Machine

Support Vector Machines (SVM) are linear models which training procedure consists in finding the maximum-margin hyperplane. The margin can be informally defined as the minimum distance between the optimal hyperplane and the closest training point. SVMs have been originally developped in the context of binary classification in order to achieve better generalization.

In the framework of binary classification, we assign label 1 to the samples belonging to the first class and label -1 to the rest of the samples. $y_i \in \{-1, 1\}$ is the label associated to sample x_i . Let's define $g(s)$ as a function returning 1 when s is positive and -1 otherwise. Given the weights w and bias b associated to the optimal hyperplane, the prediction function $h_{w,b}(x)$ can be defined as:

$$h_{w,b}(x_i) = g(w^T x_i + b) \quad (3.1)$$

where x is a new query point. The hyperplane is a decision boundary whose equation is given by $w^T x + b = 0$. Let's now define the functional margins of the model as the product $\hat{\gamma}_i = y_i(w^T x_i + b)$. This value should be maximized for each sample since $w^T x_i + b$ represents the confidence level of the model in the fact that sample x_i belongs to the first class, and $\hat{\gamma}_i$ is high when the model is both confident and right about the predicted class, given by $g(w^T x_i + b)$. However, maximizing $\hat{\gamma}_i \forall i$ does not prevent the model from scaling w to an arbitrary large value in order to maximize the functional margin. Therefore, the model should consider the *geometric margins* instead:

$$\hat{\gamma}_i = y_i \left(\left(\frac{w}{\|w\|} \right)^T x_i + \frac{b}{\|w\|} \right) \quad (3.2)$$

where w and b are now forced to have a norm of one. The geometric margin of the model with respect to the whole training set is the distance between the hyperplane and the closest point, given by $\gamma = \min_{i=1,\dots,m} \hat{\gamma}_i$. By constraining the model to have a unit length vector w and by constraining γ to be smaller than any geometric margin, we obtain the following formulation where γ must be maximized:

$$\max_{\gamma, w, b} \gamma \quad (3.3)$$

$$s.t. \quad y_i(w^T x_i + b) \geq \gamma \quad \forall i \in \{1, \dots, m\} \quad (3.4)$$

$$\|w\| = 1 \quad (3.5)$$

It can be noted that $\|w\| = 1$ is not a convex constraint and cannot be optimized in a very efficient way. This is why the previous formulation is not used in practice. However, by setting $\gamma = \hat{\gamma}/\|w\|$, arbitrary setting functional margin $\hat{\gamma}$ to 1, and by stating that maximizing $1/\|w\|$ reduces to minimizing $\|w\|^2$, we finally obtain the well-known primal formulation of **hard margin SVMs**:

$$\min_{w, b} \|w\|^2 \quad (3.6)$$

$$s.t. \quad y_i(w^T x_i + b) \geq 1 \quad \forall i \in \{1, \dots, m\} \quad (3.7)$$

This formulation is called hard margin because it does not allow any point of the training set to be misclassified. Obviously, nothing ensures that the training set is linearly separable and this is usually not the case in practice. **Soft margin SVMs** have been introduced in order to allow points to be on the wrong side of the hyperplane with respect to their classe. New variables ξ_i are introduced in order to somehow relax the constraints on the functional margins, but their values should be minimized:

$$\min_{w, b} \|w\|^2 + C \sum_{i=1}^m \xi_i \quad (3.8)$$

$$s.t. \quad y_i(w^T x_i + b) \geq 1 - \xi_i \quad \forall i \in \{1, \dots, m\} \quad (3.9)$$

$$\xi_i \geq 0 \quad \forall i \in \{1, \dots, m\} \quad (3.10)$$

Actually, this formulation is not used in practice either because its dual formulation allows to express the margins as functions of inner products of samples, which is much more interesting since it enables the representation of samples from the training set in a much higher dimensional space where it is easier to linearly separate them. This is exactly the role of kernel functions and the reason why Riemannian-based kernels were discussed in previous section. The dual formulation can be obtained either by applying the representer theorem or dualizing the constraints, introducing Lagrange multipliers α_i and using the KKT complementary conditions:

$$\max_{\alpha} \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle \quad (3.11)$$

$$s.t. \quad \alpha_i \geq 0 \quad (3.12)$$

$$\sum_{i=1}^m \alpha_i y_i = 0 \quad (3.13)$$

Introduction of the Riemannian kernel

The latest formulation represents the dual formulation of the “linear” SVM. Extending it to a “non-linear” SVM can simply be achieved by replacing the $\langle x_i, x_j \rangle$ product with the Riemannian kernel function $k_R(x_i, x_j; x_{ref})$ introduced in 2.4.6.

