

Master Thesis

# Non Linear Optimised Spatial Filter for Single-Trial Motor Related Cortical Potential Identification

Mascolini Alessio

Supervisor

Luca Mesin

Master thesis in Biomedical Engineering



Dipartimento di Elettronica e Telecomunicazioni  
Politecnico di Torino  
Italia, Torino  
Ottobre 2019



# **Non Linear Optimised Spatial Filter for Single-Trial Motor Related Cortical Potential Identification**

**Mascolini Alessio**

Supervisor:

**Luca Mesin**

## **Abstract**

The Motor Related Cortical Potential is a low frequency negative shift in the EEG signal appearing around 2 seconds before a planned or executed voluntary movement. Its detection is instrumental in the development of Brain Computer Interfaces which allow patients who are otherwise unable to do so to communicate, as well as in neurorehabilitation of patients with motor impairments. An improvement in accuracy of the detectors could cause a significant advancement in the field of neuroprosthetics.

In this thesis we develop a new technique to identify these potentials from EEG recordings of motor execution by calculating the optimal non linear combination of channels which isolates the signal we require. The method shows promising results, fast execution and requires a small amount of training examples, enabling training on a patient by patient basis to take into account the natural individual variability in how MRCs manifest themselves, and making the system agnostic to the number of channels and placement of electrodes.



# Contents

<b>1</b>	<b>Introduction</b>	<b>13</b>
1.1	What is Electroencephalography . . . . .	13
1.1.1	Electric Field . . . . .	13
1.1.2	Brain Areas . . . . .	13
1.1.3	Frequency Bands . . . . .	14
1.1.4	The 10-20 Measurement System . . . . .	15
1.2	Brain Computer Interfaces . . . . .	16
1.2.1	Techniques . . . . .	16
1.2.2	P300 . . . . .	17
	Odd-ball Paradigm . . . . .	17
	Usage in BCIs . . . . .	17
1.2.3	VEP and SSVEP . . . . .	18
	Usage in BCIs . . . . .	18
	t-VEPs . . . . .	18
	f-VEPs . . . . .	18
	c-VEPs . . . . .	18
1.2.4	SCP . . . . .	19
	Usage in BCIs . . . . .	19
1.2.5	Limitations . . . . .	19
1.3	Motor Related Cortical Potentials . . . . .	20
<b>2</b>	<b>Mathematical Concepts</b>	<b>21</b>
2.1	PCA . . . . .	22
2.1.1	Eigenvalues and Eigenvectors . . . . .	22
2.1.2	Eigenvectors of the Covariance Matrix . . . . .	22
2.1.3	Principal Component Analysis . . . . .	23
2.2	Beyond PCA . . . . .	23
2.3	Singular Value Decomposition . . . . .	24
2.4	Manifolds . . . . .	24
2.4.1	Hilbert Spaces . . . . .	24
2.4.2	Riemann Manifolds . . . . .	25
	Mapping To The Tangent Space . . . . .	25
<b>3</b>	<b>State of the art</b>	<b>27</b>
3.1	Artifact Removal . . . . .	27
3.2	Processing . . . . .	28
3.2.1	Reviews . . . . .	28
3.2.2	Overview of the Methods . . . . .	28

3.3	Reproduced Methods . . . . .	29
3.3.1	Optimal Spatial Filter . . . . .	29
3.3.2	LPP-LDA . . . . .	30
	LPP . . . . .	30
	LDA . . . . .	30
3.3.3	ARK-SVM . . . . .	31
	SPD Riemann Manifold . . . . .	31
	Classification . . . . .	31
	Geometric Median . . . . .	31
3.3.4	Final Observations . . . . .	31
<b>4</b>	<b>Methods</b>	<b>33</b>
4.1	Dataset . . . . .	33
4.2	Detrending . . . . .	33
4.3	Artifact removal . . . . .	34
4.4	Resampling . . . . .	34
4.5	Non-linear optimal spatial filter . . . . .	35
4.5.1	Intuition . . . . .	35
4.5.2	Ordinary Least Squares . . . . .	35
4.5.3	Source Estimation . . . . .	36
4.5.4	Matrix Inversion . . . . .	36
4.5.5	Underdetermined Case . . . . .	37
4.5.6	OLS Assumptions . . . . .	37
	The residuals should have zero conditional mean . . . . .	37
	The predictors should be linearly independent . . . . .	37
	The residuals should be spherical . . . . .	38
4.5.7	Whitening Transformation and SVD . . . . .	38
4.5.8	Learning Non-Linear Mappings . . . . .	39
	Kernel Trick . . . . .	39
	Random Fourier Features . . . . .	39
<b>5</b>	<b>Results</b>	<b>41</b>
5.1	Models . . . . .	41
5.1.1	State of the art . . . . .	41
5.1.2	New Models . . . . .	41
5.2	Segmentation . . . . .	41
5.2.1	Preprocessing . . . . .	42
5.3	Metrics . . . . .	42
5.4	Simulated EEG Traces . . . . .	42
5.4.1	Simulator . . . . .	42
5.4.2	Results . . . . .	43
5.5	Real EEG Traces . . . . .	45
5.5.1	Dataset . . . . .	45
5.5.2	Extracted MRCPs . . . . .	46
5.5.3	Results . . . . .	47
<b>6</b>	<b>Conclusions</b>	<b>49</b>
<b>A</b>	<b>Appendix</b>	<b>51</b>

A.1	Flow Graph of the NL-OSF algorithm and additional figures . . . . .	51
A.2	Matlab code for NL-OSF . . . . .	58
A.3	Helper Matlab class for working in the SPD matrix manifold . . . . .	58



# List of Figures

1.1	Brain regions and areas by task . . . . .	14
1.2	Electrode placement in the 10-20 measurement system . . . . .	16
1.3	MRCPs of a healthy subject for real and imaginary right ankle dorsi- flexion. Each wave is an average of 50 large Laplacian spatial filtered EEG trials . . . . .	20
2.1	Directions of highest variance for a gaussian distribution . . . . .	22
2.2	Exponential map of a sphere . . . . .	25
4.1	Mapping the predictors in a higher dimensional space makes separat- ing two classes which were previously nonseparable possible . . . . .	39
5.1	Simulated MRCP signal . . . . .	42
5.2	Performance of the models on simulated data as a function of SNR, average of 4 samples per SNR level . . . . .	44
5.3	Smoothed performance of the models on simulated data as a function of SNR, average of 4 samples per SNR level plus an additional sliding window average of 4 samples . . . . .	44
5.4	Comparison of the extracted MRCPs for the OSF algorithm on the testing set of Patient 1 . . . . .	46
A.5	Mean and Standard Error of MRCPs in the testing set . . . . .	57
A.6	Mean and Standard Error of MRCPs in the testing set after NL-OSF . . . . .	57

# List of Tables

1.1	EEG frequency bands . . . . .	14
1.2	Bandwidth of various BCI techniques . . . . .	17
5.1	Results on Real EEG Data . . . . .	47



*Thanks to my family for supporting me.*

*To my friends who made these college years the best I could ask for.*

*And to my supervisor who has always been available to help me.*



# Chapter 1

## Introduction

### 1.1 What is Electroencephalography

The brain is a complex organ yet to be fully understood. Its main functional unit, the neuron, is a cell specialised in generating and conducting electrical signals along the body, interacting with synapses which in addition to allowing communication between neurons provide a way for signals to mix and combine.

#### 1.1.1 Electric Field

Since the functionality of neurons depends on ionic currents generated inside the channels, brain activity causes an electric field which can be measured on the scalp. The recording of such potential is named electroencephalogram and is used to diagnose and monitor epilepsy, tumors, sleep disorders and various kinds of brain diseases such as Alzheimer's [1]. In some cases the field is measured more invasively directly on the brain cortex; in this case it is more appropriately defined as electrocorticogram.

The maximum amplitude of EEG potentials can be around 100  $\mu\text{V}$  which means EEG signals are difficult to record, requiring expensive equipment as well as being sensitive to artifacts and affected by extensive amounts of noise. The EEG bandwidth ranges from 1Hz to 50Hz: the most amount of information is typically around 30Hz, but the frequency range of interest can reach above 100Hz in specific applications.

#### 1.1.2 Brain Areas

From a physiological point of view the brain is the primary center of social and intellectual life to which stimuli end up after modulation from the receptors and from which impulses arise. Voluntary responses originate from the cortex, the area in which most high level reasoning occurs, reaching muscles and peripheral organs through efferent neurons.

Despite being a single organ, the brain shows several regions which serve different

purposes: visual, olfactory and auditory stimuli are processed by different areas as well as language, motor generation, motor control and high level reasoning.

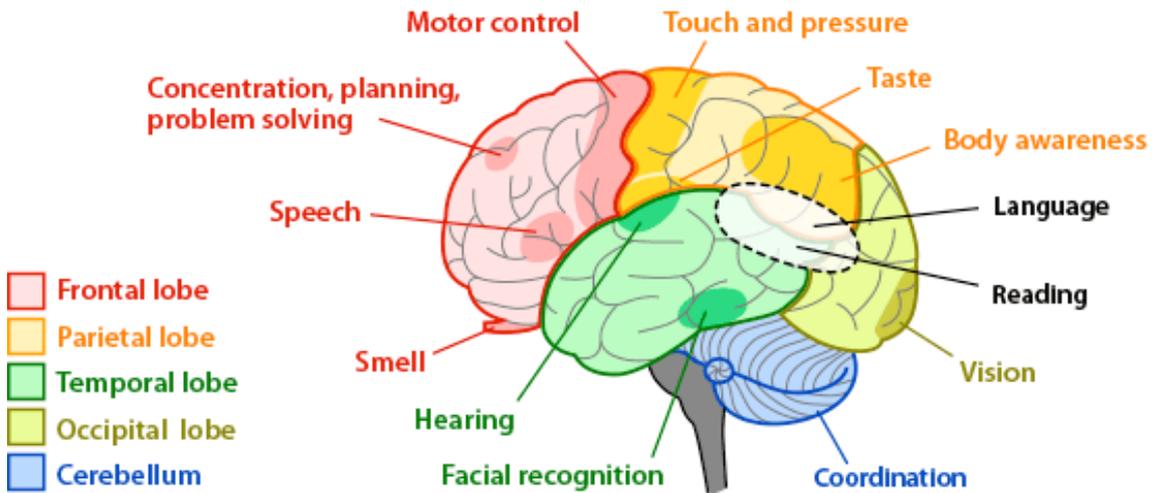


Figure 1.1: Brain regions and areas by task

### 1.1.3 Frequency Bands

Given the spontaneous voltage fluctuations and the short time constants, EEG signals are seemingly stochastic. However, they are composed of quasi sinusoidal patterns that are characterised by different frequency bands. The EEG spectrum can be divided into the frequency bands indicated in Table 1.1:

	Frequency Range
<b>Delta</b>	1-4 Hz
<b>Theta</b>	4-8 Hz
<b>Alpha</b>	8-13 Hz
<b>Beta</b>	13-30 Hz
Gamma	>30 Hz

Table 1.1: EEG frequency bands

#### Delta

Delta waves have the highest amplitude among the ones listed, in the range of about 75-200 micro-volts. They usually appear in specific stages of sleep, and are common in infants up to 12 months. The presence of delta rhythm in the waking adult EEG indicates cerebral injury or severe cerebral disease.

#### Theta

These waves are usually present in children up to the age of 12-13. Their presence in awake adults is seen as a manifestation of focal subcortical lesions.

## Alpha

This rhythm is observed as the patient has their eyes closed, commonly associated with relaxation these waves are severely attenuated when the eyes are open and when the brain is in an alerted state. It is the major rhythm seen in normal relaxed adults, and is best seen in the posterior regions of the head on each side, being higher in amplitude on the dominant side.

## Beta

Beta wave activity is usually found in the frontal region of the head. It can be registered as the dominant rhythm in individuals who are alert and attentive. It may be absent or reduced in areas in which cortical damage is present

## Gamma

Gamma wave components are normally either not recorded by EEGs or filtered out.

### 1.1.4 The 10-20 Measurement System

In order to obtain standardised and reproducible results from EEG recordings, the 10-20 measurement system has been developed.