3.2.2 Convolutional Neural Network

A Convolutional Neural Network (CNN) is a biologically-inspired variant of feed-forward neural networks which has at least one convolutional layer.

A convolutional layer uses a convolution operator to map its input to an output in a way that emulates the animal visual cortex. Let $x_n \in \mathbb{R}^N$ be a 1D input signal, $h_m \in \mathbb{R}^M$ be a 1D filter. The output of the 1D convolution of x_n through h_m is given by:

$$y_n = x_n * h_n = \sum_{m=0}^{M-1} h_m x_{n-m} \quad \forall n = 0, \dots, N-1 \quad (3.14)$$

As of today, it is considered a state-of-the-art technique for image classification and is increasingly used in signal processing and classification.

A more thorough introduction to CNN will be carried out in the thesis.

Chapter 4

Preliminary classification tests

In order to compare the performances of the Riemannian approach with the non-Riemannian and Deep Learning approaches, two choices had been made:

- a large dataset from a Kaggle competition is used for the benchmark in order to ensure that the Deep Learning approach has enough training data to produce relevant results.
- the same SVM classifier (including tuning variables) is used to compare the Riemannian with the non-Riemannian approaches so the difference in performances should only be attributed to the type of kernel used (Riemannian or linear).

4.1 Problem description

The preliminary classification tests are based on a Kaggle competition from 2015 named “Grasp-and-Lift EEG Detection” that focuses on classifying hand motions of 6 different movements: ‘HandStart’, ‘FirstDigitTouch’, ‘BothStartLoadPhase’, ‘LiftOff’, ‘Replace’, and ‘BothReleased’.

4.2 Data

The study and data to support this competition was first published in Nature and later made publicly available for the sole purpose of the Kaggle competition.

The data, totalling 3,936 grasp and lift trials, was collected from twelve participants. The participant’s task in each trial was to reach for a small object, grasp it using their index finger and thumb, and lift it a few centimetres up in the air, hold it stably for a couple of seconds, and then replace and release the object. The beginning of the reach and the lowering of the object was cued by an LED, otherwise the pace of the task was up to the participant. During all trials, we recorded 32 channels of EEG, 5 channels of EMG from the shoulder, forearm, and hand muscles, the position of the arm, thumb and index finger and the object, and the forces applied to the object by the precision

grip. We defined 16 behaviourally relevant events and extracted them for every trial.” [31]

4.3 Methodology

4.3.1 Validation

In order to perform an unbiased estimation of the model prediction accuracy and its ability to generalize, the model has to be evaluated using data that are completely separated from the model’s learning process. Therefore, as a rule of thumb, the dataset is originally split into a training set (used in the learning phase) and a validation set (used to evaluate the selected model) with a proportion of 75% and 25% respectively. Even if EEG signals are nonstationary, several factors such as external influences or the subject physiological state may have an undesirable positive influence on a classifier and lead to over-optimistic metrics if two samples from the training set and validation set were very close in the original dataset. In order to avoid this problem, from the 8 separate sessions of the subject, we selected 6 sessions to be in the training set and 2 sessions to be in the validation set.

4.3.2 Evaluation metrics

In the domain of EEG signal classification, as in many classification problems in the medical field, datasets are usually strongly unbalanced in the favor of one specific class. With 80.71% of samples with no motion at all, the hand-motion dataset used in this work is no exception to the rule. Under these constraints, the mean accuracy is not the most adequate measure to evaluate a model’s performances as it would result in highly over-optimistic metrics that would not reflect the true performances of the classifier in real-life conditions.