The 10-20 system is an international standard regarding the placement of electrodes for EEG recordings, in which, using two reference points: the *nasion*, the ridge between nose and forehead, and the *inion*, the protrusion on the back of the head along the line drawn by the ears, electrodes are arranged on a grid with either 10% or 20% of the total length or width of the recording area. In this thesis, when referring to EEG recordings, the 10-20 system will always be used. Positions and labels of the electrodes are shown in Figure 1.2

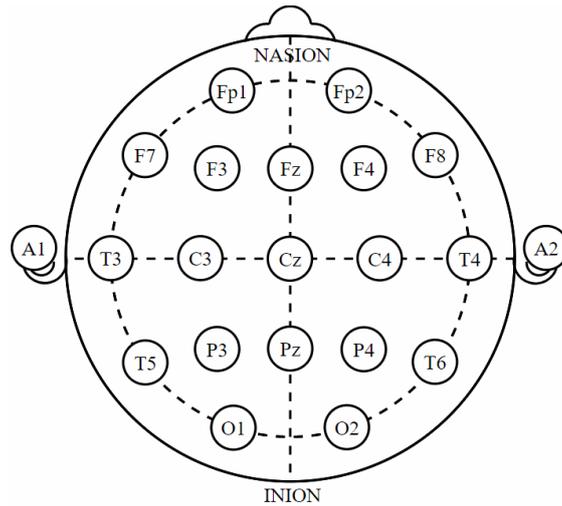


Figure 1.2: Electrode placement in the 10-20 measurement system

## 1.2 Brain Computer Interfaces

Brain Computer Interfaces (BCI) are a relatively recent subject of research, with the first paper on the topic published in 1973 [2]. The term BCI encompasses multiple types of techniques to allow machine-brain communication, which are helpful for patients with conditions which do not allow them to communicate with the external world such as locked-in syndrome, amyotrophic lateral sclerosis and cerebral palsy.

This kind of assistive technology gives these patients the ability to communicate, providing a significant improvement of their quality of life.

Nonetheless BCIs today are affected by a significant issue. Normally the surrounding environment is important as it provides feedback allowing the brain to automatically perfect the actions required to achieve a goal. Lack of touch, pressure, muscle lengthening and proprioception render the feedback less effective as the subject can only use visual stimuli to understand the difference between the desired action and its result. If the latency period between the action and its feedback is too long, it can severely affect the ability of the patient to learn and improve their ability to effectively using the BCI [3]

### 1.2.1 Techniques

The most common techniques for BCIs are non invasive and based on the EEG, in particular on four types of potential which can be observed [4]:

- P300 - A potential which appears 300ms after an unexpected stimulus
- VEP - The brain response to a rapid visual stimulus
- Steady State VEP - VEP which appears in response to a visual stimulus flashing at a steady frequency

- SCP - Very slow endogenous potential which can be controlled with adequate training

	<b>Training</b>	<b>Bandwidth</b>
<b>P300</b>	No	Medium[4]
<b>VEP</b>	No	High[5]
<b>SSVEP</b>	No	High[5]
<b>SCP</b>	Yes	Low[6]

Table 1.2: Bandwidth of various BCI techniques

### 1.2.2 P300

P300 is an evoked potential occurring 300ms to 600ms, the name is used to categorise two types of potential:

- P3a elicited in the prefrontal cortex by new and unexpected stimuli
- P3b slower response potential appearing in the parietal cortex in response to a known but rare stimulus, is used in BCIs using the odd-ball paradigm. [7]

#### Odd-ball Paradigm

To elicit a P3b response the standard protocol is to present a series of equal and predictable visual or auditory stimuli interrupted by a different, rarer stimulus. The amplitude of the P300 potential increases with the rarity of the stimulus.

#### Usage in BCIs

The P3b potential can be used in BCI applications by properly modulating the characters on a screen in order to make the evoked potential a predictor of which character the user is looking at.

### 1.2.3 VEP and SSVEP

VEPs is a catch-all term for a group of Visually Evoked Potentials, categorised by:[8]

- Morphology of the Visual Stimulus
  - VEPs caused by flash stimulation
  - VEPs caused by graphic patterns like checkerboard lattice
- Frequency of visual stimulation
  - Transient VEPs (TVEPs): VEPs with visual stimulation frequency below 6 Hz
  - Steady-state VEPs (SSVEPs): VEPs with visual stimulation frequency above 6 Hz
- Field stimulation
  - Whole field VEPs
  - Half field VEPs
  - Part field VEPs

#### Usage in BCIs

BCIs are usually based on either VEPs or SSVEPs, evoked by carefully crafted stimuli on a computer screen. The patient is shown a grid of selectable options which flash based on a predefined pattern such as the evoked potential exhibits the same pattern, allowing the machine to detect which option the patient is looking at.

#### t-VEPs

A t-VEPs based BCI modulates the options to flash on and off such as no two patterns overlap, detection of the selected option occurs by checking the time at which the VEPs occur. This method features low SNR, requiring averaging of multiple epochs and achieving low speed.

#### f-VEPs

f-VEPs based BCIs modulate the frequency each option flashes at, such as the VEP occurs at the same fundamental frequency as the target option.

#### c-VEPs

c-VEPs achieve the highest speed by using pseudo-random code modulation to significantly increase the SNR allowing patients to communicate multiple bits per second

with high accuracy. The resulting VEPs can be easily isolated from the EEG signal because it is itself modulated with the same pseudo-random code as the target option.

#### **1.2.4 SCP**

Slow Cortical Potentials, as the name suggests, are slow shifts in the EEG voltage the patient can learn to control. Learning to control SCPs is possible for patients with any level of motor impairment but requires months, sometimes years, to be done effectively [9]

##### **Usage in BCIs**

The classic protocol is named "S1-S2" and is based on two tones, one high pitched (S1) which indicates that in two seconds the patient's EEG will be read for feedback. A second low pitched tone (S2) two seconds after the first announces the feedback is being recorded, and if it exceeds a threshold will be used as input for the BCI.

#### **1.2.5 Limitations**

The first three techniques mentioned in the 'Techniques' subsection are fundamentally visual trackers, and still rely on a physical movement the patient might not be able to perform. It thus shows a significant limitation: patients need to be in front of a screen to use the BCI, in addition those who can't see or don't have full control of their eye movements are not capable of interacting with the machine.

Fully endogenous BCIs such as those based on SCP can provide a type of interface not reliant on movement or external stimuli, but SCP based BCIs require long training and have a very low bandwidth which limit their usefulness [6] [10].

### 1.3 Motor Related Cortical Potentials

The Motor Related Cortical Potential is a component of the EEG signal located around the 1-2 Hz frequency band, appearing around 2 seconds before a planned or executed movement of a skeletal muscle. This potential can also be seen when the movement is imagined and even if the patient is not physically capable of performing it, rendering its detection a good candidate as a BCI technique.

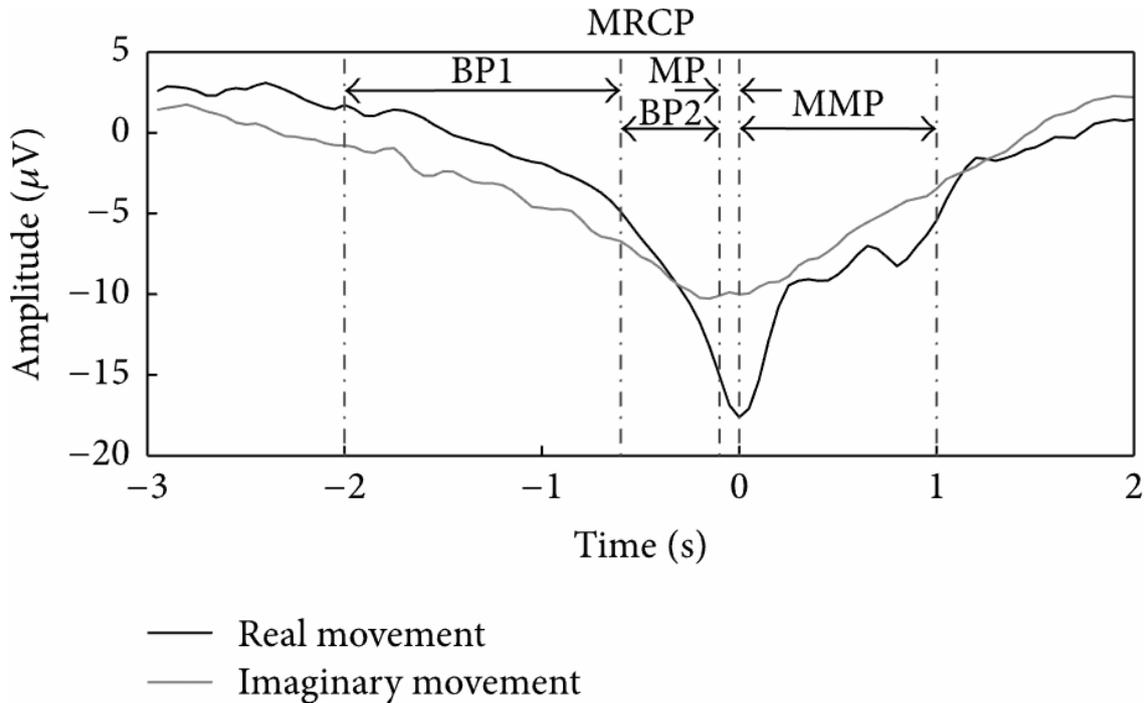


Figure 1.3: MRCPs of a healthy subject for real and imaginary right ankle dorsiflexion. Each wave is an average of 50 large Laplacian spatial filtered EEG trials

The MRCP has been studied for decades, studies have shown their size and delay are adjusted according to the participants' mental state and characteristics of the executed movement such as speed, accuracy and frequency, but only recently efforts have been devoted on developing systems for single trial MRCP detection for application in Brain Computer Interfaces.[11]

These efforts have been hindered by what is a common issue in BCIs, the signal to noise ratio, which, like most endogenous brain potentials recorded through EEG, is low making detection methods accurate enough for clinical use difficult to develop [12], though clinical studies have shown patients can learn to control and amplify MRCPs through training. [13].

# Chapter 2

## Mathematical Concepts

Before diving into the state of the art, it is important to go over the techniques commonly adapted in the analysis of multivariate data. To simplify, let's assume the data is made up of a series of measurements of temperature at  $n$  different locations at different times during the day. We can then consider each measurement as a point in an  $n+1$  dimensional space with the first  $n$  dimension representing the temperature in the  $n$  locations and the last one being time.

We can easily generalize this reasoning to any kind of measurement not involving temperature, time or any physical phenomenon. We can thus assume our data lives in an  $n$ -dimensional space with every dimension representing one particular characteristic, or *feature* of the underlying data measured by our system. The immediate question is: is this the best representation for our data? The answer depends on the nature of the data as well as our definition of best.

## 2.1 PCA

One definition of "best" can be that every component has to be uncorrelated to all others. Let's introduce the covariance matrix, a symmetrical square matrix where each (m x n) value is the variance between channel m and channel n in the original data. By finding a linear transformation which makes this matrix diagonal we can obtain a new system of reference which makes the components uncorrelated.

There are infinite transformations which satisfy this condition, so we need a second constraint to find the one we need. A sensible choice is picking the one which explains the most variance with the least amount of components.

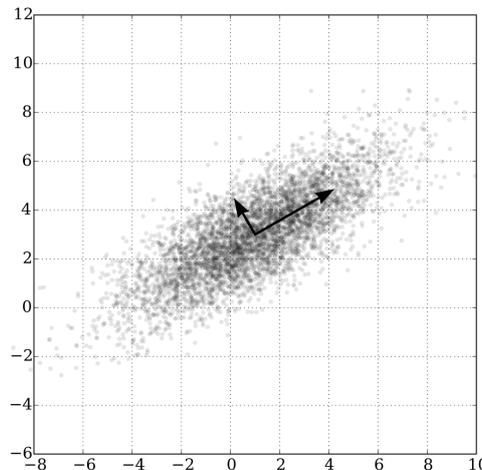


Figure 2.1: Directions of highest variance for a gaussian distribution

### 2.1.1 Eigenvalues and Eigenvectors

Given a linear transformation, an eigenvector is any vector which satisfies the following equation:

$$Av = \lambda v$$

where A is the matrix representing our transformation, v is the eigenvector, and  $\lambda$  is the eigenvalue. We can understand eigenvectors as the vectors that are only scaled by a factor equal to the eigenvalue by the linear transformation. The eigenvectors of a given matrix always form an orthogonal base, as such any transformation can have a number of eigenvectors at most equal to the number of dimensions of the space in which it applies.

### 2.1.2 Eigenvectors of the Covariance Matrix

The covariance matrix can be interpreted as a linear transformation which transforms white noise into data which features the same statistical characteristics as the

data it was calculated from. Intuitively, the eigenvectors of this matrix represent the directions of highest variance, but to prove it formally we can assert that the projection of the covariance matrix  $\Sigma$  on a vector  $v$  is equal to

$$\vec{v}^T \Sigma$$

The variance of the data is thus

$$\vec{v}^T \Sigma \vec{v}$$

Maximizing the variance equals to maximizing the above expression for vector  $\vec{v}$ , which is in a canonical form known as Rayleigh quotient. It can be demonstrated that the maximum of a Rayleigh quotient is the eigenvector associated with the largest eigenvalue of matrix  $\Sigma$ . We can repeat the calculation for the remaining eigenvectors, each component finding the direction which explains the most variance while being orthogonal to all the previous ones.

### 2.1.3 Principal Component Analysis

We have proven calculating the eigenvectors of the covariance matrix gives us a base which satisfies our first constraint by being orthogonal as well as the second one by being the directions in which the most variance is present. Eigenvalues, in this case, represent the amount of variance explained by the corresponding eigenvector, giving us the ability to arbitrarily reduce the dimensionality of the space by sorting the eigenvectors by the corresponding eigenvalues and removing the lowest ones. The new features obtained with this procedure are known as principal components, and the method is known as Principal Component Analysis (PCA).

## 2.2 Beyond PCA

As we previously implied, there is no definition of 'best' which works for every use case. The second constraint we imposed while explaining the PCA method assumes variance is a good predictor of how much information a component holds. In many cases this may not be accurate, and as such various alternatives to PCA have been proposed.

A common alternative to PCA is Independent Component Analysis (ICA)[14], which relies on the central limit theorem: signal mixtures tend to have gaussian probability density functions, so if we assume the source are non-gaussian we can use kurtosis or entropy to determine how much information the components hold.

A different approach is known as Second Order Blind source Identification (SOBI), where instead of diagonalizing the zero lag covariance matrix, the algorithm jointly diagonalizes a series covariance matrices at various time lags.[15] SOBI exploits non-whiteness as a measure of information, making it a good method to use on data where some or all the sources might be gaussian.

## 2.3 Singular Value Decomposition

Given a diagonalizable matrix  $M$ , it can be expressed as

$$M = PDP^{-1}$$

with  $P$  being the eigenvector matrix and  $D$  being the diagonal matrix which contains the corresponding eigenvalues.

It is possible to factorize any matrix including those which are not diagonalizable in a similar way using Singular Value Decomposition (SVD). This factorization takes the form

$$M = U\Sigma V^T$$

With  $U$  and  $V$  representing respectively the eigenvectors of  $MM^T$  and  $M^T M$  and  $\Sigma$  being a diagonal matrix containing the square roots of the eigenvalues of  $MM^T$  [16]. Without delving into the finer details, one can see that PCA and SVD are very closely related: since the covariance matrix of  $M$  is  $\frac{M^T M}{n-1}$   $V$  represents the principal components of  $M$  and  $\Sigma$  is the square root of the variance explained by every principal component multiplied by  $n - 1$ .

## 2.4 Manifolds

According to Encyclopedia Britannica, a manifold is "a generalization and abstraction of the notion of a curved surface; [...] a topological space that is modeled closely on Euclidean space locally but may vary widely in global properties" [17]. We will not delve into the details of manifolds as the amount of information on the topic is massive, but we will take a look at some notions about Riemann Manifolds which will be useful to understand further chapters.

### 2.4.1 Hilbert Spaces

A Hilbert space is a complete vector space in which the inner product is defined, it can have a finite or infinite amount of dimensions. From the inner product one can define a norm as:

$$\|x\|^2 = \langle x, x \rangle$$

And from a norm one can determine a way to measure distances:

$$d(x, y) = \|x - y\|$$

as well as angles:

$$\cos(\theta) = \frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}$$

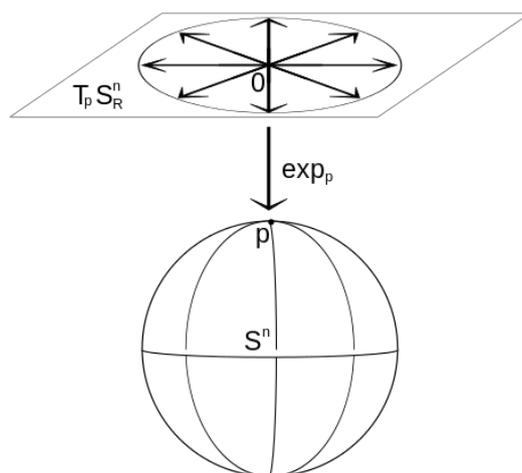


Figure 2.2: Exponential map of a sphere

## 2.4.2 Riemann Manifolds

A Riemann manifold is a manifold in which the tangent space at every point is a Hilbert space, allowing one to measure distances using what in manifold theory is called a "geodesic", a curve normal to the manifold at every point, representing the shortest distance between two points and a generalization of the notion of straight line.

Since the inner product and therefore the metrics are defined in the tangent space, performing calculations in Riemann manifolds commonly requires mapping from the manifold to its tangent space and vice versa: this is achieved using the exponential mapping and logarithmic mapping operations.

### Mapping To The Tangent Space

To map a point  $p$  on the tangent space to the manifold one needs to find the geodesic which starts at the center of the tangent space and, considering the distance between  $p$  and the center of the tangent space as a velocity, move a point along the geodesic with velocity  $p$  for 1 unit of time.

This operation is known as exponential map and its inverse as logarithmic map. Their formulation depends on the characteristics of the Riemann manifold.



# Chapter 3

## State of the art

### 3.1 Artifact Removal

EEG recordings are affected by a non negligible amount of electrical noise. Due to the small amplitude of the signal, the recordings are also sensitive to artifacts, in particular:

- Movement Artifacts: stochastic artifacts caused by the movement of the electrodes on the scalp.
- Blinks: electrical activity of high amplitude and frequency captured from the movement of the eyelid muscles during eye blinks.
- ECG: electrical activity of the heart.

Preprocessing is usually considered a necessary step for the evaluation of EEG signals. Since the frequency band of the MRCP is both known and smaller than the EEG spectrum all of the methods in this chapter employ one or more filters to isolate the appropriate frequency bands, but some methods employ additional techniques such as spatial filtering and ICA which exploit the multivariate nature of EEG recordings to further isolate the required potentials.

## 3.2 Processing

Most methods to recognize MRCPs are based on a band-pass filter, a spatial filter, optionally a feature extraction step and a classifier.

### 3.2.1 Reviews

To find the best methods developed in literature, it is imperative to analyze the most recent reviews.

The reviews used as reference in this thesis are the 2015 "Review of Techniques for Detection of Movement Intention Using Movement-Related Cortical Potentials" [18] by Shakeel et al. and the 2018 "Review of classification algorithms for EEG based brain-computer interfaces" by Lotte et al.[19]

### 3.2.2 Overview of the Methods

Niazi et al.[20] have shown an optimal linear combination of 9 EEG channels and a likelihood ratio based classifier can obtain 82.5% accuracy in the detection of the potentials on healthy subjects performing ankle dorsiflexion.

Ren Xu et al.[21] have obtained good performance in the online detection of MRCPs on non segmented EEG data, with 84% TPR and 1.5 FPM (false positives per minute) adopting Locality Preserving Projections as the spatial filter and Linear Discriminant Analysis as the classifier on 9 channel EEG data from 9 healthy subjects.

While the previous authors filtered the EEG signal between 0.1Hz and 20Hz, Lew et al.[22] obtained a 76% TPR with a narrow band zero phase IIR filter between 0.1Hz and 1Hz, only preserving very low frequency information but acquiring 34 channels.

Karimi et al.[23] obtained state of the art results with the constrained ICA method, a supervised variant of the standard ICA method, used as a preprocessing step for the LPP-LDA classifier built by Ren Xu et al.[21]. They obtained 87% TPR and 20% FPR on a dataset of 10-channel EEG data from 24 subjects filtered between 0.05Hz and 3Hz.

According to the 2018 review by Lotte et al.[19] Riemann Manifold based classifiers "clearly outperformed the other state-of-the-art methods in terms of accuracy". The review cites a paper [24] which claims: "We demonstrate that this new approach outperforms significantly state of the art results, effectively replacing the traditional spatial filtering approach". The technique modifies the classic linear SVM with a kernel which implicitly projects the data in the Riemann Manifold of Symmetric Positive Definite (SPD) matrices, achieving a 6% increase in accuracy from 80% to 86% compared to CSP+LDA on 22 channel EEG data.

### 3.3 Reproduced Methods

Three of the algorithms previously mentioned have been reproduced as a benchmark against which to compare the results obtained by the algorithm developed in this thesis.

#### 3.3.1 Optimal Spatial Filter

The method calculates a virtual channel as a linear combination of the EEG channels such as the sum of all the coefficients is equal to 0:

$$\begin{aligned} \text{maximize : } & 10 \cdot \log_{10} \left[ \frac{P(\sum_{k=1}^{nc} x_k S_k(t))}{P(\sum_{k=1}^{nc} x_k N_k(t))} \right] \\ \text{subject to : } & \sum_{k=1}^{nc} x_k = 0 \end{aligned}$$

With  $nc$  being the number of EEG channels,  $S$  being the signal and  $N$  being the noise. Since the true signal and noise are unknown, the windows in which a MRCP is present and absent are taken. Starting from the coefficients of the laplacian spatial filter

$$x_k = \left\{ \begin{array}{ll} 1, & k = 1 \\ -\frac{1}{nc-1} & k \neq 1 \end{array} \right\}$$

one can optimize the coefficient using the BFGS algorithm in order to maximize the Signal to Noise Ratio of such linear combination.

The BFGS algorithm does not allow for any type of constraint, but the original paper does not mention how they got around this issue to apply the equality constraint; it is my guess they got around the issue by implicitly parametrizing the constraint inside the loss function by using a penalty term, as this is how it has been implemented here.

$$\text{minimize : } \left( 10 \cdot \log_{10} \left[ \frac{P(\sum_{k=1}^{nc} x_k S_k(t))}{P(\sum_{k=1}^{nc} x_k N_k(t))} \right] \right)^{-1} + \left( \sum_{k=1}^{nc} x_k \right)^2$$

The classifier is based on the likelihood ratio [25], meaning it is necessary to calculate a reference signal to use in the classification process. This reference is calculated as the average of all the MRCPs in the training data, as two second windows ending in the negative peak of the potential.

### 3.3.2 LPP-LDA

#### LPP

LPP stands for Locality Preserving Projections. As the name suggests, this technique projects data in a lower dimensional space using a transformation which preserves the shape of the data [26]. It is a linear method, an alternative to PCA and other dimensionality reduction techniques, and it is used in this method to reduce the EEG data to 60% of the original number of channels.

#### LDA

Linear Discriminant Analysis is a supervised technique which aims to find a new system of reference in which the classes are maximally separable. It is a joint diagonalization problem, the two matrices to diagonalize are the class covariance matrix  $S_w$  and

$$S_b = \sum_{i=1}^c N_i (m_i - m)(m_i - m)^T$$

with  $c$  being the number of classes and  $m$  the class means. The diagonalization can be executed by finding the eigenvalues and eigenvectors of both matrices solving:

$$S_b \cdot v = \lambda S_w \cdot v$$

Such as  $v$  diagonalizes both  $S_b$  and  $S_w$ , and then projecting the data on the eigenvectors.

### 3.3.3 ARK-SVM

#### SPD Riemann Manifold

The space of all Symmetrical Positive Definite matrices forms a Riemann manifold. Barachant et al.[27] have shown the formula for distance, exponential map and logarithmic map in the Riemann manifold of SPD matrices.

#### Classification

The kernel used in this method redefines the inner product such as the distance metric is the distance in the tangent space of the SPD manifold in the point which represents the geometric median of all the training points.

Given a point  $C_i$  and the center of the tangent space  $C_{ref}$ , the kernel  $K_R$  is:

$$S_i = C_{ref}^{-1/2} \text{Log}_{C_{ref}}(C_i) C_{ref}^{-1/2}$$
$$K_R = \text{vect}(S_i)^T \text{vect}(S_j)^T$$

The original paper doesn't use the kernel directly, but uses an equivalent alternative method in which the covariance matrix of every window is transformed in the S matrix using the above formula, half vectorised and then fed to a linear SVM. For the sake of accuracy, we used the same method.

#### Geometric Median

Taking  $P(n)$  as the set of all points in the manifold, the geometric median, defined as:

$$\text{argmin}_{P \in P(n)} \sum_{i=1}^m d(P, P_i)$$

Can only be calculated by an optimisation process, as no closed form solution exists to the problem. To calculate the median, we have to know the tangent space, but to know the correct tangent space we have to know the median.

To solve this problem, we employed the algorithm shown in [27], which iteratively refines the estimation of the correct mean and thus of the correct tangent space. The stopping threshold  $\epsilon$  was set as  $10^{-5}$ .

### 3.3.4 Final Observations

Due to the lack of widely accepted standardised datasets, metrics and evaluation techniques it is fundamentally difficult to directly compare results from different authors. It is something that should be kept in mind when comparing different papers about the topic of BCIs.



# Chapter 4

## Methods

The main goal of my thesis has been to develop a technique to improve the signal to noise ratio of EEG recordings containing MRCPs.