Therefore, the metric used to evaluate the performance of our models is the Area Under the Receiver Operating Characteristic Curve (AUROCC). The Receiver Operating Characteristic (ROC) curve plots the True Positive (TP) rate (also called Sensitivity or Recall) on the Y axis against the False Positive (FP) rate (1-Specificity) on the X axis. A random classifier has an AUROCC of 0.5.

4.4 Models

4.4.1 PSD with Logistic Regression

This model is used to have a baseline representing a valid but simple approach to the problem but is not expected to yield better results than the other implemented methods.

The PSD feature extraction for this approach consists of the following steps. The PSD is estimated on the whole 7-second epoch using Welch’s method. The selected windows size and overlap are 500 samples and 25% respectively. For each of the 32 electrodes, the corresponding PSD is subsequently divided in 5 different frequency bands in each

of which the mean is computed to form a binned spectrogram of dimension 5 bands \times 1 epoch.

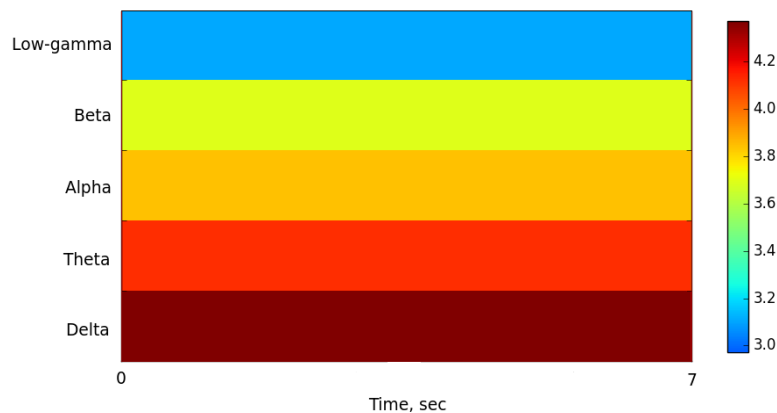


Figure 4.1: Binned spectrogram computed on a 7-second epoch

A \log_{10} is eventually applied on the 32 spectrograms. The combined spectrograms from each electrode are flattened into a one-dimensional array of 160 values to form an input to the Logistic Regression classifier.

4.4.2 Deep Learning on raw EEG data

This approach uses a CNN trained with windows of 7 seconds of raw EEG data. The values of the convolutional kernels and weights of the dense layers are initialized following the Glorot uniform distribution (Glorot and Bengio, 2010). The rectified linear unit (ReLU) (Nair and Hinton, 2010) was used as an activation function for all the layers of the CNN except for layer 1 and the output layer.

Network architecture

The following network architecture was derived from different architectures published on the Kaggle website as well as empirically.

The input to the neural network consists of a 32 channels \times 3584 raw EEG signal (7 seconds)

Layer 1 apply a linear transformation (no activation function used) using a 1D spatial convolution over the channels to condense the 32 channels into 4 convolutional filters (thus reducing the input dimension). This operation is similar to the spatial filtering introduced in 2.3.3.

Layer 2 and 3 are standard 1D temporal convolution over the condensed signal of each channel. A ReLu activation function and Max pooling operation are applied to the output of these layers. Moreover, a Dropout with probability 0.5 is used on the output of layer 3.

Layers 4 and 5 are standard Fully Connected layers with 1024 neurons each separated by a Dropout layer.

Layer 6 is the output layer. Sigmoid cross entropy was used to output the action probability because the possible actions were not mutually exclusive in the dataset.

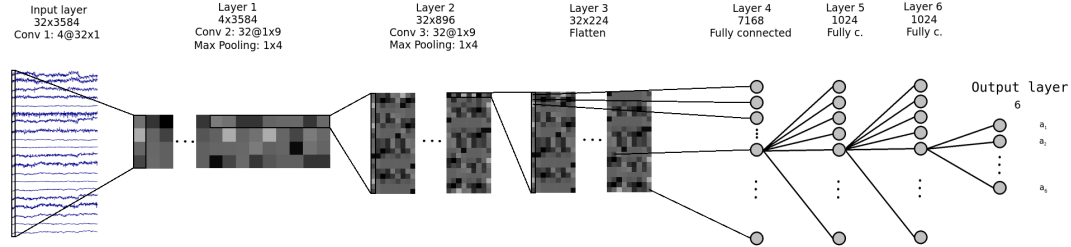


Figure 4.2: Convolutional Neural Network architecture

4.4.3 Linear and Riemannian based SVM

In this approach, covariance matrices are extracted from the training and validation sets at regular intervals. Then, the Riemannian (geometric) mean is computed from all the covariance matrices.