### 4.1 Dataset

The dataset for this thesis was provided by *Imran Khan Niazi, PhD in Biomedical Engineering and Sciences* and consists of 500Hz EEG recording of 16 healthy patients performing one session of around 50 self paced hand movements each, recorded in 9 different channels: F7-F3-Fz-T7-C3-Cz-P7-P3-Pz. The data was divided as such:

- 1 patient devoted entirely for hyperparameter optimization
- Every remaining session has been divided in a 70% training set and 30% testing set.

It is important for every testing set to be consequent in time to the corresponding training set as to simulate a real world calibration procedure.

### 4.2 Detrending

The first step of preprocessing has been to high pass filter the signal at 0.04Hz, as to remove the drift of the signal mean over time, which is a measurement artifact. The filter used was a butterworth filter with 40db per decade of attenuation outside of the passing band, as to guarantee very low ripple in the passing and stop band and a mostly linear phase response. Filtering is applied to the entire signal before any processing.

### 4.3 Artifact removal

Whilst ECG and motion artifacts lay outside the frequency band of MRCPs and thus can be removed by means of a simple low-pass filter, blink artifacts exhibit a signal power significantly higher than the rest of the signal, rendering filtering ineffective as the small frequency components overlaying the required signal are non-negligible.

Gomez-Herrero[28] has shown Second Order Blind-source Identification[15] to be capable of reliably identifying and isolating blink artifacts. Removal of this kind of artifacts can be done in an unsupervised manner by measuring the fractal dimension of each independent component, as artifacts tend to have significantly lower fractal dimension than the signals they affect. He employed Sevcik's method[29] as the estimator for the fractal dimension, separating the components in two groups (artifacts and signals) exploiting their fractal dimension and removing the artifacts before reconstructing the signal.

This method has proven effective on the dataset used in this study which was indeed affected by sharp high-frequency and high-power peaks due to the eyes blinking, and as such it has been employed as the second step in the preprocessing pipeline

### 4.4 Resampling

As the Nyquist frequency of the signal is much higher than the frequency band of the potentials we are looking to extract, resampling to a lower frequency can be useful to reduce the amount of data to process, improving the performance of every subsequent step.

Before resampling, an antialiasing filter has been applied to prevent generating additional artifacts. The filter employed is yet again a Butterworth filter with the same attenuation as the detrending filter, but in this case featuring a passband below a conservative cutoff frequency of 20Hz. The data is resampled by a factor of 10 bringing the sampling frequency down to 50Hz.

## 4.5 Non-linear optimal spatial filter

In this section the algorithm I created to process EEG data is explained in detail.

### 4.5.1 Intuition

The method assumes the multivariate EEG signal indicated by matrix  $S$  can be described as a MRCP component  $A$  plus the rest of the EEG signal  $\tau$ , such as

$$A \cdot W + \tau = S$$

where  $A$  is a vector representing the MRCP source and  $W$  is a scaling matrix indicating how strongly the MRCPs appear in each channel. This model is valid under three mathematical assumptions:

- MRCPs are perfectly additive to the EEG signal
- The process which maps the source of the MRCPs to each channel is linear
- Such process does not affect the phase of the MRCPs

Empirical research in the field of EEG source localization has shown the sources can be modeled as independent dipoles affected by a nonlinear transfer function the properties of which are determined by the electrical conductivity of tissue in the scalp as well as additive measurement noise [30].

While the first assumption holds, the other two can be considered approximations of the more complex behaviour the brain exhibits in reality. Despite most authors [20] [21] [31] managing to obtain good results with fully linear models for the extraction of MRCPs from EEG recordings, a feature mapping has been implemented in the model described by this thesis as to allow the algorithm to learn non-linear mappings, dropping the second mathematical assumption and thus improving results.

The main issue that affects our model is that the source  $A$ , the scaling matrix  $W$  and the rest of the eeg signal  $\tau$  are all unknown, making the equation impossible to solve unless we add more conditions.

### 4.5.2 Ordinary Least Squares

By rearranging the equation terms we have:

$$A \cdot W = S - \tau$$

Which is in the canonical form

$$A \cdot x = B + residual$$

. This is an Ordinary Least Squares (OLS) problem.

Under a set of mathematical assumptions, the OLS method tells us the best  $W$  is the one which minimizes the residual  $\tau$ . The mathematical assumptions are:

- The residuals have zero conditional mean
- The predictors are linearly independent
- The residuals are spherical

The validity of the above assumptions will be analysed in section 4.5.5

### 4.5.3 Source Estimation

At this stage, the equation is not particularly useful, as to calculate the scaling vector we need to know the signal generated by the MRCP source, meaning to calculate  $A$  we have to find  $W$ , but to find  $W$  we have to know  $A$ .

A calibration process can solve this issue. Assuming the  $W$  vector is not a function of time, which is reasonable considering the MRCP generating dipole is primarily affected by the physical properties of the skull and of the measuring system which are supposed to be time-invariant, we can estimate  $A$  by knowing the instant in which a movement was executed during a training session and generating a reference signal as a 1s triangular wave peaking in that instant, then calculate  $W$  by solving the model and obtain  $A$  for new unseen EEG recordings as

$$A = W^{-1} \cdot S$$

### 4.5.4 Matrix Inversion

The solution of the above equation does not exist as  $W$ , being a vector, does not allow a multiplicative inverse.

To remove the need of inverting the  $W$  vector we need to understand the meaning of  $W^{-1}$ : it's the vector which when premultiplied by  $S$  gives us the closes possible approximation of  $A$ . By rearranging once again our equation as:

$$S \cdot W^{-1} - \tau = A$$

We can change the meaning of our  $W$  variable and invert  $\tau$  to simplify to:

$$S \cdot W + \tau = A$$

Our model gets better the closer  $\tau$  becomes to zero, as our linear combination of EEG channels gets closer to our MRCP source  $A$ . We can set  $\tau = 0$ , and provided  $S$  is invertible solve the system as:

$$\text{training set : } W = S^{-1} \cdot A$$

$$\text{testing set : } A = S \cdot W$$

### 4.5.5 Underdetermined Case

$S$  is only invertible if it is square, which entails recording as many measurement channels as time samples. This is at best extremely expensive and time consuming, being impossible in the vast majority of cases.

If the number of channels is less than the number of timesteps solving for  $W$  becomes an underdetermined linear problem which has infinite solutions.

There exist a number of methods to solve this kind of problem, in this case we used the Moore-Penrose inverse of  $S$ , also known as pseudo-inverse, since it allows us to use a performance optimization which will be described in a further subsection. Replacing  $S^{-1}$  with its pseudo-inverse and solving the same equation as the determined case is equivalent to minimizing the square of the residual  $\tau$  while obtaining the minimum norm solution. As expected in this case the solution is unlikely to feature a residual  $\tau = 0$ , but still represents the best linear combination of channels to extract the MRCPs.

### 4.5.6 OLS Assumptions

Now that we have the solution to our model, it is imperative we analyse whether the main OLS assumptions are verified:

#### **The residuals should have zero conditional mean**

This is also known as the Exogeneity constraint. Without going into detail about the causes that can make this assumption fail, I will direct the reader which wishes to go into further detail into this topic to the work of Singh et al. [32], which lists the "main causes of failure of exogeneity" as

- Measurement Error
- Reverse Causality
- Omitted Variables
- Omitted Sample Selection
- Lagged Dependent Variables

We can easily see that our predictor matrix  $S$  should not be affected by the first four under proper measurement conditions, and provided the assumption that the process which maps the source of the MRCPs to each channel does not affect its phase is verified, the fifth cause is also not verified.

#### **The predictors should be linearly independent**

There is no guarantee of this, in fact since we assume every channel is the EEG recording to be the sum of all the sources in the brain we expect the predictors to be strongly correlated amongst each other.

## The residuals should be spherical

This implies the variance of the residual is diagonal and not dependent on time. If we assume the MRCPs are small compared to the matrix  $S$  and thus the EEG signal, we can ensure this assumption is close to being verified by ensuring the matrix  $S$  is spherical itself.

### 4.5.7 Whitening Transformation and SVD

Applying a transformation to  $S$  that makes it spherical, meaning it diagonalizes its covariance matrix and normalises the variance of each channel, can ensure the model satisfies the last two assumptions improving the reliability of the results.

To do so the whitening transformation has been employed, in particular SVD whitening. This choice is rather unusual, as techniques involving PCA are much more common, but since there is an infinite amount of possible whitening transformations this choice comes as an important performance optimization due to two main reasons:

- SVD is a common algorithm in data processing, and as such is implemented in historical and well optimised software packages and libraries for most existing programming languages. Due to the popularity of SVD, there exist a number of extremely performing algorithms to factorise a matrix in as little time as possible, which is a desirable feature considering MRCP detection should be performed in real time.
- The calculation of the Moore-Penrose inverse  $S^{w+}$  of the whitened matrix  $S^w$  can be executed at no additional cost by simply inverting the left and right singular values. Not only this allows us to skip a computationally expensive step of our algorithm, it also proves to be one of the most numerically robust algorithms to execute this calculation.

Executing SVD whitening is a simple process: the factorisation of matrix  $M$  is

$$M = U\Sigma V^T$$

Since  $U$  and  $V$  are orthonormal, multiplying them together leads to a matrix with an identity covariance matrix. We can prove this matrix still contains the information needed by our regression model by noticing  $\Sigma$  is a square diagonal matrix, meaning it simply represents a scaling transformation. We can thus calculate :

$$S = U\Sigma V^T$$

$$S^w = UV^T$$

$$S^{w+} = VU^T$$

The final linear model is now:

$$\text{training set : } W = S^{w+} \cdot A$$

$$\text{testing set : } A = S^w \cdot W$$

## 4.5.8 Learning Non-Linear Mappings

Up until now, the method we devised is only able to infer linear mappings between the EEG signal and the MRCs. In the field of machine learning, a common strategy to allow SVM classifiers to classify non linear data is known as the kernel trick.

### Kernel Trick

A Support Vector Machine (SVM) is a linear classifier which separates data in two classes by the hyperplane which ensures the maximum separation between the classes. SVM fails in case the data is not linearly separable, but non linearly separable data can be linearly separable in a different, usually higher dimensional, feature space, as shown in Fig 4.1

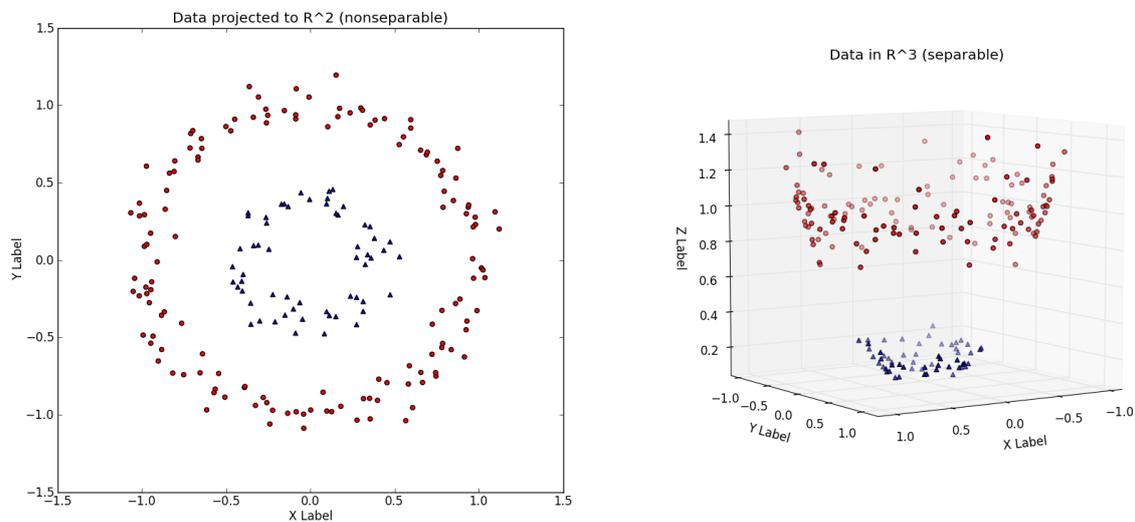


Figure 4.1: Mapping the predictors in a higher dimensional space makes separating two classes which were previously nonseparable possible

Kernels are particularly common in SVMs because it is possible to define the classification formula such as the predictor matrix  $x$  only appears in the form  $x_i^T \cdot x_j$ . In this case it is possible to execute the operation in a feature space without ever having to explicitly visit it, by simply replacing  $x_i^T \cdot x_j$  with  $K(x_i, x_j)$  with  $K$  being a kernel function.

This trick is not just a performance optimization, as it allows operating in infinite dimensional spaces which would be impossible to map the data to. In particular the most common kernel, the Radial Basis Function, is one which maps the data in an infinite dimensional space and it is particularly useful as it allows a linear classifier to learn any smooth non linear function.

### Random Fourier Features

The classifier we have devised is not in the form expressed above, and as such the kernel trick can't be applied to our linear model. We can't explicitly map our

predictors to an infinite dimensional feature space such as the one given by the RBF kernel because it would require infinite memory to store the data, but we can approximate the kernel. In 2008 Ali Rahimi and Benjamin Recht [33] have proposed a method to approximate kernels which satisfy certain mathematical properties with a finite dimensional feature space.

The idea is simple: shift-invariant kernels such as the RBF one, can be expressed as a convex combination of rank one kernels, whereas a rank 1 kernel is a kernel which maps data to a 1 dimensional feature space.

Given the fourier transform of the RBF kernel it is possible to consider its frequency spectrum as a probability density function to sample  $n$  points from it, each point being a rank one kernel: the paper shows that for every pair of transformed data points  $x_i^f, x_j^f$ ,  $x_i^f \cdot x_j^f$  is an unbiased estimator of  $k(x_i, x_j)$ , the approximation error decreasing with  $n$  asymptotically approaching infinity. The new kernel has finite dimensionality and can be simply reconstructed from the  $n$  sampled points, so we can use it to explicitly map the EEG data before feeding it to the algorithm.

As shown in the *Results* section, 200 dimensions are enough to provide a very significant performance boost to the algorithm without overfitting.

# Chapter 5

## Results

### 5.1 Models

#### 5.1.1 State of the art

From the state of the art, three models have been reproduced for comparison:

- Optimised Spatial Filter (OSF) with quasi-newton BFGS optimiser + likelihood ratio based detector[20]
- LPP-LDA with 60% dimensionality reduction [22]
- ARK-SVM by Barachant et al. [27]

#### 5.1.2 New Models

From the previous chapter, two models have been implemented:

- Optimised Spatial Filter using Moore-Penrose inverse and pre-whitening
- Non-Linear Optimised Spatial Filter (NL-OST) using Moore-Penrose inverse, pre-whitening and a 200d approximated RBF kernel

At the end of these two models, a single decision tree computes the final result.

## 5.2 Segmentation

Training is performed on continuous traces, while the results are computed on 2s segmented windows of EEG data taken from the testing set. For every movement of the user a single window is taken containing the 2 seconds before the motion execution and a second window is taken from 4 to 6 seconds before the movement in an area in which there are no MRCPs, the algorithms are thus asked to solve a balanced classification problem. The optimal threshold for the classic OSF algorithm has been calculated using cross-validation on the training data and the Receiver Operating Characteristics (ROC) curve.

### 5.2.1 Preprocessing

Every EEG recording has been filtered and processed to remove artifacts as detailed in the "Methods" section.

## 5.3 Metrics

The metrics chosen for the evaluation of the results are the True Positive Rate (TPR) and False Positive Rate (FPR). They have been reported per-patient alongside the global mean and standard deviation. The time taken to analyze the dataset on an intel i5-7200U processor is also shown.

## 5.4 Simulated EEG Traces

A simulated EEG signal provides a controlled setup in which to test the models as well as a source of potentially infinite data on which to experiment

### 5.4.1 Simulator

The simulator considers the EEG as a simple stationary stochastic process with a gaussian probability density function of mean 0 and standard deviation 1, while the MRCP wave is modeled as the first half of a cosine wave, lasting one second, as shown in Figure 5.1

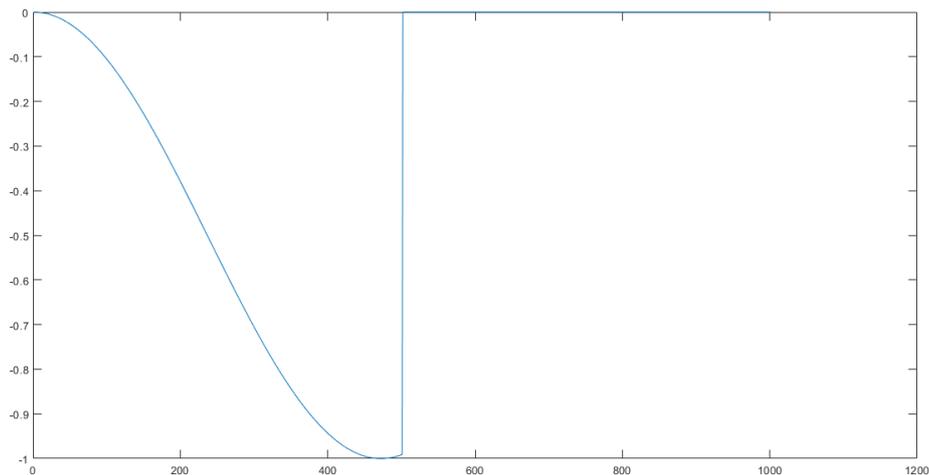


Figure 5.1: Simulated MRCP signal

To generate the MRCP component of the signal a vector of zeros with a kronecker delta every  $10 \pm 3s$  has been convolved with the modeled MRCP wave and then multiplied by a random number sampled from the same pdf the EEG is generated

from for every different channel.

This leads to a 9 channel simulated EEG in which the Signal to Noise Ratio can be easily tuned, which approximates the head as a linear system which only performs a frequency independent scaling.

## 5.4.2 Results

The aim of the simulator was to test the performance of the models as the SNR varies, as such the percentage of correctly classified windows has been used as the performance metric. The results are presented in Figure 5.2. For every SNR level 4 samples have been averaged, but despite this the variance is very high among measurements; as such to make the results easier to understand a 4 samples sliding window average has been computed and shown in Figure 5.3.

While all algorithms experience a sharp drop in performance around the -35db mark, the LPP-LDA proves to be the most effective algorithm at every noise level, followed by the RSVM method at low noise and the linear W-OSF when the noise increases.

It is interesting to notice that the NL-OSF model exhibits lower performance than its simpler linear variant W-OSF: this can be explained by recalling the simulator is perfectly linear, rendering the increased number of coefficients of the NL-OSF useless and therefore making the model prone to overfitting on the training data. This is also the likely reason why the most performant method according to literature, RSVM, falls short of the expectations.

In fact, both the non linear methods (NL-OSF, RSVM) experience a sharper drop in performance as soon as the noise increases, with RSVM being surpassed in accuracy by W-OSF and NL-OSF reaching the same performance as the standard OSF.

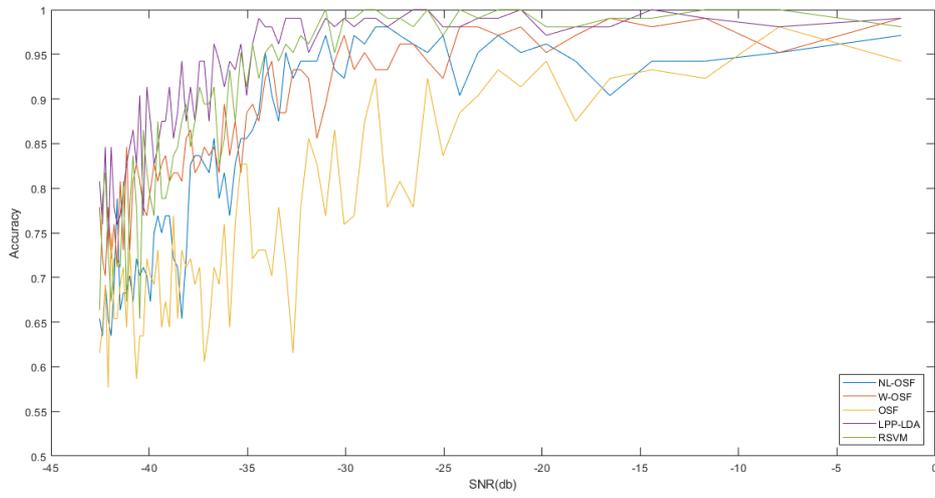


Figure 5.2: Performance of the models on simulated data as a function of SNR, average of 4 samples per SNR level

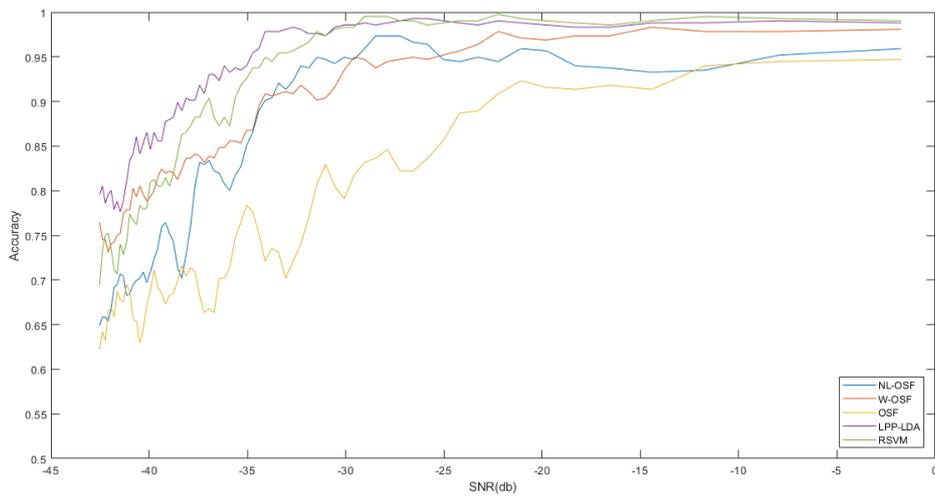


Figure 5.3: Smoothed performance of the models on simulated data as a function of SNR, average of 4 samples per SNR level plus an additional sliding window average of 4 samples

## 5.5 Real EEG Traces

To properly evaluate the performance of the models it is imperative to use real recordings of humans performing motor executions.

### 5.5.1 Dataset

As a reminder the dataset for this thesis was provided by *Imran Khan Niazi PhD in Biomedical Engineering and Sciences* and consists of 9 channel, 500Hz EEG recordings of 16 healthy patients each performing one session containing around 400 seconds and 50 self paced hand movements, totaling 1 hour and 46 minutes of EEG recordings. The data was divided as such:

- 1 patient devoted entirely for hyperparameter optimization
- Every remaining session has been divided in a 70% training set and 30% testing set.

## 5.5.2 Extracted MRCPs

A comparison of the various OSF algorithm against the reference signal on the testing set of Patient 1 is shown in Figure 5.4.

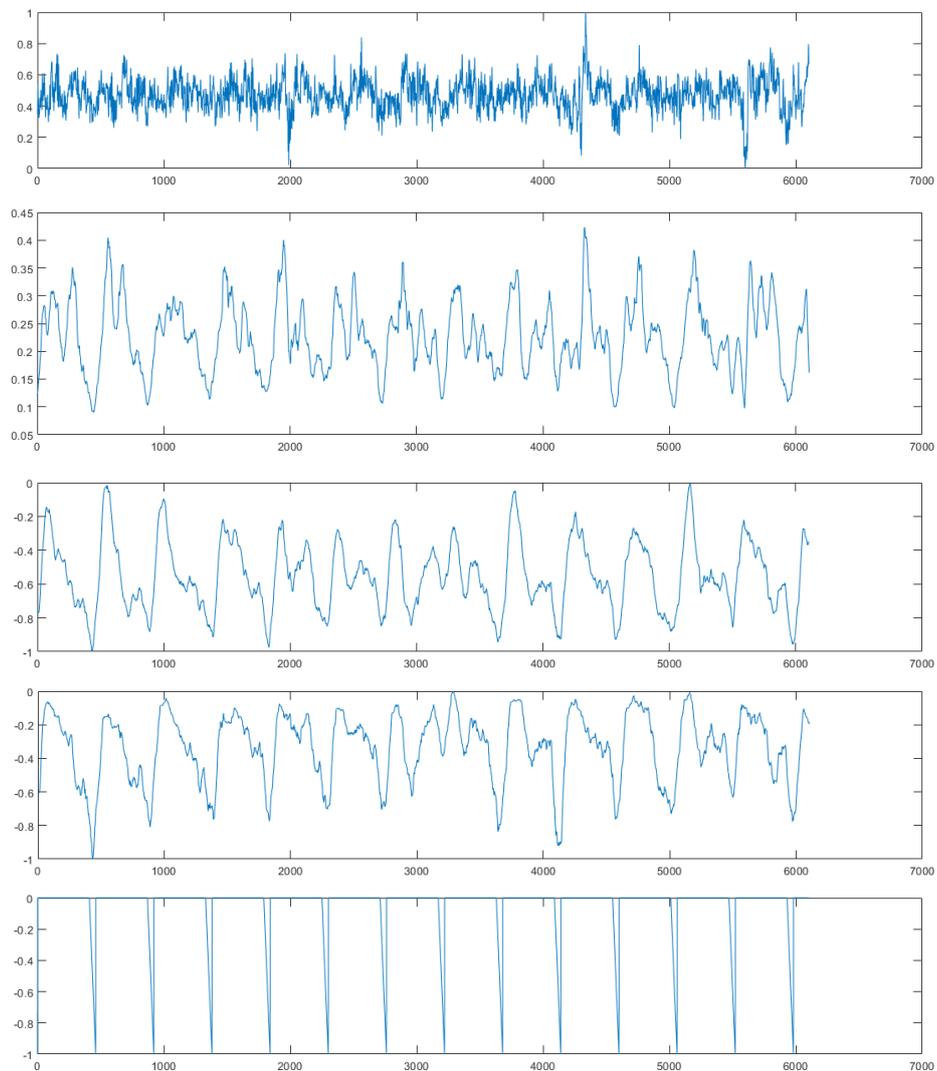


Figure 5.4: Comparison of the extracted MRCPs for the OSF algorithm on the testing set of Patient 1:

1. Filtered EEG channel 4, the one which shows the highest SNR for the MRCPs
2. Classical OSF
3. Whitened OSF (ours)
4. NL-OSF (ours)
5. Reference Signal

### 5.5.3 Results

After running every model on every patient, the data collected is available in Table 5.1.