Two different Support Vector Machines are trained, one with a linear kernel and the other with a Riemannian kernel in order to objectively quantify the gain, if any, when using a Riemannian kernel function.

4.5 Results

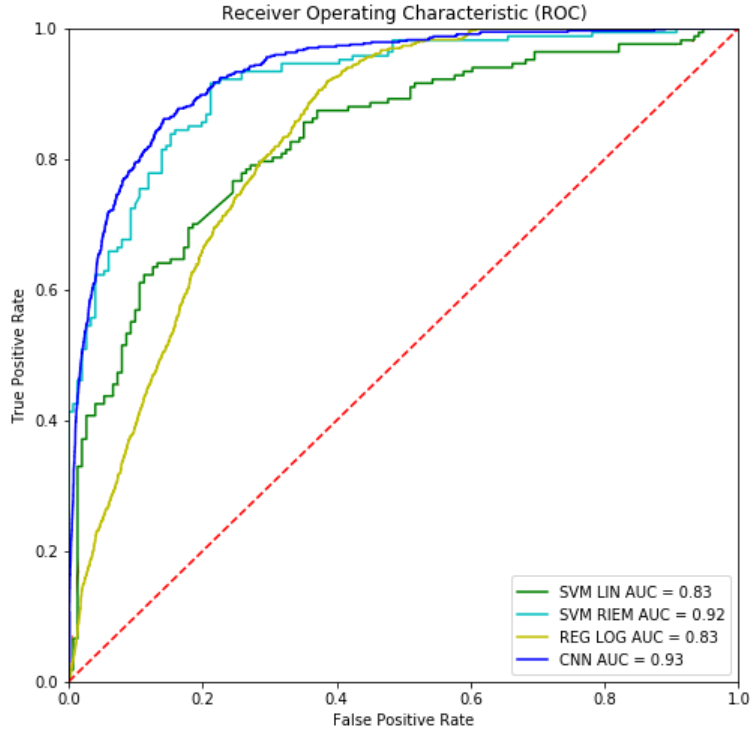


Figure 4.3: Plot of the Receiver Operating Characteristic (ROC) curves with the Area Under the Receiver Operating Characteristic Curve (AUROCC) obtained by the different models (i.e., from top to bottom, Linear SVM, Riemannian SVM, Logistic Regression and CNN).

4.6 Discussion

First, we clearly see that two models, SVM with a Riemannian kernel and CNN, achieved significantly higher performances, 0.92 and 0.93 AUC respectively, than the other models. Furthermore, as expected, the SVM with a Riemannian kernel significantly outperformed the SVM with the linear kernel, 0.92 and 0.83 AUC respectively. These results, as well as the first place obtained by models based on Riemannian geometry in 5 recent international BCI competitions, confirm that the Riemannian geometry is a critical tool to achieve state-of-the-art results in EEG signal classification.

Although the results displayed hereabove look very promising and correspond to single-model performances reported by top Kaggle competitors in 2015, caution is however advised because important features selected by the classifiers were reported to include frontal electrodes that are known to contain less information about the movement but are more affected by EMG artefacts. Thus, the results achieved in the framework of this competition might not be reproducible for a BCI based on motor imagery.

Conclusion

The main purpose of this work was to describe succinctly the state of the art in non-invasive electroencephalographic signal processing, types BCIs and the Machine Learning algorithms applied in this domain.