Patient	NL-OSF (ours)		W-OSF (ours)		Classic OSF		LPP-LDA		A-RKSVM	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
1	0,93	0,07	0,93	0,07	0,79	0,07	0,93	0,07	1	0,07
2	0,86	0,07	0,57	0,21	0,86	0,29	0,71	0,43	1	0,07
3	0,85	0,08	0,77	0,31	0,62	0,23	0,77	0,54	0,54	0,08
4	0,69	0,23	0,62	0,23	0,46	0,08	0,69	0,31	0,85	0,31
5	0,79	0,07	0,57	0,29	0,50	0,21	0,86	0,07	0,86	0,07
6	0,79	0	0,71	0,07	0,93	0	0,79	0	0,93	0
7	0,92	0	0,92	0,08	0,92	0,69	1	0,08	0,85	0,15
8	0,79	0,21	0,79	0,42	0,93	0,79	0,93	0	0,79	0,07
9	0,85	0,15	0,85	0,23	0,15	0	0,85	0,31	0,92	0,31
10	0,53	0,07	0,67	0	0,27	0	0,73	0,07	0,80	0,27
11	0,79	0	0,86	0,29	0,93	0,86	0,57	0,14	0,86	0,21
12	0,86	0,14	0,57	0,36	0,50	0,36	0,64	0,36	1	0
14	0,79	0,14	0,71	0,29	0,79	0,21	0,93	0,50	0,93	0,14
15	0,79	0	1	0	0	0	1	0	1	0
16	0,85	0,23	0,62	0,54	0,85	0	0,69	0,23	0,92	0,31
Mean	<b>0,81</b>	<b>0,10</b>	<b>0,74</b>	<b>0,23</b>	<b>0,63</b>	<b>0,25</b>	<b>0,81</b>	<b>0,21</b>	<b>0,88</b>	<b>0,14</b>
Std	0.10	0.08	0.14	0.16	0.31	0.30	0.13	0.19	0.12	0.12
Time	63s		44s		183s		110s		91s	

Table 5.1: Results on Real EEG Data

We can see that while LPP-LDA performs significantly better than both linear Optimised Spatial Filter algorithms, our non linear model which performed poorly on linear data due to overfitting is now able to make use of the additional parameters and non linearity to obtain a good fit on the training data, outperforming LPP-LDA on the testing set while proving faster and more consistent among patients.

AK-SVM proves better performing than my method, achieving a 1.5% increase in accuracy at the expense of a 50% increase in calculation time. It is to be noted that the higher specificity and lower computation time my method provides are highly desirable features in BCIs[13][34], making the method a faster yet less accurate alternative to Riemann manifold based A-RKSVM.



# Chapter 6

## Conclusions

The field of BCIs is still in the early days, but as Károly Zsolnai-Fehér often says, when a field is still young "two more papers down the line" are enough to provide a significant improvement on any method or technology.

The classifier I have developed proved efficient, consistent, and significantly faster than the other algorithms tested.

My hope is that the Non-Linear Optimised Spatial Filter, coupled with a classification algorithm more sophisticated than a single decision tree or a preprocessing step such as constrained Independent Component Analysis, may prove helpful in developing newer and better technologies to help those who currently lack the capability to communicate with the outside world.

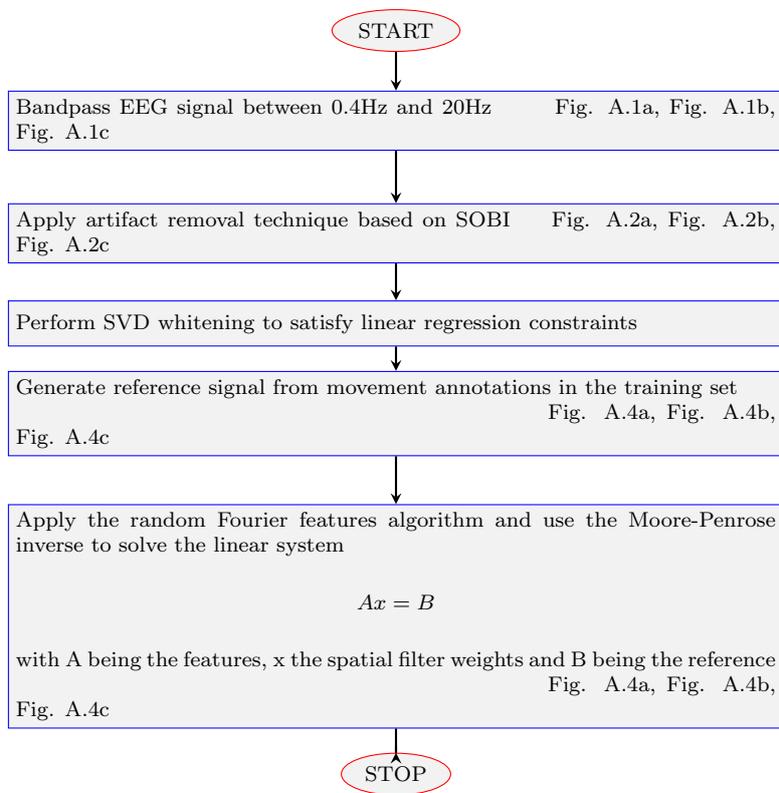
This thesis has been a fantastic learning experience, possibly teaching me better than any course how to read a research paper, understand the mathematics behind it and reproduce the technique; how to work autonomously to develop something that didn't exist before, and most of all how to deal with a massive, unspecified amount of work in a limited time.

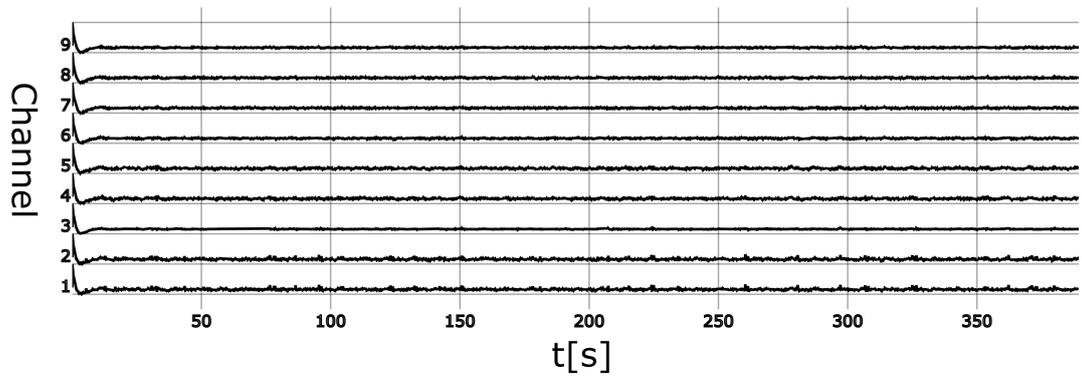


# Appendix A

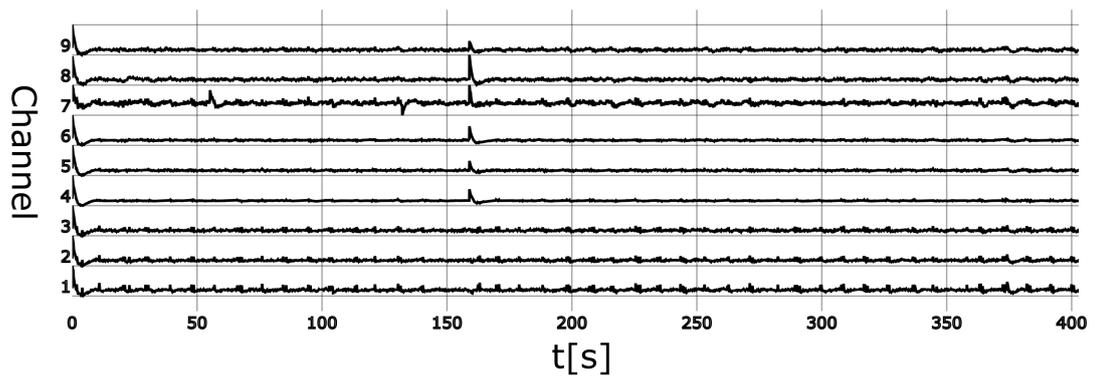
## Appendix

### A.1 Flow Graph of the NL-OSF algorithm and additional figures

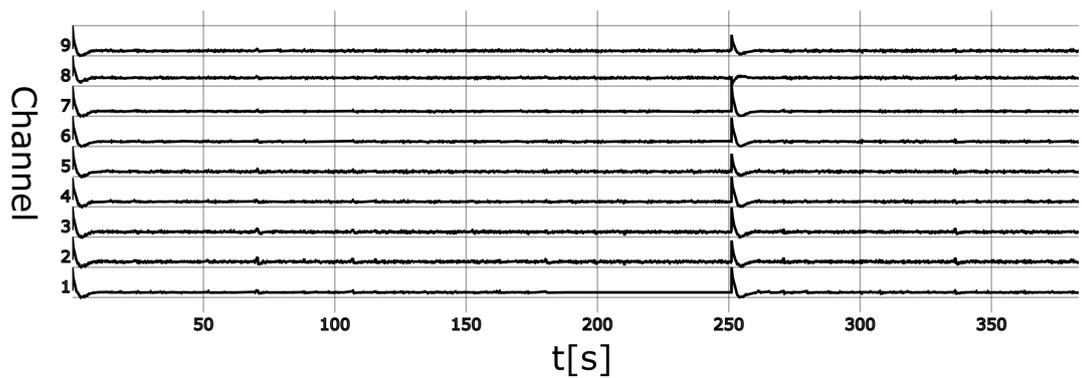




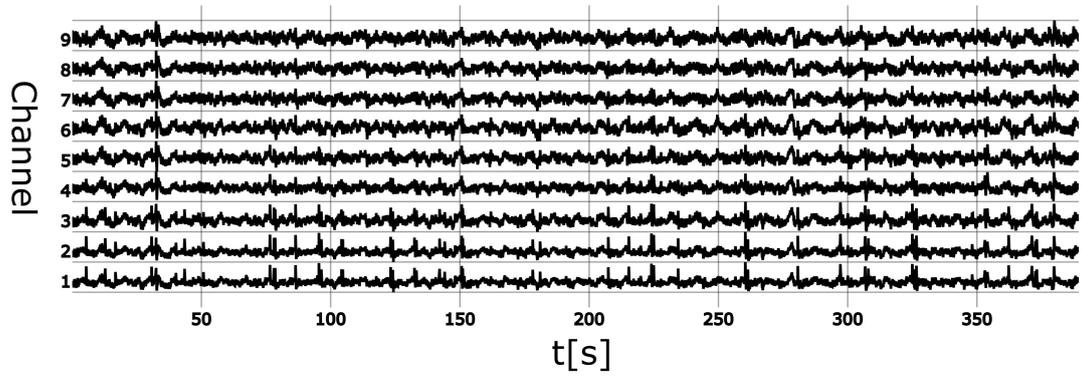
(a) EEG, Bandpass between 0.4Hz and 20Hz - Patient 1



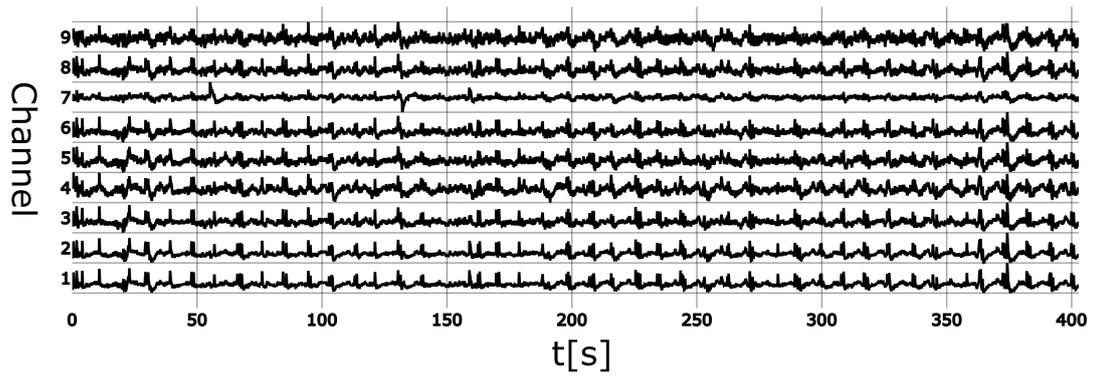
(b) EEG, Bandpass between 0.4Hz and 20Hz - Patient 2



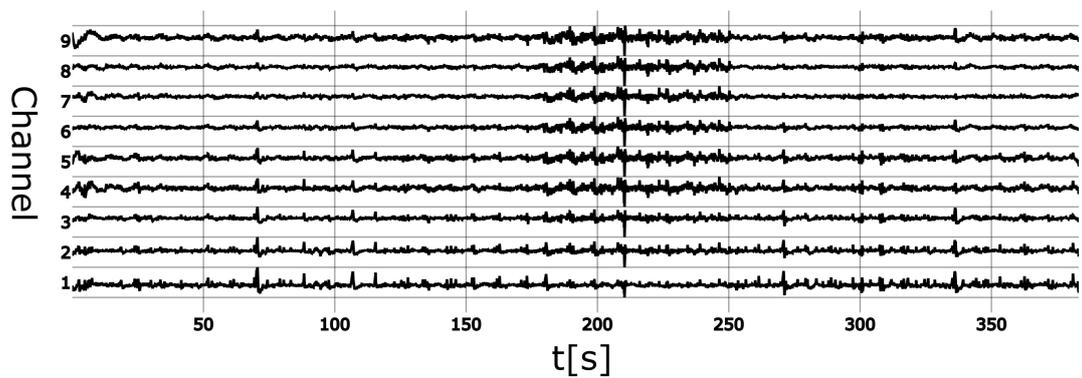
(c) EEG, Bandpass between 0.4Hz and 20Hz - Patient 3



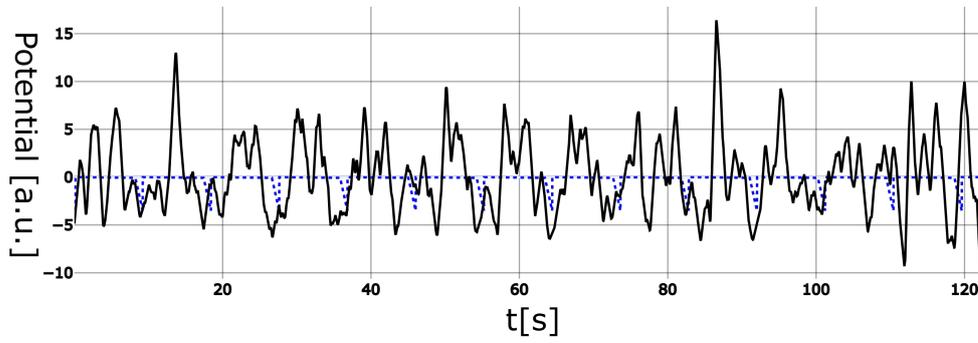
(a) EEG, Bandpass and Artifact Removal - Patient 1



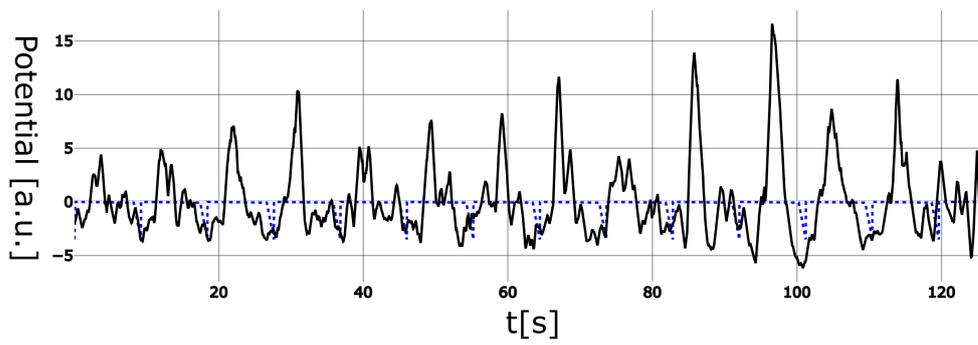
(b) EEG, Bandpass and Artifact Removal - Patient 2



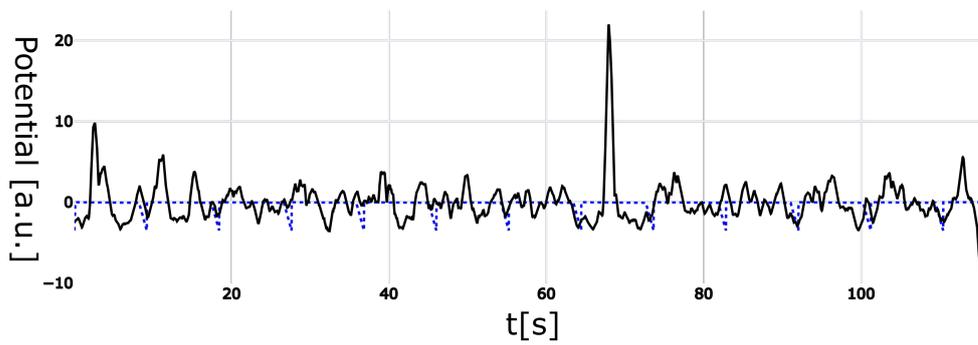
(c) EEG, Bandpass and Artifact Removal - Patient 3



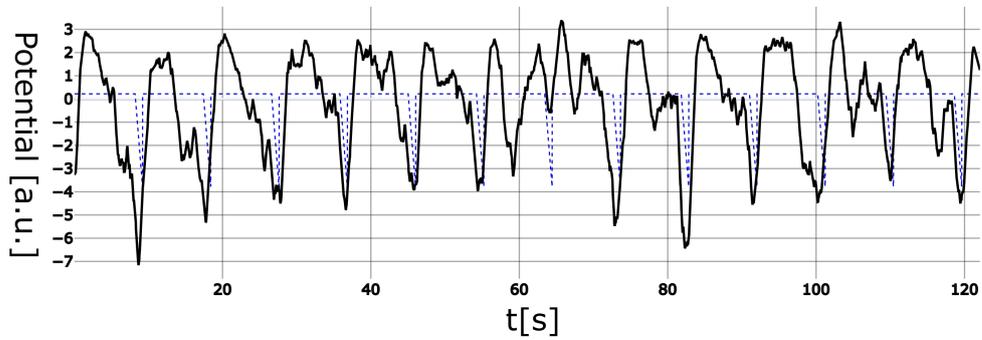
(a) Result of the OSF Algorithm - Testing Set - Reference in dashed blue - Patient 1



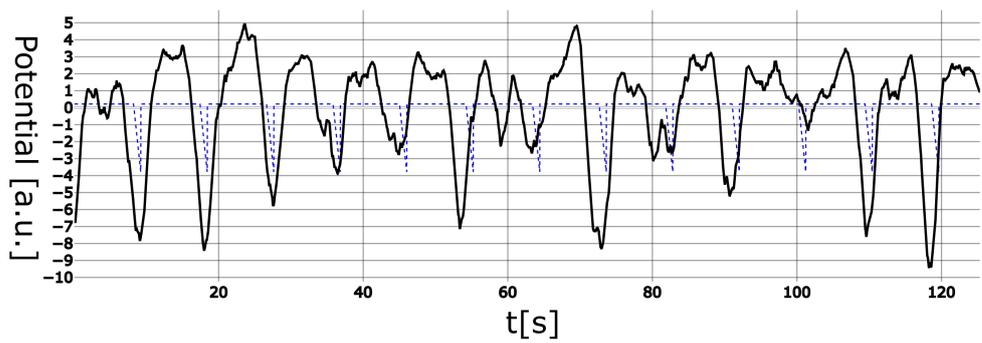
(b) Result of the OSF Algorithm - Testing Set - Reference in dashed blue - Patient 2



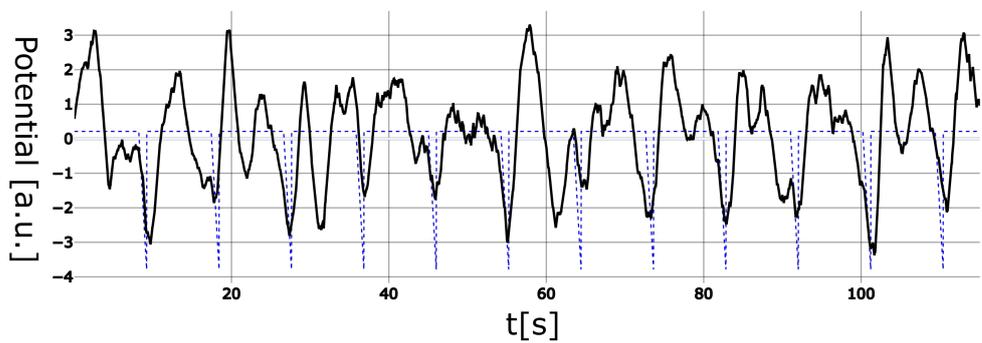
(c) Result of the OSF Algorithm - Testing Set - Reference in dashed blue - Patient 3



(a) Result of the NL-OSF Algorithm - Testing Set - Reference in dashed blue - Patient 1



(b) Result of the NL-OSF Algorithm - Testing Set - Reference in dashed blue - Patient 2



(c) Result of the NL-OSF Algorithm - Testing Set - Reference in dashed blue - Patient 3

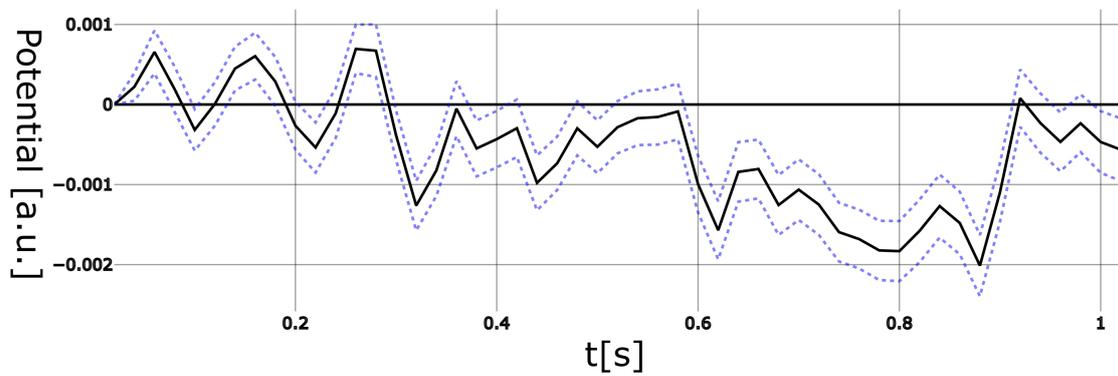


Figure A.5: Mean and Standard Error of MRCPs in the testing set

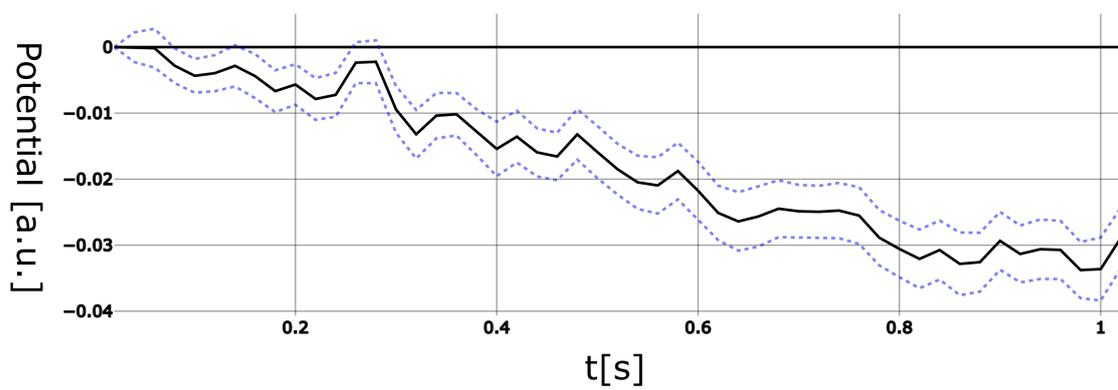


Figure A.6: Mean and Standard Error of MRCPs in the testing set after NL-OSF

## A.2 Matlab code for NL-OSF

Given EEG as the EEG data, MovementOnset as the labeled instants in which the movement is imagined or execute and training as the number of activations which should be in the training set:

```
%Ground truth vector
gt=zeros(1,length(EEG));
for i=1:length(MovementOnset)
    gt(MovementOnset(i)-1*fc:MovementOnset(i))=0:1/fc:1;
end

gt=centerRows(gt);
train=gt(1:MovementOnset(training));

%Random Fourier Features
w = randn(200,size(EEG,1));
b = 2 * pi * rand(200, 1);
EEG = cos(16 * w * EEG + b * ones(1,size(EEG,2)));

%Whitening
EEG=centerRows(EEG);
[U, S, V] = svd(EEG,'econ');
EEG=U'*V';

X=EEG(:,1:MovementOnset(training));

%System solution
mpinv= V(1:length(train),:)*U;
W=train*mpinv;

s=(EEG'*W')';
s=nlfilter(s,[1 80],@sum);
extracted = s;

tree=fitrtree(s(:,1:MovementOnset(training))',train');
s=tree.predict(s')';

surrogate = nlfilter(s(1,:),[1 50],@mean);
```