Next year, the Master thesis will use the knowledge and results obtained in this work in order to initiate the development of a functioning state-of-the-art BCI system and obtain sufficiently encouraging results that can hopefully lead to a PhD thesis

Appendix A

Covariance matrices are symmetric positive definite

Let $x_i = (x_{i1}, \dots, x_{im})^T$ be a sequence of raw samples in a fixed time window, where $i \in \{1, \dots, n\}$ is a time step, m is the number of electrodes and n is the number of samples within the time step (depending on the sampling frequency). Let μ_x be the vector of empirical means of those samples defined as follows:

$$\mu_x = \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{A.1})$$

where the size of μ_x is equal to m and element μ_{x_i} is the empirical mean of samples from electrodes i . Then the corresponding SCM Σ with respect to the raw EEG sample is computed using the following formula:

$$\Sigma = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_x) (x_i - \mu_x)^T \quad (\text{A.2})$$

More specifically, Σ_{ij} (i.e. the covariance between electrodes i and j) is computed using $\frac{1}{n} \sum_{k=1}^n (x_{ki} - \mu_{x_i}) (x_{kj} - \mu_{x_j})^T$.

For Σ to be positive definite, the scalar $z^T \Sigma z$ must be strictly positive for any non-zero vector $z \in \mathbb{R}^m$. We have:

$$z^T \Sigma z = z^T \left(\frac{1}{n} \sum_{i=1}^n (x_i - \mu_x) (x_i - \mu_x)^T \right) z \quad (\text{A.3})$$

$$= \frac{1}{n} \sum_{i=1}^n z^T (x_i - \mu_x) (x_i - \mu_x)^T z \quad (\text{A.4})$$

$$= \frac{1}{n} \sum_{i=1}^n \left((x_i - \mu_x)^T z \right)^T \left((x_i - \mu_x)^T z \right) \quad (\text{A.5})$$

$$= \frac{1}{n} \sum_{i=1}^n \left((x_i - \mu_x)^T z \right)^2 \geq 0 \quad (\text{A.6})$$

This result shows that matrix Σ is semi-positive definite. Furthermore, in the case of EEG data samples, it is reasonable to assume that samples values are not constant and that product $(x_i - \mu_x)^T z$ is different from zero for at least one time step. Thus confirming that Σ is definite positive.

Moreover, the symmetric property can also be derived from the definition of Σ as follows:

$$\Sigma^T = \left(\frac{1}{n} \sum_{i=1}^n (x_i - \mu_x) (x_i - \mu_x)^T \right)^T \quad (\text{A.7})$$

$$= \frac{1}{n} \sum_{i=1}^n \left((x_i - \mu_x) (x_i - \mu_x)^T \right)^T \quad (\text{A.8})$$

$$= \frac{1}{n} \sum_{i=1}^n \left((x_i - \mu_x)^T \right)^T \left((x_i - \mu_x) \right)^T \quad (\text{A.9})$$

$$= \frac{1}{n} \sum_{i=1}^n (x_i - \mu_x) (x_i - \mu_x)^T \quad (\text{A.10})$$

$$= \Sigma \quad (\text{A.11})$$

SCMs are thus both symmetric and positive definite.

Bibliography

- [1] L. Senhadji A. Kachenoura, L. Albera and P. Comon. Ica: A potential tool for bci systems. *IEEE Signal Processing Magazine*, 25(1):57–68, 2008.
- [2] Marco Congedo Christian Jutten Alexandre Barachant, Stphane Bonnet. Classification of covariance matrices using a riemannian-based kernel for bci applications. *Neurocomputing*, 112:172–178, 2013. hal-00820475.
- [3] Waldemar Ratajczak Andrzej Mackiewicz. Principal components analysis (pca). *Computers and Geosciences*, 19(3):303–342, 1993.
- [4] Fetz E. E. Shupe L. Lettich E. Aoki, F. and G. A. Ojemann. Increased gamma-range activity in human sensorimotor cortex during performance of visuomotor tasks. *Clinical Neurophysiology*, 110:524–537, 1999.
- [5] Del Percio C. Vecchio F. Sebastiano F. Di Gennaro G. Quarato P. P. et al. Babiloni, C. Alpha, beta and gamma electrocorticographic rhythms in somatosensory, motor, premotor and prefrontal cortical areas differ in movement execution and observation in humans. *Clinical Neurophysiology*, 127:641–654, 2016.
- [6] H. Berger. ber das elektrenkephalogramm des menschen. *Arch. Psychiatr. Nervenkr.*, 87:527–570, 1929.
- [7] Faraone SV. Biederman J. Attention-deficit hyperactivity disorder. *The Lancet*, 366:237–248, 2005.
- [8] Lemm S Kawanabe M Mller KR Blankertz B, Tomioka R. Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal Processing Magazine*, 25(1):41–56, 2008b.
- [9] S. Butterworth. On the theory of filter amplifiers. *Wireless Engineer*, 7:536–541, 1930.
- [10] A. M. Cebolla and G. Cheron. Sensorimotor and cognitive involvement of the beta-gamma oscillation in the frontal n30 component of somatosensory evoked potentials. *Neuropsychologia*, 79:215–222, 2015.
- [11] Zarka D. Hoellinger T. Leroy A. Dan B. Cevallos, C. and G. Cheron. Oscillations in the human brain during walking execution, imagination and observation. *Neuropsychologia*, 79:223–232, 2015.
- [12] Fabien Lotte Chang S. Nam, Anton Nijholt. *L^AT_EX: Brain-Computer Interfaces Handbook: Technological and Theoretical Advances*. CRC Press, first edition, 2018.