## A.3 Helper Matlab class for working in the SPD matrix manifold

```

classdef RiemannSpace<handle
    properties
        matrix
        sqP
        sqmP
    end

    methods(Static)

        function out=geometricMean(matrixArray)
            eps=10e-5;
            Ps=mean(matrixArray,3);
            PsC=RiemannSpace(Ps);
            quality=9999;
            niter=0;
            while quality>eps && niter<1000
                S=zeros(size(matrixArray(:,:,1)));
                for j=1:size(matrixArray,3)
                    S=S+PsC.logMap(matrixArray(:,:,j)+eps);
                end
                S=S./size(matrixArray,3);
                quality=norm(S,'fro');
                niter=niter+1;
                Ps=PsC.expMap(S);
                PsC=RiemannSpace(Ps);
            end
            out=Ps;
        end

        function out=distance(P1,P2)
            peg=eig(P1\P2);
            peg=peg(peg>0);
            dist=sum(log(peg).^2);
            out=norm(dist,'fro');
        end
    end

    methods(Access=public)
        function self=RiemannSpace(SPDMatrix)
            self.matrix=SPDMatrix;
            self.sqP=SPDMatrix^0.5;
            self.sqmP=SPDMatrix^-0.5;
        end

        function S = logMap(self,point)
            sqP=self.sqP;
            sqmP=self.sqmP;
            S=sqP*logm(sqmP*point*sqmP)*sqP;
        end
    end
end

```

```
end
```

```
function S = expMap(self,point)
    sqP=self.sqP;
    sqmP=self.sqmP;
    S=sqP*expm(sqmP*point*sqmP)*sqP;
end
```

```
end
```

```
function out = kernel(self,U)
    v1=self.sqmP*self.logMap(U)*self.sqmP;
    v1=tril(v1);
    out=real(v1(:));
end
```

```
end
```

```
end
```

```
end
```

# Bibliography

- [1] U.S. National Library of Medicine. *EEG*. URL: <https://medlineplus.gov/ency/article/003931.htm>.
- [2] J J Vidal. “Toward Direct Brain-Computer Communication”. In: *Annual Review of Biophysics and Bioengineering* 2.1 (1973). PMID: 4583653, pp. 157–180. DOI: 10.1146/annurev.bb.02.060173.001105. eprint: <https://doi.org/10.1146/annurev.bb.02.060173.001105>. URL: <https://doi.org/10.1146/annurev.bb.02.060173.001105>.
- [3] Moritz Grosse-Wentrup, Donatella Mattia, and Karim Oweiss. “Using brain-computer interfaces to induce neural plasticity and restore function”. In: *Journal of neural engineering* 8 (Mar. 2011), p. 025004. DOI: 10.1088/1741-2560/8/2/025004.
- [4] Fabrizio Beverina et al. “User adaptive BCIs: SSVEP and P300 based interfaces”. In: *PsychNology Journal* 1 (2003), pp. 331–354.
- [5] Stavros I. Dimitriadis and Avraam D. Marimpis. “Enhancing Performance and Bit Rates in a Brain-Computer Interface System With Phase-to-Amplitude Cross-Frequency Coupling: Evidences From Traditional c-VEP, Fast c-VEP, and SSVEP Designs”. In: *Frontiers in Neuroinformatics* 12 (2018), p. 19. ISSN: 1662-5196. DOI: 10.3389/fninf.2018.00019. URL: <https://www.frontiersin.org/article/10.3389/fninf.2018.00019>.
- [6] J. R. Wolpaw et al. “Brain-computer interface technology: a review of the first international meeting”. In: *IEEE Transactions on Rehabilitation Engineering* 8.2 (June 2000), pp. 164–173. ISSN: 1063-6528. DOI: 10.1109/TRE.2000.847807.
- [7] Nancy K Squires, Kenneth C Squires, and Steven A Hillyard. “Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man”. In: *Electroencephalography and Clinical Neurophysiology* 38.4 (1975), pp. 387–401. ISSN: 0013-4694. DOI: [https://doi.org/10.1016/0013-4694\(75\)90263-1](https://doi.org/10.1016/0013-4694(75)90263-1). URL: <http://www.sciencedirect.com/science/article/pii/0013469475902631>.
- [8] Rajesh Singla. “SSVEP-Based BCIs”. In: *Evolving BCI Therapy*. Ed. by Denis Larrivee. Rijeka: IntechOpen, 2018. Chap. 6. DOI: 10.5772/intechopen.75693. URL: <https://doi.org/10.5772/intechopen.75693>.
- [9] T. Hinterberger et al. “Brain-computer communication and slow cortical potentials”. In: *IEEE Transactions on Biomedical Engineering* 51.6 (June 2004), pp. 1011–1018. ISSN: 0018-9294. DOI: 10.1109/TBME.2004.827067.
- [10] Ioulietta Lazarou et al. “EEG-Based Brain-Computer Interfaces for Communication and Rehabilitation of People with Motor Impairment: A Novel Approach of the 21st Century”. In: *Frontiers in Human Neuroscience* 12 (2018),

- p. 14. ISSN: 1662-5161. DOI: 10.3389/fnhum.2018.00014. URL: <https://www.frontiersin.org/article/10.3389/fnhum.2018.00014>.
- [11] Ying Gu et al. “Comparison of Movement Related Cortical Potential in healthy people and Amyotrophic Lateral Sclerosis patients”. In: *Frontiers in neuroscience* 7 (May 2013), p. 65. DOI: 10.3389/fnins.2013.00065.
  - [12] N.Jiang et al. *Brain-Computer Interface Research: A State-of-the-Art Summary*. Vol. 3, p. 58.
  - [13] K. Dremstrup et al. “Rehabilitation Using a Brain Computer Interface Based on Movement Related Cortical Potentials – A Review”. In: *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*. Ed. by Laura M. Roa Romero. Cham: Springer International Publishing, 2014, pp. 1659–1662.
  - [14] Pierre Comon. “Independent component analysis, A new concept?” In: *Signal Processing* 36.3 (1994). Higher Order Statistics, pp. 287–314. ISSN: 0165-1684. DOI: [https://doi.org/10.1016/0165-1684\(94\)90029-9](https://doi.org/10.1016/0165-1684(94)90029-9). URL: <http://www.sciencedirect.com/science/article/pii/0165168494900299>.
  - [15] Lieven De Lathauwer and Joséphine Castaing. “Second-Order Blind Identification of Underdetermined Mixtures”. In: *Independent Component Analysis and Blind Signal Separation*. Ed. by Justinian Rosca et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 40–47. ISBN: 978-3-540-32631-1.
  - [16] Oregon Willamette University. “The Singular Value Decomposition in Symmetric (Lowdin) Orthogonalization and Data Compression”. In:
  - [17] Encyclopedia Britannica. URL: <https://www.britannica.com/science/manifold>.
  - [18] Aqsa Shakeel et al. “A Review of Techniques for Detection of Movement Intention Using Movement-Related Cortical Potentials”. In: *Computational and Mathematical Methods in Medicine* 2015 (Dec. 2015). DOI: 10.1155/2015/346217.
  - [19] Fabien Lotte et al. “A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces: A 10-year Update”. In: *Journal of Neural Engineering* 15 (Feb. 2018). DOI: 10.1088/1741-2552/aab2f2.
  - [20] Imran Khan Niazi et al. “Detection of movement intention from single-trial movement-related cortical potentials”. In: *Journal of Neural Engineering* 8.6 (Oct. 2011), p. 066009. DOI: 10.1088/1741-2560/8/6/066009. URL: <https://doi.org/10.1088%2F1741-2560%2F8%2F6%2F066009>.
  - [21] R. Xu et al. “Enhanced Low-Latency Detection of Motor Intention From EEG for Closed-Loop Brain-Computer Interface Applications”. In: *IEEE Transactions on Biomedical Engineering* 61.2 (Feb. 2014), pp. 288–296. ISSN: 0018-9294. DOI: 10.1109/TBME.2013.2294203.
  - [22] Eileen Lew et al. “Detection of self-paced reaching movement intention from EEG signals”. In: *Frontiers in Neuroengineering* 5 (2012), p. 13. ISSN: 1662-6443. DOI: 10.3389/fneng.2012.00013. URL: <https://www.frontiersin.org/article/10.3389/fneng.2012.00013>.
  - [23] Fatemeh Karimi et al. “Detection of Movement Related Cortical Potentials from EEG Using Constrained ICA for Brain-Computer Interface Applications”. In: *Front. Neurosci.* 2017.

- [24] Alexandre Barachant et al. “Classification of covariance matrices using a Riemannian-based kernel for BCI applications”. In: *Neurocomputing* 112 (July 2013), pp. 172–178. DOI: 10.1016/j.neucom.2012.12.039.
- [25] Science Direct. *Likelihood Ratio - an overview*. URL: <https://www.sciencedirect.com/topics/earth-and-planetary-sciences/likelihood-ratio>.
- [26] Xiaofei He and Partha Niyogi. *Locality Preserving Projections*. 2002.
- [27] Alexandre Barachant et al. “Riemannian Geometry Applied to BCI Classification”. In: Sept. 2010. DOI: 10.1007/978-3-642-15995-4\_78.
- [28] G. Gomez-Herrero et al. “Automatic Removal of Ocular Artifacts in the EEG without an EOG Reference Channel”. In: *Proceedings of the 7th Nordic Signal Processing Symposium - NORSIG 2006*. June 2006, pp. 130–133. DOI: 10.1109/NORSIG.2006.275210.
- [29] Carlos Sevcik. “A procedure to Estimate the Fractal Dimension of Waveforms”. In: *arXiv e-prints*, arXiv:1003.5266 (Mar. 2010), arXiv:1003.5266. arXiv:1003.5266 [nlin.CD].
- [30] B Neil Cuffin. “EEG localization accuracy improvements using realistically shaped head models”. In: *IEEE transactions on bio-medical engineering* 43 (Apr. 1996), pp. 299–303. DOI: 10.1109/10.486287.
- [31] Ou Bai et al. “Exploration of computational methods for classification of movement intention during human voluntary movement from single trial EEG”. In: *Clinical Neurophysiology* 118.12 (2007), pp. 2637–2655. ISSN: 1388-2457. DOI: <https://doi.org/10.1016/j.clinph.2007.08.025>. URL: <http://www.sciencedirect.com/science/article/pii/S1388245707004312>.
- [32] Kishore K Das N. Uttam Singh and Aniruddha Roy. “HOW TO TEST ENDOGENEITY OR EXOGENEITY: AN E-LEARNING HANDS ON SAS”. In: p. 3.
- [33] Ali Rahimi and Benjamin Recht. “Random Features for Large-Scale Kernel Machines”. In: *Advances in Neural Information Processing Systems 20*. Ed. by J. C. Platt et al. Curran Associates, Inc., 2008, pp. 1177–1184. URL: <http://papers.nips.cc/paper/3182-random-features-for-large-scale-kernel-machines.pdf>.
- [34] Y. Liu, C. Huang, and Y. Hsiao. “Controlling the False Positive Rate of a Two-State Self-Paced Brain-Computer Interface”. In: *2013 IEEE International Conference on Systems, Man, and Cybernetics*. Oct. 2013, pp. 1476–1481. DOI: 10.1109/SMC.2013.255.