- [13] Bosnyak D. Manning F. C. Spinelli C. Marie C. Fujioka T. et al. Cirelli, L. K. Beat-induced fluctuations in auditory cortical beta-band activity: using eeg to measure age-related changes. *Audit. Cogn. Neurosci.*, 5:742, 2014.
- [14] M. X. Cohen. *Analyzing Neural Time Series Data: Theory and Practice*. The MIT Press, 2014.
- [15] Jonathan R. Wolpaw Dean J. Krusienski. Chapter 11 brain-computer interface research at the wadsworth center: Developments in noninvasive communication and control. *International Review of Neurobiology*, 86:147–157, 2009.
- [16] W.A. Sarnacki D.J. McFarland and J.R. Wolpaw. Electroencephalographic (eeg) control of three-dimensional movement. *Journal of Neural Engineering*, 7, 2010.
- [17] F. Lotte F. Yger and M. Sugiyama. Averaging covariance matrices for eeg signal classification based on the csp: An empirical study. *European Signal Processing Conference (EUSIPCO)*, pages 2721–2725, 2015.
- [18] L. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70:510–523, 1988.
- [19] Guger C Sellers EW Kleih SC Kbler A Fazel-Rezai R, Allison BZ. P300 brain computer interface: current challenges and emerging trends. *Frontiers in Neuro-engineering*, 5:14, 2012.
- [20] Pierre Ferrez and Jos del R. Milln. *Error-related EEG potentials in brain-computer interfaces*. PhD thesis, EPFL, 2007.
- [21] H. J. Gastaut and J. Bert. Eeg changes during cinematographic presentation; moving picture activation of the eeg. *Electroencephalography and Clinical Neurophysiology*, 6:433–444, 1954.
- [22] Keil A. Gruber, T. and M. M. Mller. Modulation of induced gamma band responses and phase synchrony in a paired associate learning task in the human eeg. *Neuroscience Letters*, 316:29–32, 2001.
- [23] Mller M. M. Gruber, T. and A. Keil. Modulation of induced gamma band responses in a perceptual learning task in the human eeg. *Cognitive Neuroscience*, 14:732–744, 2002.
- [24] T. Gruber and M. M. Mller. Oscillatory brain activity in the human eeg during indirect and direct memory tasks. *Brain Research*, 1097:194–204, 2006.
- [25] Yuval-Greenberg S. Keren, A. S. and L. Y. Deouell. Saccadic spike potentials in gamma-band eeg: characterization, detection and suppression. *NeuroImage*, 49:2248–2263, 2010.
- [26] Karmos G.-Mehta A. D. Ulbert I. Lakatos, P. and C. E. Schroeder. Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, 320:110–113, 2008.
- [27] Bottou L.-Bengio Y. Lecun, Y. and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86:2278–2324, 1998.

- [28] F. Lotte. A tutorial on eeg signal processing techniques for mental state recognition in brain-computer interfaces. 2014. hal-01055103.
- [29] Ripper B.-Busse L. Birbaumer N. Lutzenberger, W. and J. Kaiser. Dynamics of gamma-band activity during an audiospatial working memory task in humans. *Journal of Neuroscience*, 22:5630–5638, 2002.
- [30] Alexandre Barachant Marco Congedo and Rajendra Bhatia. Riemannian geometry for eeg-based brain-computer interfaces; a primer and a review. *Brain-Computer Interfaces*, 4(3):155–174, 2017.
- [31] Ewa Jarocka Matthew D Luciw and Benoni B Edin. Multi-channel eeg recordings during 3,936 grasp and lift trials with varying weight and friction. *Scientific Data (Nature)*, 1, 2014. Article number: 140047.
- [32] M. Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *Journal on Matrix Analysis and Applications*, 26(3):735–747, 2005.
- [33] B. Blankertz M.S. Treder. (c)overt attention and visual speller design in an erp-based brain-computer interface. *Behavioral and Brain Functions*, 6:28, 2010.
- [34] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27 th International Conference on Machine Learning*, 1:807–814, 2010.
- [35] Federal Standard 1037C: Glossary of Telecommunications Terms. Band-stop filter. <https://www.its.bldrdoc.gov/fs-1037/fs-1037c.htm>, last visited on August 17, 2018.
- [36] Coombes S. A. Ofori, E. and D. E. Vaillancourt. 3d cortical electrophysiology of ballistic upper limb movement in humans. *NeuroImage*, 115:30–41, 2015.
- [37] Peter Oostenveld, Robert; Praamstra. The five percent electrode system for high-resolution eeg and erp measurements. *Clinical Neurophysiology*, 112:713–719, 2001.
- [38] S. Briskin J.R. Wolpaw H. Bischof G. Schalk P. Brunner, S. Joshi. Does the 'p300' speller depend on eye gaze? *Journal of Neural Engineering*, 7(5):056013, 2010.
- [39] D. Plass-Oude Bos. Automated artifact detection in brainstream; an evaluation of an online eye and muscle artifact detection method. 2012.
- [40] JB Plum, F; Posner. The diagnosis of stupor and coma. *F.A. Davis*, page 197, 1966.
- [41] A. Rakotomamonjy and V. Guigue. Bci competition iii: Dataset ii- ensemble of svms for bci p300 speller. *IEEE Transactions on Biomedical Engineering*, 55(3):1147–1154, March 2008.
- [42] Pfurtscheller G Ramoser H, Muller-Gerking J. Optimal spatial filtering of single trial eeg during imagined hand movement. *IEEE Transactions on Rehabilitation Engineering*, 8(4):441–446, 2000.
- [43] M. Steriade. Grouping of brain rhythms in corticothalamic systems. *Neuroscience*, 137:1087–1106, 2006.

- [44] Zubin J. Sutton S., Tueting P. and John E. Information delivery and the sensory evoked potential. *Science*, 155:1436–1439, 1967.
- [45] Li Chao Tang, Zhichuan and Shouqian Sun. Single-trial eeg classification of motor imagery using deep convolutional neural networks. *International Journal for Light and Electron Optics*, 130, 2016.
- [46] Porikli F. Meer P. Tuzel, O. Pedestrian detection via classification on riemannian manifolds. *Pattern Analysis and Machine Intelligence*, 30:1713–1727, 2008.
- [47] X. Pennec V. Arsigny, P. Fillard and N. Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *Journal on Matrix Analysis and Applications*, 2007.
- [48] P. D. Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15:70–73, 1967.
- [49] K. Whittingstall and N. K. Logothetis. Frequency-band coupling in surface eeg reflects spiking activity in monkey visual cortex. *Neuron*, 64:281–289, 2009.
- [50] Yoshua Bengio Xavier Glorot. Understanding the difficulty of training deep feed-forward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 9:249–256, 2010.
- [51] M. Berar Yger and F. Lotte. Riemannian approaches in brain-computer interfaces: A review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10):1753–1762, 2017.
- [52] Cevallos C. Petieau M. Hoellinger T. Dan B. Zarka, D. and G. Cheron. Neural rhythmic symphony of human walking observation: upside-down and uncoordinated condition on cortical theta, alpha, beta and gamma oscillations. *Frontiers in Systems Neuroscience*, 8:169, 2014.
- [53] Koles ZJ. The quantitative extraction and topographic mapping of the abnormal components in the clinical eeg. *Electroencephalography and Clinical Neurophysiology*, 79:440–447, 1991.